UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics
of the University of Lübeck
Director: Prof. Dr. Jürgen Prestin

Symbolic-based analysis techniques and their information contents

Symbolbasierte Analysetechniken und ihr Informationsgehalt

Dissertation

for Fulfillment of Requirements
for the Doctoral Degree of the University of Lübeck
— from the Department of Computer Sciences —

submitted by
Inga Lisbeth Stolz
from Bobingen, Germany

Lübeck 2018

First referee: Prof. Dr. Karsten Keller

Second referee: Prof. Dr. Alexander Schnurr

Chairman: Prof. Dr. Jan Modersitzki

Date of oral examination: July 17, 2018

Approved for printing. Lübeck, July 18, 2018.

Abstract

The abundance of data that is gathered every day with the ultimate goal to understand and to predict the world around us requires techniques to analyze the data efficiently and accurately. Methods that are widely used in this context involve a decoding of the collected data into a sequence of symbols following a certain symbolic scheme. These symbolic-based analysis techniques facilitate not only the analysis and increase the numerical efficiency in most cases, but also have a foundation in the mathematical fields of information and system theory. For instance, the distribution of symbols or the determination of entropy changes can be used to detect temporal characteristics in the data.

In this thesis, symbolic schemes are formalized and information-theoretic similarities and differences of symbolic-based analysis techniques are highlighted. In particular, a formalization of symbolic schemes is presented that is natural and sufficiently general. It includes, for instance, threshold crossings, the relatively new ordinal symbolic approach which goes back to innovative works of Bandt and Pompe, and variants of both approaches. Moreover, results achieved in ordinal symbolic dynamics are generalized via this formalization to show that symbolic schemes which regard a dependency between two measured values provide, under very natural conditions, a route to the Kolmogorov-Sinai entropy. This is a substantial advantage over symbolic schemes such as threshold crossings or other variants where the decoding is only performed on a one-dimensional level, and a scheme has to be found that characterizes the Kolmogorov-Sinai entropy directly.

The results show that a theoretical study of different symbolic schemes at once gives further insight into differences and similarities of symbolic-based analysis techniques and increases the spectrum of methods and potential applications.

Zusammenfassung

Symbolbasierte Analysetechniken finden ein breites Echo in der Zeitreihenanalyse und werden vielfältig eingesetzt und weiterentwickelt. Bekannte Verfahren sind zum Beispiel Schwellwert-Methoden oder Methoden, die auf der ordinalen Idee basieren. Die ordinalen Verfahren gehen auf innovative Arbeiten von Bandt und Pompe zurück und werden unter anderem in der Analyse von EEG-Daten zur Identifikation von Schlafphasen oder epileptischen Anfällen genutzt. Im Mittelpunkt einer symbolbasierten Analysetechnik steht eine Übersetzung experimenteller oder simulierter Daten in eine Symbolsequenz, indem ein bestimmtes Symbolisierungsschema angewendet wird. Diese Symbolfolge kann meistens effizienter und auch robuster gegenüber den Rohdaten analysiert werden. Außerdem können durch Symbolverteilungen oder Änderungen der Entropie zeitliche Merkmale der Zeitreihe schneller identifiziert werden. Das ist insbesondere interessant, wenn zum Beispiel ein drohender Herzinfarkt, ein epileptischer Anfall, ein Wetterumschwung oder eine seismologische Aktivität vorhergesagt werden soll.

In dieser Arbeit werden Symbolisierungsverfahren natürlich formalisiert sowie auf informationstheoretische Gemeinsamkeiten und Unterschiede untersucht. Hierbei wird angenommen, dass ein zeitabhängiges System als maßerhaltendes dynamisches System modelliert werden kann und dass jedes Symbolisierungsschema eine Partition des zugrunde liegenden Zustandsraumes liefert. Durch die Formalisierung wird gezeigt, dass Symbolisierungsschemata, die eine Abhängigkeit zwischen zwei gemessenen Werten berücksichtigen, unter sehr natürlichen Bedingungen einen Weg zur Kolmogorov-Sinai-Entropie liefern. Dies ist ein wichtiger Unterschied zu Schemata, die bei der Dekodierung nur eine Beobachtung berücksichtigen. Der Unterschied ist insbesondere interessant, wenn kein Schema gefunden werden kann, welches die Kolmogorov-Sinai-Entropie direkt charakterisiert, d.h. wenn die zugrunde liegende Partition nicht generierend unter der Dynamik des Systems ist.

Die Ergebnisse zeigen, dass eine simultane Untersuchung verschiedener Verfahren zur Symbolisierung dem Anwender nicht nur eine größere Anzahl an Symbolisierungsschemata zur Verfügung stellt, sondern auch einen zusätzlichen Einblick in theoretische Unterschiede und Gemeinsamkeiten schafft.

Acknowledgments

To my dearly-loved parents Christiane and Manfred: because I owe it all to you.

First I would like to express my sincere gratitude to my supervisor, Prof. Dr. Karsten Keller, for his continuous support, motivation and patience. Further, I would like to thank the departments of mathematics at the University of Lübeck and the Europa-Universität Flensburg for giving me the opportunity to conduct this research. There was always room for helpful advice and joyful breaks.

Moreover, I thank the DAAD RISE Program for giving me the chance to work with Paige Hillen, and I thank Paige Hillen herself. At the time of her internship she collected, for example, the seismological data and related metadata considered in this thesis by using the facilities of IRIS Data Services, and specifically the IRIS Data Management Center. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681 and the DOE National Nuclear Security Administration. The seismic instruments were provided by IRIS through the PASSCAL Instrument Center at New Mexico Tech.

Last but not least, I would like to thank all my family and friends for proofreading my thesis and giving me the thumbs up. I do not really know how to express my love for your support and trust. In particular, I wish to express my gratitude to Claudine Leonhard, Sabrina Kombrink and Christian Wülker for their precious time and for their many comments and helpful suggestions.

Contents

Chapter 1

Introduction

Symbolic-based analysis techniques are efficient and popular research tools for analyzing real-world time series that are continuously refined and improved. The data mostly stems from non-linear and possibly chaotic time-dependent systems, and therefore, is not easy to understand. In the application of these techniques, the central task is to decode the experimental data into a sequence of symbols following a certain symbolic scheme, and subsequently, conflating successive symbols into symbol words. In most cases, the analyst analyzes this symbol sequence or symbol word sequence not only more efficiently and robustly than the original data, but can also rely on tools of information and system theory (see for instance Cover and Thomas [25], Choe [23] and Lind and Marcus [64]). Thus interesting temporal patterns that are hidden in the data can be exposed by determining, for example, the symbol distribution or changes in the entropy. This is particularly interesting when trying to predict an impending epileptic seizure, cardiac infarction, weather change, seismological activity or the like.

Popular methods of symbolic-based analysis techniques are, for instance, based on threshold crossings and the relatively new ordinal symbolic approach, which goes back to the innovative works of Bandt and Pompe [11] and Bandt et al. [10]. Further techniques and application examples are listed in the review papers of Kurths et al. [60], Daw et al. [27], Zanin et al. [88] and the examples in biology, medicine, artificial intelligence, data mining—just to mention a few—given therein. Moreover, see the contributions to the special topic *Recent Progress in Symbolic Dynamics and Permutation Complexity. Ten Years of Permutation Entropy* of The European Physical Journal [6] and to the special issue *Symbolic Entropy Analysis and Its Applications* of Entropy [65] (see also Section 1.1).

In fact, there exists a great pool of symbolic schemes to choose from, where the simplest variants result in sequences of single bit values. Hence the analyst faces the challenge to pick symbolic schemes and word lengths that are most suitable and reliable for the specific data that has to be evaluated and interpreted. Overall, the analyst can rely on know-how, practical experience or on theoretical results and ideas. However, although there exists a lot of literature with convincing results and application examples for each symbolic scheme (see the references given previously), little attention has been paid to the study of different symbolic schemes and their information content at once. The main objective of this thesis is to give the analyst an additional overview of symbolic-based analysis techniques.

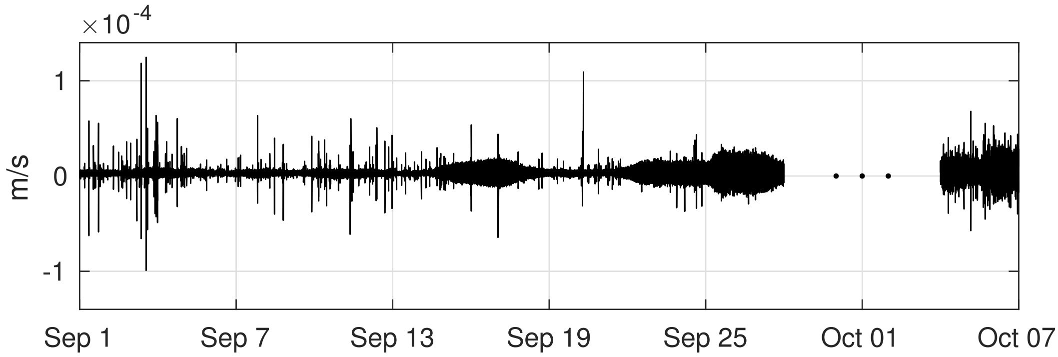
In fact, by studying the different symbolic schemes at once via a natural formalization, we generalize results achieved in ordinal symbolic dynamics and show thereby that symbolic schemes that regard a dependency between two measured values, such as the techniques based on the ordinal idea, provide, under very natural conditions, a route to the Kolmogorov-Sinai entropy (KS entropy). This means that the analyst who picks such a scheme can skip the search for a technique that characterizes the KS entropy directly. The search for such a scheme is hopeless in a lot of cases anyway, for instance, in the presence of noise (see Crutchfield and Packard [26], Bollt et al. [15], Daw et al. [27], Kennel and Buhl [56] and the references given therein). Nevertheless, applying symbolic-based analysis techniques is a trade-off between computational capacity and computational accuracy. Thus we still recommend to compare different schemes in the finite setting of applications since even if the KS entropy is not characterized directly, the underlying symbolic scheme can still be efficient and sufficient in order to find and quantify relevant information in the analysis process (see Daw et al. [27]).

Note that some of the results presented in this thesis are published in Keller et al. [50, 49] and Stolz and Keller [79].

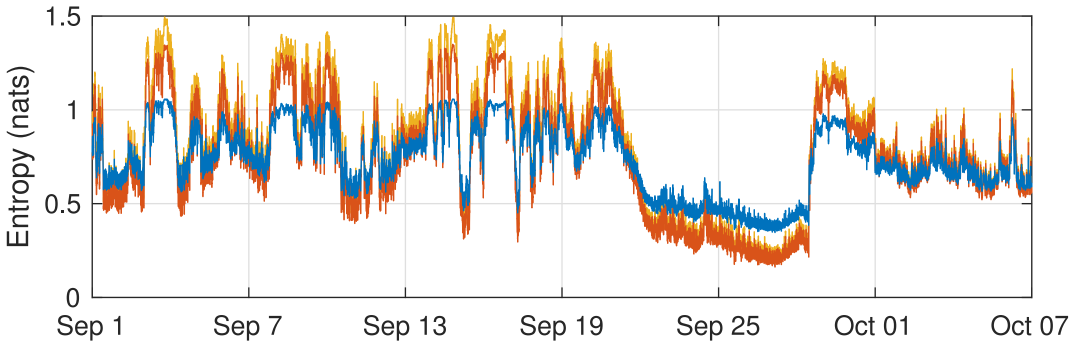
1.1 An application example: a symbolic-based analysis of seismic data

The aim of this section is to motivate symbolic-based analysis techniques as a tool to analyze real-world data. Symbolic based-techniques are applied in many research areas as the authors summarize in Kurths et al. [60], Daw et al. [27], Zanin et al. [88], in the special topic [6] and the special issue [65] (see also Section 1.1). In particular, there exist many scientific publications on analyzing electroencephalogram-data (EEG data) by applying techniques that are based on the ordinal idea. For instance, the empirical Permutation entropy or the ordinal Conditional entropy are determined in order to detect complexity changes in the EEG data. The results are used to describe the considered EEG data, to detect epileptic seizures or different sleeping states. For a fuller treatment, we refer the reader to Keller and Lauffer [48], Cao et al. [21], Keller et al. [54], Unakafova [85], Unakafov [82], Berger et al. [12] and the references given therein. The authors show that in a lot of cases—depending on the patients, the choice of the channel, etc.—symbolic-based complexity measures decrease during an epileptic seizure.

These results raise the question, whether analog results can be obtained by analyzing data that has similar characteristics as EEG data, such as audio data, physiological data (heart rate, blood pressure, stroke volume, systemic vascular resistance, etc.), climate data and so on. The question is answered by the researchers Glynn and Konstantinou in their publication *Reduction of Randomness in Seismic Noise as a Short-term Precursor to a Volcanic Eruption* [36]. The authors discuss the decrease of empirical Permutation entropy in seismic noise, eight days prior to the 1996 Gjálp eruption in Iceland. The eruption, accompanied by an intense earthquake swarm, took place in the most productive area of Iceland's hotspot and was preceded by an earthquake at the northern rim of the Bárdarbunga caldera of magnitude 5.4 (Richter scale) on the 29th of September in 1996 (see Einarsson et al. [30]). The data was collected with 20 samples per second by the temporary seismic network



(a) Original seismic noise (the data of the earthquake and the first days of the eruption are left out).



(b) Empirical Permutation entropy (blue), ordinal Conditional entropy (red) and a weak version of empirical Permutation entropy (yellow) for $d = 3$ and a window length of 12000 samples (for more details see Section 3.2.2, Section 3.5 and Chapter 4).

Figure 1.1: The curves in **(b)** reflect the 1996 Gjálp eruption in Iceland by a decrease of their values eight days before the eruptive activity.

HOTSPOT with a PASSCAL Guralp CMG3-ESP Broadband Sensor (see the FDSN entry [40] and Allen et al. [2, 3, 4]).

In the following, we touch only on a few aspects of the data acquisition, and mainly focus on the usage of symbolic-based analysis techniques in order to identify complexity changes. For a deeper discussion of the experimental setup and the results in the seismological context, we refer the reader to the work of Glynn and Konstantinou [36] and the references given therein.

We analyzed the vertical component of broadband seismic data collected by station HOT23 at Grímsvötn that is eight kilometers south of Gjálp from the 1st of September to the 7th of October in 1996 (see Figure 1.1**(b)**). We used the facilities of IRIS Data Services and, in particular, the IRIS Data Management to access the data and related metadata (`http://ds.iris.edu/mda/XD/HOT23?timewindow=1996-1998`). In Figure 1.1**(a)**, we skipped the days of high seismic activity in order to zoom in on the seismic noise, as otherwise the fluctuations would not be visible to the naked eye.

We computed the empirical Permutation entropy (see Unakofova [85] and Unakafova and Keller [84]), the empirical ordinal Conditional entropy (see Unakafov [82] and

Unakafov and Keller [83]) and a weak version of empirical Permutation entropy, as described in Keller and Sinn [51], for the ordinal order $d = 3$, the time delay $\tau = 1$ and a sliding window of 12000 samples (see also Section 3.2.2, Section 3.5 and Chapter 4). The window length is equivalent to 10 minutes of data acquisition. The choice of $d, \tau \in \mathbb{N}$ and the window length was motivated by our results presented in Figure 4.1 and facilitates the comparison of the three methods. Note that Glynn and Konstantinou picked the ordinal order $d = 5$, a sliding window of 20 minutes and the time delay $\tau = 5$. In fact, the choice and role of the delay $\tau \in \mathbb{N}$ is an interesting theme in itself. More details on this topic can be found in Keller et al. [50], in Berger et al. [12], in the supplementary information published along with the manuscript of Glynn and Konstantinou [36] (see also Section 3.2.2 and the references given therein). For our numerical computations, we used MATLAB R2017b [67] and the MATLAB scripts given by Unakafova [86] in order to compute the empirical Permutation and ordinal Conditional entropy in sliding windows in a fast way (see also Unakafova and Keller [84]).

Indeed, the curves of empirical entropy reflect the 1996 Gjálp eruption in Iceland by a decrease of their values. The decrease starts, as Glynn and Konstantinou also point out, on the 22nd of September and reaches a trough on the 27th of September. The curves increase rapidly on the day of the Bárdarbunga earthquake. Moreover, it is noticeable that the magnitude of fluctuations between the 22nd and 29th of September differs from the magnitude previous to September 22nd. Overall, the results, as the ones presented here, emphasize that a symbolic-based analysis of experimental data can detect interesting complexity changes. However, since all three curves behave similarly, but the peak-to-through fluctuations differ in their height, this section also shows that the choice of a symbolic-based measure is not evident and should be made, for instance, in accordance with the application purpose and the algorithmic efficiency and accuracy.

1.2 Outline of this thesis

In the following, we give a brief overview of the content of this thesis with some complementary remarks and close this chapter with Figure 1.2 in order to show the leitmotif of this thesis and to put the results into context.

In Chapter 2, we first introduce the mathematical framework: We assume that an underlying time-dependent system can be modeled as a measure-preserving dynamical system, and a measuring process by a sequence of real-valued random variables that we call observables. Further, a symbolic-based analysis technique or rather the underlying symbolic scheme entails a finite partition of the underlying state space of the system, and thus we review symbolic dynamics in this context. Secondly, we introduce the concept of ergodicity and recall Birkhoff's ergodic theorem. Ergodicity is needed in several statements in this thesis to ensure that the analysis of the underlying system cannot be simplified by studying subsystems and that properties of the considered system can be recovered from the outcome of a time-infinite measurement (in a measurable way).

A main part of Chapter 2 is devoted to the KS entropy, i.e. we review some of the standard facts on the KS entropy and summarize different possibilities to theoretically determine this complexity measure. In this respect, we also look more closely at the wide range of theoretical entropy measures accompanying symbolic dynamics. We close Chapter 2 by stating sufficient conditions on the measuring process such that the information of the underlying system is preserved. This is a basic prerequisite for estimating the complexity of a time-dependent system by a symbolic-based analysis technique since information that is lost during the measuring process cannot be restored. Moreover, we present sufficient conditions on a discretization of the state space such that the information of the underlying system is preserved.

Chapter 3 is devoted to the study of which symbolic-based analysis techniques provide a route to the KS entropy. In this context, the main results of this thesis are stated. At first, we present a formalization of symbolic schemes that is natural and sufficiently general, i.e. it includes threshold crossings, the ordinal symbolic approach and variants of both approaches (see Section 3.2). Secondly, in Theorem 3.2, we state sufficient conditions on symbolic-based analysis techniques such that they preserve, under relatively weak assumptions, the information given by the observables, and in Theorem 3.3, we state conditions such that even the information given by the measuring process is preserved. In doing so, we demonstrate, in Theorem 3.1, a main advantage of the ordinal approach: The ordinal approach and variants of it provide, under very natural conditions, a route to the KS entropy. In order to show our key findings, we assume that underlying dynamics are ergodic. For completeness, we include a study of the non-ergodic case. Moreover, we give an example of a symbolic-based analysis technique that does not preserve the information given by the measuring process in general. We close Chapter 3 with a summary of the results and concluding remarks.

In Chapter 4, we show how the presented theory can be applied by analyzing simulated data. We apply different empirical complexity measures based on different symbolic schemes that play a role in this thesis, and study the dependence on the orbit length. In doing so, we reflect on the practical difficulty to choose a symbolic-based measure from a big pool of possibilities. We consider a one-dimensional orbit of a logistic map and a two-dimensional orbit of Arnold's cat map.

This thesis includes an appendix where the interested reader finds definitions and properties from measure theory and topology that are relevant in our discussions. Moreover, at this point, we would like to refer to two books that assisted the process of writing: The booklet by Trzeciak [81] provides a major support in writing mathematical theses by summarizing common phrases and problems. The handbook of mathematics by Bronstein et al. [18] is a good guide of mathematical knowledge.

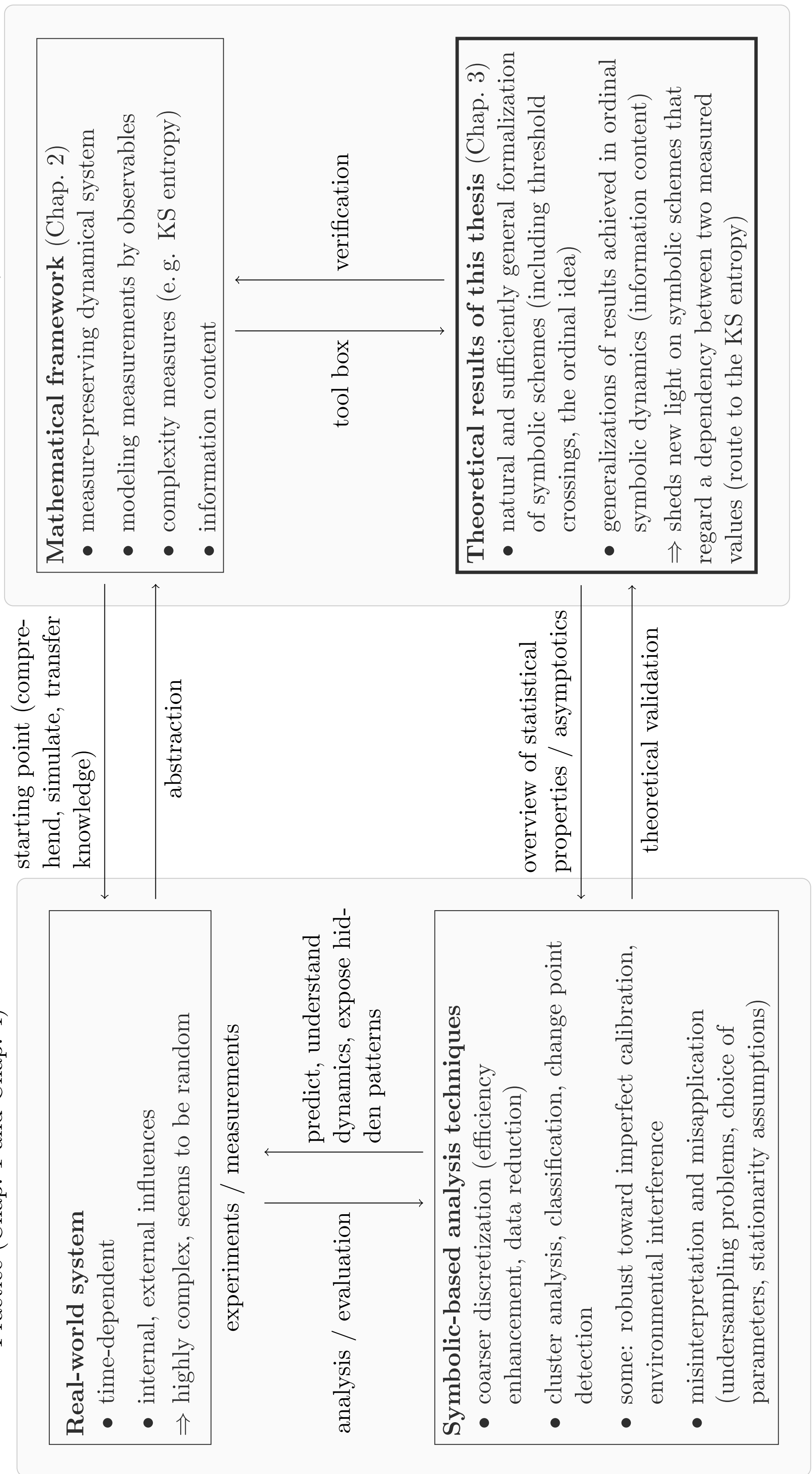


Figure 1.2: Results of this thesis put into context.

Chapter 2

The mathematical framework

This chapter introduces the mathematical framework of this thesis along with some basic notation and concepts (see also the appendix). For a more detailed treatment of measure theory and topology, we refer the reader to Brin and Struck [17], Billingsley [13], Munkres [68] and the references given therein. Further, Section 2.2 is devoted to the Kolmogorov-Sinai entropy which is an important complexity measure of dynamical systems (see for instance Walters [87] and Amigó [5]). Moreover, we look more closely at the information content of metric-based analysis techniques and symbolic dynamics, and study whether the information of the underlying system is preserved. For a detailed review on system, estimation and ergodic theory, the theory of signal processing, differential equations, symbolic dynamics, recurrence and sensitivity analysis, we refer the reader to Marx and Vogt [66], Denker [28] and the references given therein.

2.1 Measure-preserving and ergodic dynamical systems

This section is organized as follows. Firstly, we introduce how we model both a real-world system that evolves over time and a measuring process of this system. Secondly, we look more closely at symbolic-based analysis techniques or rather at the underlying symbolic schemes in this mathematical framework. Note that this framework is generally used to study time-dependent systems and symbolic-based analysis techniques (see for instance Gutman [39], Amigó et al. [7], Keller et al. [50] and the references given therein).

In this thesis, we model a real-world system by a *(discrete) dynamical system* (Ω, T), which consists of a non-empty set Ω, and a map $T : \Omega \hookleftarrow$. For $t \in \mathbb{N}_0$, the t-th iterate of ω under T is given by a t-fold composition

$$T^{\circ t}(\omega) := \overbrace{T(T(T(\ldots T}^{t \text{ times}}(\omega)))) \text{ for } t \in \mathbb{N} \text{ and } T^{\circ 0}(\omega) := \omega.$$

A dynamical system is usually used as a basic model in order to study a time-dependent system. We call the sequence $(T^{\circ t}(\omega))_{t \in \mathbb{N}_0}$ the *orbit* of $\omega \in \Omega$ with respect to T. A state $\omega \in \Omega$ is *fixed* if $T^{\circ t}(\omega) = \omega$ for all $t \in \mathbb{N}_0$, and *periodic of period* $t \in \mathbb{N}$ if $T^{\circ t}(\omega) = \omega$. The *minimal period* of a periodic state $\omega \in \Omega$ is the smallest $t \in \mathbb{N}_0$ such that $T^{\circ t}(\omega) = \omega$. We denote by $T^{\circ -t}(A) = (T^{\circ t})^{-1}(A)$ the preimage of $A \subseteq \Omega$ under $T^{\circ t}$, i.e. the set $\left\{\omega \in \Omega \mid T^{\circ t}(\omega) \in A\right\}$.

In accordance to application, which includes a discrete data acquisition starting at a finite time, we consider $t \in \mathbb{N}_0$ and talk about a *(discrete-time) dynamical system.* Thus we interpret $(T^{\circ t}(\omega))_{t\in\mathbb{N}_0}$ as the *dynamical evolution of a state* $\omega \in \Omega$ *over time*, and call Ω *state space* and $t \in \mathbb{N}_0$ *time point.* Henceforth, we consider a dynamical system (Ω, T) within a probability space $(\Omega, \mathcal{A}, \mu)$ where the map $T : \Omega \hookleftarrow$ is $\mathcal{A}$-$\mathcal{A}$-measurable, and denote it by $(\Omega, \mathcal{A}, \mu, T)$. Moreover, we say the *information content of* $(\Omega, \mathcal{A}, \mu, T)$ is based on the measure μ on the σ-algebra $\mathcal{A}$.

The following definition, for instance, can be found in Walters [87, Definition 4.4].

Definition 2.1. Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. We say a sub-σ-algebra $\mathcal{F} \subset \mathcal{A}$ *preserves almost surely the information* given by a sub-σ-algebra $\mathcal{F}' \subseteq \mathcal{A}$, and write $\mathcal{F}' \overset{\mu}{\subset} \mathcal{F}$ if for each $F' \in \mathcal{F}'$ there exists some $F \in \mathcal{F}$ with

$$\mu\left(F \bigtriangleup F'\right) = 0,$$

where $\bigtriangleup$ denotes the symmetric difference of two sets.

♣ **Example 2.1** (Logistic maps on the unit interval and rotations on the unit circle)**.** Classical and simple examples of discrete-time and non-linear dynamical systems are given on the unit interval and on the unit circle. For a two-dimensional example, we refer the reader to Remark 4.3.

Let $r \in (0, 4]$, $\Omega = [0, 1)$, and consider a transformation $T_r : \Omega \hookleftarrow$, defined by

$$T_r(\omega) = r\omega(1 - \omega)$$

for all $\omega \in \Omega$, i.e. T_r is a *logistic map* on the unit interval. The graph of the map $T_4 : [0, 1) \hookleftarrow$ and the orbit $\left(T^{\circ t}(0.4854)\right)_{t=0}^{9}$ are given in a cobweb plot in Figure 2.1 on the left, and, on the right, the preimage $T_4^{\circ -1}([a, b))$ of an interval $[a, b) \subset [0, 1)$ is exemplified. Note that, for all $\omega \in [0, 1)$, it holds

$$T_4^{\circ -1}(\omega) = \frac{1 \mp \sqrt{1 - \omega}}{2}.$$

Another popular example of a discrete-time dynamical system is given on the unit circle $\Omega = \{z \in \mathbb{C} \mid |z| = 1\}$ by the transformation $T_a : \Omega \hookleftarrow$ with $a \in \mathbb{R}$ and

$$T_a(z) = e^{2\pi \mathrm{i} a} z$$

for all $z \in \Omega$, i.e. T_a is a *rotation* on the unit circle. For simplicity, the states in Ω are identified with those in the unit interval $[0, 1)$ in the obvious way $\omega \mapsto z = \mathrm{e}^{2\pi \mathrm{i}\omega}$. Then the rotation reads as follows

$$T_a(\omega) = (\omega + a) \mod 1$$

for all $\omega \in \Omega$. Note that for $[b, c) \subset [0, 1)$ with $b + a < 1 \leq c + a$ the following holds

$$T_a([b, c)) = [0, c + a) \cup [b + a, 1). \qquad ♣$$

The following definition, for instance, can be found in Walters [87, Definition 1.1].

Definition 2.2. An $\mathcal{A}$-$\mathcal{A}$-measurable map $T : \Omega \hookleftarrow$ is called *μ-preserving* if

$$\mu\left(T^{\circ-1}(A)\right) = \mu(A)$$

for all $A \in \mathcal{A}$. Equivalently, the measure μ is called *T-invariant.*

From now on, we make the assumption: $(\Omega, \mathcal{A}, \mu, T)$ is a measure-preserving dynamical system, i.e. T is μ-invariant.

♣ **Example 2.2** (Measure-preserving dynamical systems)**.** The rotation T_a with $a \in \mathbb{R}$ is invariant with respect to the Lebesgue measure λ. In order to see this, recall that the family of half-closed intervals $\xi := \{[m, n) \mid m, n \in \mathbb{Q} \cap [0, 1), m < n\}$ is closed under the formation of finite intersections and generates $\mathcal{B}([0, 1))$ (see Appendix B.5). Since

$$T_a^{\circ-1}([b, c)) = \begin{cases} [b-a, c-a) & \text{if } 0 \le b-a < c-a, \\ [0, c-a) \cup [b-a+1, 1) & \text{if } b-a < 0 \le c-a, \\ [b-a+1, c-a+1) & \text{if } b-a < c-a \le 0 \end{cases}$$

for any $[b, c) \in \xi$, it holds $T_a^{\circ-1}(A) \in \mathcal{B}([0, 1))$ for all $A \in \xi$, i.e. T_a is $\mathcal{B}([0, 1))$-$\mathcal{B}([0, 1))$ measurable (see Billingsley [13, Theorem 13.1.]). Moreover, $\lambda\left(T_a^{\circ-1}(A)\right) = \lambda(A)$ for all $A \in \xi$, i.e. $\lambda\left(T_a^{\circ-1}(B)\right) = \lambda(B)$ for all $B \in \mathcal{B}([0, 1))$ (see Billingsley [13, Theorem 3.3.]). Consequently, T_a is μ-preserving.

In the case of the logistic map T_4, there exists an invariant measure μ_4 with density $p : [0, 1) \to [0, \infty)$ given by

$$p(\omega) = \frac{1}{\pi\sqrt{\omega(1-\omega)}}$$

for all $\omega \in (0, 1)$. In particular, the measure μ_4 is *absolutely continuous* with respect to the Lebesgue measure, i.e. $\mu_4(A) = 0$ for all $A \in \mathcal{B}([0, 1))$ whenever $\lambda(A) = 0$ (see for instance Amigó [5, Section 1.1.3] and Chan and Tong [22] and the references given therein). ♣

The following definition, for instance, can be found in Walters [87, Definition 1.2].

Definition 2.3. A μ-preserving map $T : \Omega \hookleftarrow$ is called *ergodic* with respect to the probability measure μ (or rather μ is ergodic with respect to T) if for all $A \in \mathcal{A}$ with $T^{\circ-1}(A) = A$ either $\mu(A) = 0$ or $\mu(A) = 1$. In this case $(\Omega, \mathcal{A}, \mu, T)$ is called an *ergodic dynamical system.*

Ergodicity means that the study of the dynamics given by T cannot be simplified by restricting T to the sets $A \in \mathcal{A}$ with $T^{\circ-1}(A) = A$ (see for instance Walters [87], Krengel [58] and Remark 3.4).

★ **Remark 2.1** (Birkhoff's ergodic theorem)**.** Major achievements in ergodic theory are the various ergodic theorems. Most relevant for our purposes is *Birkhoff's ergodic theorem.* The theorem legitimates the use of

$$\left(T^{\circ t}(\omega)\right)_{t \in \mathbb{N}_0}$$

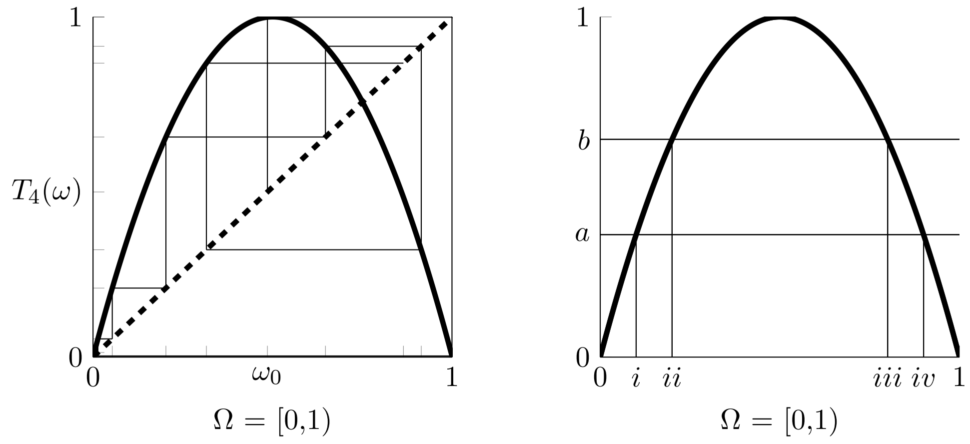


Figure 2.1: Graph of the logistic map $T_4 : [0,1) \hookleftarrow$. Left: the orbit $\left(T_4^{\circ t}(\omega_0)\right)_{t=0}^{9}$ with $\omega_0 = 0.4854$ illustrated by a cobweb plot. Right: preimage $T_4^{\circ -1}([a,b))$ of an interval $[a,b) \subset [0,1)$, whereby $i, iv = \frac{1 \mp \sqrt{1-a}}{2}$ and $ii, iii = \frac{1 \mp \sqrt{1-b}}{2}$. Note how shifting the interval $[a,b)$ within the unit interval determines the size of the intervals $[i, ii)$ and $[ii, iv)$ with respect to the Lebesgue measure λ.

for almost every $\omega \in \Omega$ in order to recover properties of the underlying system. Thus ergodicity has a great physical significance (see for instance Bogachev [14], and Collet and Eckmann [24] for more details). The following theorem, for instance, can be found in Choe [23].

Theorem 2.1 (Birkhoff's ergodic theorem)**.** *Let $(\Omega, \mathcal{B}(\Omega), \mu, T)$ be a measure-preserving dynamical system and*

$$X : (\Omega, \mathcal{B}(\Omega), \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

some μ-integrable function, i.e.

$$\int_\Omega |X| \, \mathrm{d}\mu(\omega) < \infty.$$

Then there exists some μ-integrable function X^* with $X^*(T(\omega)) = X^*(\omega)$ for all $\omega \in \Omega$ such that

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} X\left(T^{\circ s}(\omega)\right) = X^*(\omega)$$

for almost every $\omega \in \Omega$. Supplementary, if $(\Omega, \mathcal{B}(\Omega), \mu, T)$ is ergodic, then X^* is constant and

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} X\left(T^{\circ s}(\omega)\right) = \int_\Omega X \, \mathrm{d}\mu(\omega) \tag{2.1}$$

for almost every $\omega \in \Omega$.

Consider the characteristic function $\mathbf{1}_A : \Omega \to \{0,1\}$ for all $A \in \mathcal{A}$ in (2.1). This gives

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{1}_{\{T^{\circ s}(\omega) \in A\}} = \int_\Omega \mathbf{1}_A(\omega)\, \mathrm{d}\mu(\omega) = \mu(A)$$

for almost every $\omega \in \Omega$. By Birkhoff's ergodic theorem, the measure of A coincides with the relative frequency of how often the state ω visits A over time. ★

If T is not ergodic, it can be interesting to ask whether $(\Omega, \mathcal{A}, \mu, T)$ can be decomposed into ergodic subsystems. An answer to this question is given by the *ergodic decomposition theorem*, which we introduce in Remark 3.4 (consult also Einsiedler and Ward [33], Einsiedler et al. [31], Quas [72] and the references given therein).

♣ **Example 2.3** (Non-ergodic and ergodic dynamical systems, Example 2.2 continued)**.** The rotation T_a is not ergodic with respect to the Lebesgue measure λ if a is rational. In order to see this, denote by id the *identity map*, and let $p \in \mathbb{N}$ and $q \in \mathbb{N}$ be coprime with $p < q$, then $a := \frac{p}{q}$ is rational. It follows easily that $T_a^{\circ q} = \mathrm{id}$, and thus every state $\omega \in \Omega$ is periodic of period q. Let now, for instance,

$$A = \bigcup_{t \in \{0,1,\ldots,q-1\}} T_a^{\circ t}\left(\left[0, \frac{p}{q^2}\right)\right),$$

then $T_a^{\circ -1}(A) = A$, but $0 < \lambda(A) = q\frac{p}{q^2} = \frac{p}{q} < 1$, and thus T_a with $a = \frac{p}{q}$ is not ergodic as claimed. However, λ can be decomposed into ergodic components (see Remark 3.4 for a brief exposition of the topic).

In contrast, if a is irrational, then the rotation T_a is ergodic with respect to λ (see for instance Brin and Stuck [17, Chapter 4]). In order to show the difference between rational and irrational rotations, we illustrate in Figure 2.2 different orbits for $a = \frac{1}{4}$ and an orbit segment for $a = \sqrt{2}$. Moreover, the irrational rotation T_a holds *dense orbits* $\left(T_a^{\circ t}(\omega)\right)_{t \in \mathbb{N}_0}$ in $[0,1)$ for all $\omega \in \Omega$ (see for instance Hasselblatt and Katok [42, Section 4.1.2]). Hence, for any $\omega \in \Omega$ and howsoever small interval $[b,c)$ with $b, c \in \Omega$ and $b < c$, there exists a point $s \in \mathbb{N}_0$ such that $T_a^{\circ s}(\omega) \in [b,c)$. This implies that $[0,1)$ is the only closed non-empty element $A \in \mathcal{B}([0,1))$ fulfilling $T^{\circ -1}(A) = A$. In fact, the following holds (see for instance Walters [87, Chapter 1]).

Proposition 2.1. *Let (Ω, d) be a compact metric space and μ be a Borel measure on Ω (see Appendix A.3, A.5 and B.8). If T is ergodic and $\mu(A) > 0$ for all open non-empty subsets A of Ω, then almost every orbit of $\omega \in \Omega$ is dense in Ω.* ♣

2.1.1 Modeling measurements

In general, the laws of a time-dependent system to be analyzed are unknown, i.e., in most cases, it is impossible to study the dynamics of a real-world system directly. Therefore, measurements or rather *observations* have to be made and further investigated (see for instance the introductory remarks of Isermann [44], and the natural

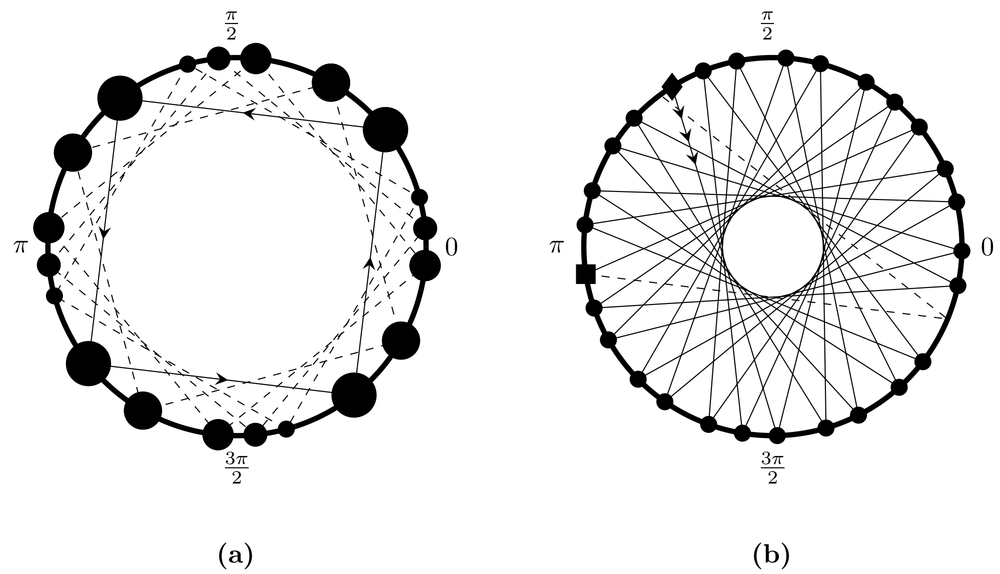


Figure 2.2: Graph of the rotation map on the unit circle. **(a)**: five random starting states and their orbits with $a = \frac{1}{4}$ (indicated by markers of the same size). **(b)**: an orbit segment of length 28 (square) for $a = \sqrt{2}$ and a random starting state $\omega_0 \in \Omega$ (diamond).

scientific discussion of Hively et al. [43]). In the following, these measurements are modeled by a countably infinite set of random variables $X_i : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and by the map T, i.e.

$$\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} : \Omega \to \left(\mathbb{R}^{\mathbb{N}\times\mathbb{N}_0}, \mathcal{B}\left(\mathbb{R}^{\mathbb{N}\times\mathbb{N}_0}\right)\right),$$

where $\mathbf{X} = (X_i)_{i\in\mathbb{N}} : (\Omega, \mathcal{A}, \mu) \to \left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)\right)$. The sequence $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ is called *measuring process*, $\left(\mathbf{X}\left(T^{\circ t}(\omega)\right)\right)_{t\in\mathbb{N}_0}$ a *realization* of $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ for $\omega \in \Omega$ and the random variables X_i are called *observables*.

Note that, in accordance with our framework, we keep definitions regarding the measuring process $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ as general as possible. The finite case, meaning that there is no information gain when additionally $\left(X_i \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ with $i \geq n$ and $n \in \mathbb{N}$ is considered, fits into our framework by assuming that all $\left(X_i \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ with $i \geq n$ coincide. Nevertheless, in some proofs, we work with $\mathbf{X} = (X_i)_{i=1}^n$ and $\mathbf{X} = X$, respectively, where $X : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable.

In order to emphasize the structure, we write $\left(\mathbb{R}^{\mathbb{N}\times\mathbb{N}_0}, \mathcal{B}\left(\mathbb{R}^{\mathbb{N}\times\mathbb{N}_0}\right)\right)$, though by the countability of $\mathbb{N} \times \mathbb{N}_0$, we restrict discussions to $\mathbb{R}^{\mathbb{N}}$ (see also Appendix B.8). Subsequently, we consider the following sub-σ-algebras (see Appendix B.5):

$$\sigma(\mathbf{X}) = \bigvee_{i\in\mathbb{N}} \sigma(X_i) = \bigvee_{i\in\mathbb{N}} \left\{X_i^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\right\} \subset \mathcal{A},$$

$$\sigma\left(\mathbf{X} \circ T^{\circ s}\right) = \bigvee_{i\in\mathbb{N}} \sigma\left(X_i \circ T^{\circ s}\right) = \bigvee_{i\in\mathbb{N}} \left\{\left(X_i \circ T^{\circ s}\right)^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\right\} \subset \mathcal{A} \text{ and}$$

$$\sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right) = \bigvee_{t\in\mathbb{N}_0} \sigma\left(\mathbf{X} \circ T^{\circ t}\right) = \bigvee_{t\in\mathbb{N}_0} \bigvee_{i\in\mathbb{N}} \sigma\left(X_i \circ T^{\circ t}\right) \subset \mathcal{A}$$

for $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ being a sequence of observables and $s \in \mathbb{N}$.

★ **Remark 2.2** (The reason why we introduce observables)**.** Commonly, measurements are applied to determine quantitative properties, for instance, the linear momentum or the speed of movement of an examined object. In doing so, it is assumed that measurements provide the state of a system directly. In this thesis, we assume that hidden states are measured (compare to Gutman [39] and the references given therein) in order to determine whether information of the underlying system is lost during the measuring process or during a symbolic-based analysis. Even though observables are superfluous if states are measured directly, we model these measurements by a sequence $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ of random variables X_i where X_i is the i-th coordinate projection, i.e. $X_i(\omega) = x_i$ for $\omega = (x_1, x_2, \ldots)$. ★

★ **Remark 2.3** (Random process)**.** In this thesis, we model measurements by a measuring process in the aforementioned way. Another possibility is to model measurements by a sequence $(Y_t)_{t\in\mathbb{N}}$ of random variables $Y_t : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (see for instance Unakafov [82] and the references given therein). Mostly, it is assumed that the *random process* $(Y_t)_{t\in\mathbb{N}}$ is *stationary*, i.e. the distribution of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_d})$ and $(Y_{t_1+\tau}, Y_{t_2+\tau}, \ldots, Y_{t_d+\tau})$ coincide for all $\tau, t_1, t_2, \ldots, t_d \in \mathbb{N}_0$.

Note that each measuring process on a measure-preserving dynamical system induces a stationary random process on the probability space $(\Omega, \mathcal{A}, \mu)$ and each stationary random process $(Y_t)_{t\in\mathbb{N}}$ with random variables $Y_t : (\Omega, \mathcal{A}, \mu) \to (A, \mathcal{B}(A))$ and at most countable $A \subset \mathbb{R}$ induces a measure-preserving dynamical system. In order to see this, let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system and X be an observable. Then $(Y_t)_{t\in\mathbb{N}}$ with $Y_t = X \circ T^{\circ(t-1)}$ is a real-valued stationary random process on $(\Omega, \mathcal{A}, \mu)$. On the contrary, let $(Y_t)_{t\in\mathbb{N}}$ be a stationary random process on $(\Omega, \mathcal{A}, \mu)$ with values in $A^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$, where A is at most countable. Then $(Y_t)_{t\in\mathbb{N}}$ induces the measure-preserving dynamical system $\left(A^{\mathbb{N}}, \mathcal{B}\left(A^{\mathbb{N}}\right), \mu_{(Y_t)_{t\in\mathbb{N}}}, \sigma\right)$. The map $\sigma : A^{\mathbb{N}} \hookleftarrow$, defined by

$$(\sigma r)_{t-1} = r_t$$

for all $t \in \mathbb{N}$ and $r = (r_1, r_2, \ldots) \in A^{\mathbb{N}}$, is called *shift-map*, $\mathcal{B}\left(A^{\mathbb{N}}\right)$ is the sigma-algebra generated by the family of *cylinder sets* $Z_t(a_1, a_2, \ldots, a_t)$ with $t \in \mathbb{N}$ and $a_1, a_2, \ldots, a_t \in A$ that are given by

$$Z_t(a_1, a_2, \ldots, a_t) := \left\{ r = (r_1, r_2, \ldots) \in A^{\mathbb{N}} \,\middle|\, r_1 = a_1, r_2 = a_2, \ldots, r_t = a_t \right\}.$$

Moreover, $\mu_{(Y_t)_{t\in\mathbb{N}}}$ is the probability measure defined on the cylinder sets by

$$\mu_{(Y_t)_{t\in\mathbb{N}}}(Z_t(a_1, a_2, \ldots, a_t)) := \mu\left(\{\omega \in \Omega \mid Y_t(\omega) = a_t, Y_{t-1}(\omega) = a_{t-1}, \ldots, Y_1(\omega) = a_1\}\right). ★$$

2.1.2 Symbolic schemes in the framework

The opening chapter of this thesis motivates the study of time-dependent, non-linear systems by encoding outcomes of measurements by sequences of symbols following a certain symbolic scheme. Therefore, a symbolic scheme is nothing else than a rule to classify measured values (see Section 3.2). Further, applying this rule to a dynamical system $(\Omega, \mathcal{A}, \mu, T)$ in a measurable way gives a finite partition of Ω. A *finite partition* $\mathcal{C} \subset \mathcal{A}$ of Ω is a set

$$\mathcal{C} := \left\{ C^{(1)}, C^{(2)}, \ldots, C^{(q)} \right\} \subset \mathcal{A}$$

with $q \in \mathbb{N}$ elements, where $\bigcup_{l=1}^{q} C^{(l)} = \Omega$ and $C^{(l)} \cap C^{(k)} = \emptyset$ for any $l, k \in \{1, 2, \ldots, q\}$ and $l \neq k$. That is why we say that a symbolic scheme entails a finite partition in the following. Usually, a symbolic scheme entails not only one partition but a sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions

$$\mathcal{C}_r = \left\{ C_r^{(1)}, C_r^{(2)}, \ldots, C_r^{(|\mathcal{C}_r|)} \right\} \subset \mathcal{A};\ r \in \mathbb{N},$$

i.e., in each step $r \in \mathbb{N}$ of the symbolic analysis, more information of the underlying system is accessed, however, the nature of symbolic encoding is not changed.

★ **Remark 2.4** (Symbolic dynamics)**.** We are mainly interested in the properties of partitions that arise from different symbolic schemes. Note that the actual idea of symbolic dynamics is to pick an arbitrary initial partition $\mathcal{C} \subset \mathcal{A}$ of Ω and to study the dynamical evolution of $(\Omega, \mathcal{A}, \mu, T)$ with respect to this partition. For the theory of symbolic dynamics, we refer the reader to Choe [23] and to Lind and Marcus [64]. ★

Note that we are only interested in the partitions and not in the symbols in the following. In particular, we have a special interest in the information content of a symbolic-based analysis technique and whether the technique preserves the information of the underlying system. In the following, let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be a sequence of finite partitions of Ω entailed by a symbolic-based analysis technique or rather by the underlying symbolic scheme. We say that the *information content* of the technique is based on the measure μ on the sub-σ-algebra $\sigma((\mathcal{C}_r)_{r\in\mathbb{N}})$ of $\mathcal{A}$. The sub-σ-algebra $\sigma((\mathcal{C}_r)_{r\in\mathbb{N}})$ of $\mathcal{A}$ is given by

$$\sigma((\mathcal{C}_r)_{r\in\mathbb{N}}) := \bigvee_{r\in\mathbb{N}} \sigma(\mathcal{C}_r) = \bigvee_{r\in\mathbb{N}} \left\{ \bigcup_{l\in L} C_r^{(l)} \,\middle|\, L \subset \{1, 2, \ldots, |\mathcal{C}_r|\} \right\}, \tag{2.2}$$

i.e. $\sigma((\mathcal{C}_r)_{r\in\mathbb{N}})$ is the minimal sub-σ-algebra containing all $\sigma(\mathcal{C}_r)$ (see Appendix B.5). Moreover, we say that the symbolic-based technique *preserves the information of the underlying system* if $(\mathcal{C}_r)_{r\in\mathbb{N}}$ has the following property.

Definition 2.4. Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be a sequence of finite partitions. If $\sigma((\mathcal{C}_r)_{r\in\mathbb{N}}) \overset{\mu}{\supset} \mathcal{A}$ (see Definition 2.1), then we call $(\mathcal{C}_r)_{r\in\mathbb{N}}$ *generating*.

Note that each finite partition $\mathcal{C}_r$ of $(\mathcal{C}_r)_{r\in\mathbb{N}}$ gives again a coarse-grained observation of $(\Omega, \mathcal{A}, \mu, T)$ by the random variable $Y_r : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that assigns to all elements of $C_r^{(l)}$ the value $l \in \{1, 2, \ldots, q\}$. In other words,

$$Y_r(\omega) := \sum_{l=1}^{|\mathcal{C}_r|} l \, \mathbf{1}_{C_r^{(l)}}(\omega)$$

for all $\omega \in \Omega$, where for each $l \in \{1, 2, \ldots, \mathcal{C}_r\}$

$$\mathbf{1}_{C_r^{(l)}}(\omega) = \begin{cases} 1 & \text{if } \omega \in C_r^{(l)}, \\ 0 & \text{otherwise} \end{cases}$$

denotes the characteristic function. We call the sequence $(Y_r)_{r\in\mathbb{N}}$ a *sequence of coarse-grained observations* with respect to $(\mathcal{C}_r)_{r\in\mathbb{N}}$, and the orbit $\big(Y_r(T^{\circ t}(\omega))\big)_{t\in\mathbb{N}_0}$ with $\omega \in \Omega$ a *symbolic path* with respect to $\mathcal{C}_r$. In fact, our whole discussion is concerned with the information content as well as the complexity of possible symbolic paths with respect to symbolic schemes. However, before we start, we summarize some basic concepts and properties of finite partitions.

We are interested, in particular, in such sequences $(\mathcal{C}_r)_{r\in\mathbb{N}}$ that are increasing, meaning that the finite partition $\mathcal{C}_{r+1} \subset \mathcal{A}$ is finer than the finite partition $\mathcal{C}_r \subset \mathcal{A}$ for all $r \in \mathbb{N}$. The following definition, for instance, can be found in Walters [87, Definition 4.2].

Definition 2.5. Let p and q be two natural numbers. A partition

$$\mathcal{D} := \left\{D^{(1)}, D^{(2)}, \ldots, D^{(p)}\right\} \subset \mathcal{A}$$

is *finer* than a partition

$$\mathcal{C} := \left\{C^{(1)}, C^{(2)}, \ldots, C^{(q)}\right\} \subset \mathcal{A}$$

or, equivalently, $\mathcal{C}$ is *coarser* than $\mathcal{D}$ if for all $l \in \{1, 2, \ldots, q\}$, there exists a non-empty set $K \subset \{1, 2, \ldots, p\}$ such that

$$C^{(l)} = \bigcup_{k\in K} D^{(k)}.$$

If a partition $\mathcal{D} \subset \mathcal{A}$ is finer than a partition $\mathcal{C} \subset \mathcal{A}$, we say that $\mathcal{D}$ is *refining* $\mathcal{C}$, and denote it by $\mathcal{C} \prec \mathcal{D}$. Note that the relation $\prec$ is a partial order on the set of finite partitions of $(\Omega, \mathcal{A})$, i.e. $\prec$ is reflexive, antisymmetric and transitive. If we consider an *increasing* sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions, then there exists for all $l \in \{1, 2, \ldots, |\mathcal{C}_{r-1}|\}$ a non-empty set $K \subset \{1, 2, \ldots, |\mathcal{C}_r|\}$ such that

$$C_{r-1}^{(l)} = \bigcup_{k\in K} C_r^{(k)}.$$

Moreover, we consider the *join* $\bigvee_{r=1}^{m} \mathcal{C}_r$ of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ with $m \in \mathbb{N}$ and $r \in \{1, 2, \ldots, m\}$ which is defined by

$$\bigvee_{r=1}^{m} \mathcal{C}_r := \left\{ \bigcap_{r=1}^{m} C_r^{(l_r)} \neq \emptyset \,\middle|\, l_r \in \{1, 2, \ldots, |\mathcal{C}_r|\} \text{ for } r \in \{1, 2, \ldots, m\} \right\}.$$

Hence the join is the coarsest partition refining all $\mathcal{C}_r$; $r \in \{1, 2, \ldots, m\}$. If $m = 2$, we also write $\mathcal{C}_1 \vee \mathcal{C}_2$ instead of $\bigvee_{r=1}^{2} \mathcal{C}_r$. For a fuller treatment, in particular, if m approaches infinity, we refer the reader to Itô [45, Section 3.3].

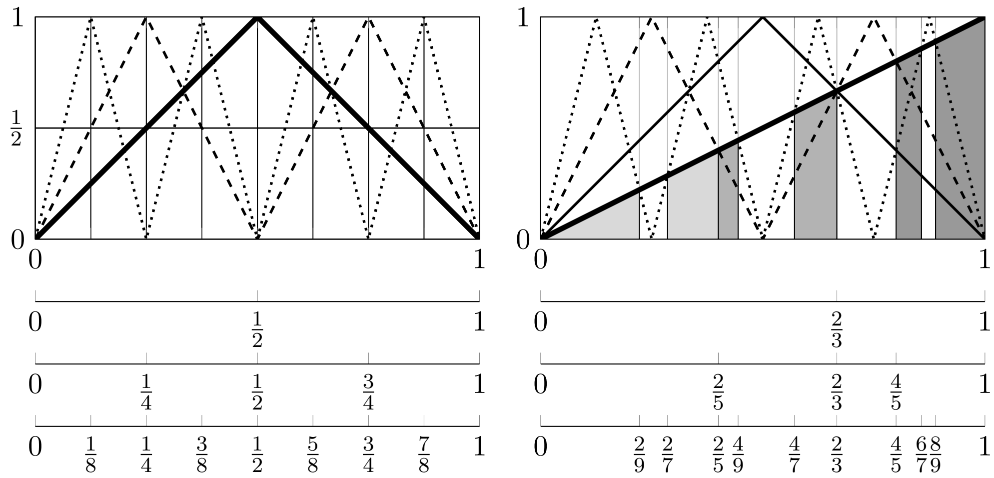


Figure 2.3: Graph of the full tent map T and the iterations $T^{\circ 2}$ and $T^{\circ 3}$. Left: refining partitions entailed by a threshold crossing method. Right: refining partitions entailed by an ordinal approach (see Example 2.4 for more details). This figure is a derivative of Stolz and Keller [79, Figure 4].

♣ **Example 2.4** (Increasing sequences of finite partitions)**.** Let Ω be the unit interval $[0, 1)$, and consider the transformation $T : [0, 1) \hookleftarrow$ defined by

$$T(\omega) = \begin{cases} 2\omega & \text{if } 0 \leq \omega < \frac{1}{2}, \\ 2 - 2\omega & \text{if } \frac{1}{2} \leq \omega < 1. \end{cases}$$

The map T is called the *full tent map* on $[0, 1)$, and is ergodic with respect to the Lebesgue measure λ (see for instance Chan and Tong [22]). Moreover, we have that

$$T^{\circ t}(\omega) = \begin{cases} 2^t\omega - 2l & \text{if } \frac{2l}{2^t} \leq \omega < \frac{2l+1}{2^t}, \\ 2(l+1) - 2^t\omega & \text{if } \frac{2l+1}{2^t} \leq \omega < \frac{2l+2}{2^t} \end{cases}$$

with $l \in \{0, 1, \ldots, 2^{t-1} - 1\}$.

In Figure 2.3, we display two different symbolic schemes that entail increasing sequences of finite partitions. Firstly, we apply a threshold crossing method with threshold $\frac{1}{2}$, and secondly, we apply an ordinal approach where we relate a state ω to the iterates $T(\omega)$, $T^{\circ 2}(\omega)$ and $T^{\circ 3}(\omega)$ (compare also to Section 2.2.1 and Section 3.2). In the case of the threshold crossing method, we have

$$\mathcal{C}_1 = \left\{\left\{\omega \in \Omega \mid \tfrac{1}{2} \leq \omega\right\}, \left\{\omega \in \Omega \mid \tfrac{1}{2} > \omega\right\}\right\},$$

$$\mathcal{C}_2 = \mathcal{C}_1 \vee \left\{\left\{\omega \in \Omega \mid \tfrac{1}{2} \leq T(\omega)\right\}, \left\{\omega \in \Omega \mid \tfrac{1}{2} > T(\omega)\right\}\right\}$$

and

$$\mathcal{C}_3 = \mathcal{C}_2 \vee \left\{\left\{\omega \in \Omega \mid \tfrac{1}{2} \leq T^{\circ 2}(\omega)\right\}, \left\{\omega \in \Omega \mid \tfrac{1}{2} > T^{\circ 2}(\omega)\right\}\right\}.$$

In the case of the ordinal approach, we obtain the following finite partitions

$$\mathcal{D}_1 = \{\{\omega \in \Omega \mid T(\omega) \leq \omega\}, \{\omega \in \Omega \mid T(\omega) > \omega\}\},$$

$$\mathcal{D}_2 = \mathcal{D}_1 \vee \left\{ \left\{ \omega \in \Omega \mid T^{\circ 2}(\omega) \leq \omega \right\}, \left\{ \omega \in \Omega \mid T^{\circ 2}(\omega) > \omega \right\} \right\}$$

and

$$\mathcal{D}_3 = \mathcal{D}_2 \vee \left\{ \left\{ \omega \in \Omega \mid T^{\circ 3}(\omega) \leq \omega \right\}, \left\{ \omega \in \Omega \mid T^{\circ 3}(\omega) > \omega \right\} \right\}.$$

Note that some elements of the partition $\mathcal{D}_3$ are unions of intervals, this is indicated by different grayscale values whereas white represents non-united intervals that are elements of $\mathcal{D}_3$. ♣

2.2 The Kolmogorov-Sinai entropy

What is complexity? In various natural and social fields, in particular, if time-dependent systems are analyzed, one meets the challenge to define and to measure complexity (see for instance Adami [1] and Sprott [77, Section 15.7]). For a deeper discussion of the previous question, we refer the reader to Ladyman et al. [61] and the references given therein.

In this thesis, we utilize the characterization given from information theory, and say a system is the more complex the more unpredictable its dynamics are or, in other words, the less we know about the underlying dynamics before we perform an experiment. Having symbolic-based analysis techniques at hand, that is nothing else than determining the complexity of possible symbolic paths or rather the complexity which is independent of any such discretization. In fact, the latter one is the traditional and well-defined complexity measure called Kolmogorov-Sinai entropy (KS entropy for short). Recall, that we assume that $(\Omega, \mathcal{A}, \mu, T)$ is a measure-preserving dynamical system.

It is not an easy task to determine the KS entropy since usually an uncountable set of finite partitions has to be considered (see Section 2.2.1). However, some special partitions yield properties and concepts which facilitate the determination (see Section 2.2.2).

2.2.1 Definition of the Kolmogorov-Sinai entropy

Let $\mathcal{C} = \left\{ C^{(1)}, C^{(2)}, \ldots, C^{(q)} \right\} \subset \mathcal{A}$ be a finite partition entailed by a symbolic scheme. In order to determine the complexity of possible symbolic paths with respect to the considered symbolic scheme, one assigns to every part of $\mathcal{C}$ a letter of the alphabet $A = \{1, 2, \ldots, q\}$, and classifies, stepwise, the states $\omega \in \Omega$ with respect to their itinerary (compare to Figure 2.4). In other words, one determines for each word $(a_1, a_2, \ldots, a_t)$ of length $t \in \mathbb{N}$ the sets

$$C^{(a_1, a_2, \ldots, a_t)} := \left\{ \omega \in \Omega \,\middle|\, \left(\omega, T(\omega), \ldots, T^{\circ t-1}(\omega)\right) \in C^{(a_1)} \times C^{(a_2)} \times \ldots \times C^{(a_t)} \right\}. \quad (2.3)$$

All non-empty sets $C^{(a_1, a_2, \ldots, a_t)}$ provide a finite partition $(\mathcal{C})_t \subset \mathcal{A}$ of Ω. The finite partition $(\mathcal{C})_t$ describes the dynamical behavior of $(\Omega, \mathcal{A}, \mu, T)$ up to a point $t \in \mathbb{N}$ with respect to an *initial partition* $\mathcal{C}$. In particular, let $(\mathcal{C})_1 = \mathcal{C}$. We use the notation $(\mathcal{C})_t$ to emphasize that the partition is constructed with respect to T, and call the sequence $((\mathcal{C})_t)_{t\in\mathbb{N}}$ the *partition sequence of* $\mathcal{C}$ *under* T (see Example 2.5).

The complexity of possible symbolic paths with respect to $\mathcal{C}$ is now given by the *entropy rate* $h_\mu(T,\mathcal{C})$ defined by

$$h_\mu(T,\mathcal{C}) = \lim_{t\to\infty} \frac{1}{t} H_\mu((\mathcal{C})_t), \tag{2.4}$$

where $H_\mu((\mathcal{C})_t)$ is the Shannon entropy of $(\mathcal{C})_t$. The Shannon entropy of a finite partition $\mathcal{C}$ is defined by

$$H_\mu(\mathcal{C}) := -\sum_{l=1}^{q} \mu\left(C^{(l)}\right) \ln\left(\mu\left(C^{(l)}\right)\right) \tag{2.5}$$

(with $0\ln(0) := 0$), and measures, in our context, the complexity of the considered symbolic scheme without considering the dynamics. Here, we use the natural logarithm, however, taking the logarithm to other basis is also possible (see Amigó [5, Annex B]). In computer science, for instance, mostly the base two is chosen in order to measure complexity in binary digits. Some more notes on the Shannon entropy are given in Remark 2.5. However, for a fuller treatment of the topic, for instance, that the limit in Equation (2.4) exists and $\frac{1}{t}H_\mu((\mathcal{C})_t)$ decreases to $h_\mu(T,\mathcal{C})$, we refer the reader to Walters [87, Chapter 4]. Note that the limit in Equation (2.4) can also be infinite.

★ **Remark 2.5** (Notes on the Shannon entropy and the entropy rate)**.** The more we know about the outcomes of an experiment beforehand the less the complexity of a symbolic scheme (without considering the dynamics) should be. This agrees with the definition of the Shannon entropy. In order to see this, let us assume that the outcomes of an experiment are represented by $\mathcal{C}$ and that the probability that $C^{(l)}$ with $l \in \{1,2,\dots,q\}$ occurs is given by $\mu\left(C^{(l)}\right)$.

In fact, the expression in (2.5) is maximal if $\mu\left(C^{(l)}\right) = \frac{1}{q}$ for all $l \in \{1,2,\dots,q\}$, and minimal if for one $l \in \{1,2,\dots,q\}$ the element $C^{(l)}$ has measure one (see Walters [87]). Hence the Shannon entropy is bounded, i.e. $0 \leq H_\mu(\mathcal{C}) \leq \ln(q)$.

Moreover, for two finite partitions $\mathcal{C} \subset \mathcal{A}$ and $\mathcal{D} \subset \mathcal{A}$ of Ω, we have that $\mathcal{C} \prec \mathcal{D}$ if and only if $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$. Further, if $\mathcal{C} \prec \mathcal{D}$, then $H_\mu(\mathcal{C}) \leq H_\mu(\mathcal{D})$ and $(\mathcal{C})_t \prec (\mathcal{D})_t$ for all $t \in \mathbb{N}$, and thus $h_\mu(T,\mathcal{C}) \leq h_\mu(T,\mathcal{D})$ (see for instance Walters [87, Section 4.3 and 4.4]). ★

In order to make a statement about the complexity of the considered time-dependent system independently of any discretization by a symbolic scheme, one determines the KS entropy, i.e.

$$h_\mu^{\mathrm{KS}}(T) := \sup_{\mathcal{C}\text{ finite partition}} h_\mu(T,\mathcal{C}).$$

We close this section with some remarks on the partition sequence of a partition $\mathcal{C}$ under T. First of all, note that $((\mathcal{C})_t)_{t\in\mathbb{N}}$ is increasing by construction (compare to Equation (2.3) and Example 2.5).

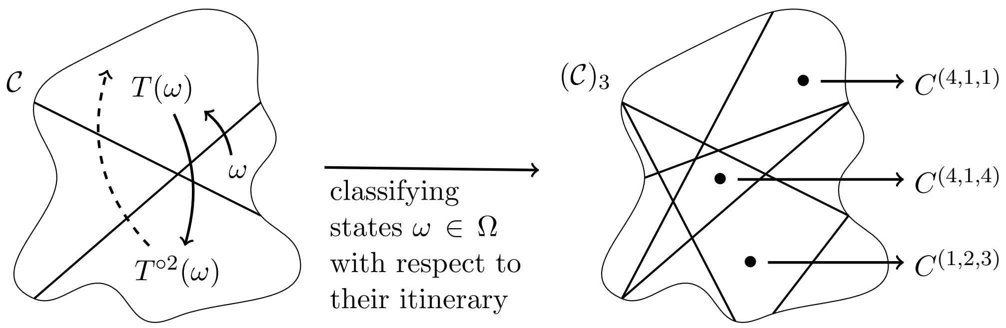


Figure 2.4: Process underlying the determination of the entropy rate $h_\mu(T, \mathcal{C})$ for $t = 3$ and a finite partition $\mathcal{C}$ with four elements (see Section 2.2.1). This figure is published in Stolz and Keller [79].

Definition 2.6. Let $\mathcal{C} \subset \mathcal{A}$ be a finite partition of Ω. If the sequence $((\mathcal{C})_t)_{t\in\mathbb{N}}$ is generating (see Definition 2.4), then we call $\mathcal{C}$ *generating under* T.

Let $\mathcal{C} \subset \mathcal{A}$ and $\mathcal{D} \subset \mathcal{A}$ be two finite partitions of Ω. If $\mathcal{C} \prec \mathcal{D}$ and $\mathcal{C}$ is generating under T, then $\mathcal{D}$ is generating under T since $(\mathcal{C})_t \prec (\mathcal{D})_t$ for all $t \in \mathbb{N}$, and thus $\sigma(((\mathcal{C})_t)_{t\in\mathbb{N}}) \subset \sigma(((\mathcal{D})_t)_{t\in\mathbb{N}})$ (see Remark 2.5).

♣ **Example 2.5** (Partition sequence of a finite partition $\mathcal{C}$ under the full tent map)**.** Let Ω be the unit interval $[0, 1)$, λ the Lebesgue measure, and $T : [0, 1) \hookleftarrow$ be the full tent map (see Example 2.4). The partition sequence of

$$\mathcal{C} := \left\{ \left[0, \tfrac{1}{2}\right), \left[\tfrac{1}{2}, 1\right) \right\}$$

under T is given by

$$((\mathcal{C})_t)_{t\in\mathbb{N}} = \left(\left\{ \left[\frac{l}{2^t}, \frac{l+1}{2^t} \right) \,\middle|\, l \in \{0, 1, \ldots, 2^t - 1\} \right\} \right)_{t\in\mathbb{N}}.$$

Thus each $(\mathcal{C})_t$ consists of 2^t dyadic intervals of $[0, 1)$, and

$$H_\lambda((\mathcal{C})_t) = t \ln(2)$$

(see Remark 2.5, Figure 3.2 and Example 3.1, and compare also to Bollt et al. [15]).

♣

The determination of the entropy rate, and subsequently, of the KS entropy can be simplified (see Section 2.2.2) if the symbolic scheme entails a finite partition with the following property.

Definition 2.7. Let $\mathcal{C} \subset \mathcal{A}$ be a partition of Ω with $q \in \mathbb{N}$ elements and $((\mathcal{C})_t)_{t\in\mathbb{N}}$ the partition sequence of $\mathcal{C}$ under T. If for each word $(a_1, a_2, \ldots, a_t) \in \{1, 2, \ldots, q\}^t$

with $t, s \in \mathbb{N}$, $t > s$ and $\mu\left(C^{(a_1,a_2,\ldots,a_{t-1})}\right) > 0$ we have that

$$\frac{\mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)}{\mu\left(C^{(a_1,a_2,\ldots,a_{t-1})}\right)} = \frac{\mu\left(C^{(a_{t-s},a_{t-(s-1)},\ldots,a_t)}\right)}{\mu\left(C^{(a_{t-s},a_{t-(s-1)},\ldots,a_{t-1})}\right)},$$

then we say that $\mathcal{C}$ has the *Markov property of order s with respect to μ.*

Note that a partition satisfies the Markov property of order $s \in \mathbb{N}$ if the probability of $\left\{\omega \in \Omega \mid T^{\circ t-1}(\omega) \in C^{(a_t)}\right\}$ with $t \in \mathbb{N}$, $t > s$ and $a_t \in \{1, 2, \ldots, q\}$ does not depend on the visited elements up until $T^{\circ(t-s)}$ (see Parry and Williams [70] and Unakafov [82]).

♣ **Example 2.6** (Partitions with the Markov property of order $s \in \mathbb{N}$)**.** There are many examples of partitions with the Markov property of order one. For instance, consider $([0, 1), \mathcal{B}([0, 1)), \lambda, T)$, where T is the full tent map on $[0, 1)$, and choose the initial partition

$$\mathcal{C} := \left\{\left[0, \tfrac{1}{2}\right), \left[\tfrac{1}{2}, 1\right)\right\}.$$

Since $\lambda\left(\left[\frac{l}{2^t}, \frac{l+1}{2^t}\right)\right) = \frac{1}{2^t}$, we have

$$\frac{\lambda\left(C^{(a_1,a_2,\ldots,a_t)}\right)}{\lambda\left(C^{(a_1,a_2,\ldots,a_{t-1})}\right)} = \frac{2^{t-1}}{2^t} = \frac{2}{2^2} = \frac{\lambda\left(C^{(a_{t-1},a_t)}\right)}{\lambda\left(C^{(a_{t-1})}\right)}$$

for all $t > 1$ (see Example 2.5). Moreover, $\mathcal{C}$ is generating under T (see Billingsley [13, A31. *Dyadic expansions.*]). Further examples of partitions with the Markov property of order one are given in Unakafov [82, Section 2.1.3].

An example of some arbitrary order is designed in the following way: Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and $(Y_t)_{t\in\mathbb{N}}$ be a stationary random process with values in $A^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$, where A is at most countable (see Remark 2.3). Moreover, let $(Y_t)_{t\in\mathbb{N}}$ be a *Markov chain* of order $s \in \mathbb{N}$, i.e. for each natural number $t > s$

$$\begin{aligned}\mu\left(Y_t = a_t \mid Y_{t-1}\right. &= \left.a_{t-1}, Y_{t-2} = a_{t-2}, \ldots, Y_1 = a_1\right) \\ &= \mu\left(Y_t = a_t \mid Y_{t-1} = a_{t-1}, Y_{t-2} = a_{t-2}, \ldots, Y_{t-s} = a_{t-s}\right).\end{aligned} \tag{2.6}$$

Then $(Y_t)_{t\in\mathbb{N}}$ induces a dynamical system $\left(A^{\mathbb{N}}, \mathcal{B}\left(A^{\mathbb{N}}\right), \mu_{(Y_t)_{t\in\mathbb{N}}}, \sigma\right)$ that is measure-preserving (see Remark 2.3). Moreover, it holds

$$\begin{aligned}&\mu_{(Y_t)_{t\in\mathbb{N}}}(Z_t(a_1, a_2, \ldots, a_t)) \\ &\quad= \mu\left(\{\omega \in \Omega \mid Y_t(\omega) = a_t, Y_{t-1}(\omega) = a_{t-1}, \ldots, Y_1(\omega) = a_1\}\right) \\ &\quad= \mu_{(Y_t)_{t\in\mathbb{N}}}(Z_t(a_1, a_2, \ldots, a_{t-1})) \\ &\qquad\qquad \mu\left(Y_t = a_t \mid Y_{t-1} = a_{t-1}, Y_{t-2} = a_{t-2}, \ldots, Y_1 = a_1\right)\end{aligned}$$

and

$$\begin{aligned}\mu\,(\;Y_t &= a_t \mid Y_{t-1} = a_{t-1}, Y_{t-2} = a_{t-2}, \ldots, Y_1 = a_1) \\ &\overset{(2.6)}{=} \mu\left(Y_t = a_t \mid Y_{t-1} = a_{t-1}, Y_{t-2} = a_{t-2}, \ldots, Y_{t-s} = a_{t-s}\right) \\ &= \frac{\mu\left(\{\omega \in \Omega \mid Y_t(\omega) = a_t, Y_{t-1}(\omega) = a_{t-1}, \ldots, Y_{t-s}(\omega) = a_{t-s}\}\right)}{\mu\left(\{\omega \in \Omega \mid Y_{t-1}(\omega) = a_{t-1}, Y_{t-2}(\omega) = a_{t-2}, \ldots, Y_{t-s}(\omega) = a_{t-s}\}\right)}\end{aligned}$$

for each $(a_1, a_2, \ldots, a_t) \in \{1, 2, \ldots, q\}^t$ with $t > s$ and $\mu_{(Y_t)_{t\in\mathbb{N}}}(Z_t(a_1, a_2, \ldots, a_t)) > 0$. Thus, for some $t > s$, the cylinder sets $Z_t(a_1, a_2, \ldots, a_t)$ correspond to a partition with the Markov property of order $s \in \mathbb{N}$. ♣

2.2.2 Determining the Kolmogorov-Sinai entropy

As already stated, the determination of the KS entropy is not easy. However, in the rare case, that one can find a finite partition $\mathcal{C}$ which is generating under T, one can characterize the KS entropy directly (see Walters [87, Theorem 4.10]):

Lemma 2.1. *Let $\mathcal{C} \subset \mathcal{A}$ be a finite partition of Ω which is generating under T, then*

$$h_\mu^{\mathrm{KS}}(T) = h_\mu(T, \mathcal{C}).$$

The determination is even more facilitated if the initial partition $\mathcal{C}$ is generating under T and has the Markov property of some order $s \in \mathbb{N}$. Before we write down the relevant statement, note that the entropy rate can be determined alternatively to (2.4) by

$$h_\mu(T, \mathcal{C}) = \lim_{t\to\infty} \Big(H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t) \Big). \tag{2.7}$$

In Remark 2.6, we show, using our notation, that (2.4) and (2.7) are equivalent for the sake of completeness.

Lemma 2.2. *Let $\mathcal{C} \subset \mathcal{A}$ be a partition of Ω with $q \in \mathbb{N}$ elements and $((\mathcal{C})_t)_{t\in\mathbb{N}}$ the partition sequence of $\mathcal{C}$ under T. If $\mathcal{C}$ is generating under T and has the Markov property of order $s \in \mathbb{N}$, then*

$$h_\mu^{\mathrm{KS}}(T) = H((\mathcal{C})_{s+1}) - H((\mathcal{C})_s).$$

The previous theorem is a generalization of a statement given in Unakafov [82, Theorem 2.1], which says that $h_\mu^{\mathrm{KS}}(T) = H((\mathcal{C})_2) - H((\mathcal{C})_1)$ if $\mathcal{C}$ is generating under T and has the Markov property of order one (compare also to Kitchens [57, Observation 6.2.10]). We give the proof at the end of this section for illustrative purposes in our notation. Moreover, in order to reinforce the legibility, we switch to the landscape format.

♣ **Example 2.7** (Concluding remarks to the full tent map)**.** Consider the dynamical system $([0,1), \mathcal{B}([0,1)), \lambda, T)$, where T is the full tent map on $[0,1)$, and choose the initial partition

$$\mathcal{C} := \left\{ \left[0, \tfrac{1}{2}\right), \left[\tfrac{1}{2}, 1\right) \right\}.$$

By the results stated in Example 2.6, we have that

$$h_\mu^{\mathrm{KS}}(T) = H((\mathcal{C})_2) - H((\mathcal{C})_1) = 2\ln(2) - \ln(2) = \ln(2). \qquad ♣$$

As already stated, generating partitions generally do not exist or are not accessible (see Chapter 3, in particular, Section 3.2.1). The following lemma provides a route to the KS entropy via an arbitrary increasing sequence of finite partitions (see Walters [87, Theorem 4.14], and Chapter 3 for examples).

Lemma 2.3. *Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be an increasing sequence of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ of Ω with respect to $\prec$ and generating, then*

$$h_\mu^{\mathrm{KS}}(T) = \lim_{r\to\infty} h_\mu(T, \mathcal{C}_r) = \sup_{r\in\mathbb{N}} h_\mu(T, \mathcal{C}_r).$$

Thus, in general, a double-limit has to be evaluated (see Figure 2.5), where for a fixed $r \in \mathbb{N}$ the sequence $\Big(H_\mu((\mathcal{C}_r)_{t+1}) - H_\mu((\mathcal{C}_r)_t)\Big)_{t\in\mathbb{N}}$ is monotonically decreasing to the entropy rate $h_\mu(T, \mathcal{C}_r)$, and the sequence $(h_\mu(T, \mathcal{C}_r))_{r\in\mathbb{N}}$ is increasing since $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is refining.

As a direct consequence of the last results, we obtain the following proposition (compare also to Unakafov [82, Lemma 3.8] and Figure 2.5).

Proposition 2.2. *Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be an increasing sequence of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ of Ω which is generating. Moreover, let $((\mathcal{C}_r)_t)_{t\in\mathbb{N}}$ be the corresponding partition sequence of each $\mathcal{C}_r$ under T.*

(1) If $\mathcal{C}_r$ has the Markov property of order $s \in \mathbb{N}$ for all $r \geq m$, then

$$h_\mu^{\mathrm{KS}}(T) = \lim_{r\to\infty} \Big(H((\mathcal{C}_r)_{s+1}) - H((\mathcal{C}_r)_s)\Big).$$

(2) If $\mathcal{C}_r$ is generating under T and has the Markov property of order $s \in \mathbb{N}$ for some $r \in \mathbb{N}$, then

$$h_\mu^{\mathrm{KS}}(T) = H((\mathcal{C}_r)_{s+1}) - H((\mathcal{C}_r)_s).$$

The previous proposition requires again good knowledge about the properties of the considered partitions. In general, it is difficult to state whether these partitions are generating or have the Markov property. In Unakafov [82, Lemma 3.9], examples for the idea of Bandt and Pompe [11] are given, and conditions are named that have to be fulfilled such that the corresponding partitions are generating and have the Markov property of order one.

★ **Remark 2.6** (Notes on the conditional entropy)**.** Let $\mathcal{C}$ be a partition with $q \in \mathbb{N}$ elements and $((\mathcal{C})_t)_{t\in\mathbb{N}}$ be the partition sequence of $\mathcal{C}$ under T. Recall, that $((\mathcal{C})_t)_{t\in\mathbb{N}}$ is an increasing sequence with respect to $\prec$. In the following, we show that

$$\lim_{t\to\infty} \Big(H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t)\Big) = \lim_{t\to\infty} \frac{1}{t} H_\mu((\mathcal{C})_t). \tag{2.8}$$

The following, for instance, can be found in Cover and Thomas [25] and in Amigó [5]. We repeat it, using our notation, for the sake of completeness and in preparation for the proof of Lemma 2.2. In order to show (2.8), we need the following definition (see for instance Walters [87, Chapter 4]). Let p and q be two natural numbers, and consider the two finite partitions

$$\mathcal{C} := \left\{C^{(1)}, C^{(2)}, \ldots, C^{(q)}\right\} \subset \mathcal{A} \text{ and } \mathcal{D} := \left\{D^{(1)}, D^{(2)}, \ldots, D^{(p)}\right\} \subset \mathcal{A}$$

of Ω with $\mu(C^{(l)}) > 0$ for each $l \in \{1, 2, \ldots, q\}$. The *conditional entropy of $\mathcal{D}$ given $\mathcal{C}$* is defined by

$$H_\mu(\mathcal{D}|\mathcal{C}) = -\sum_{k=1}^{p}\sum_{l=1}^{q} \mu\left(D^{(k)} \cap C^{(l)}\right) \ln\left(\frac{\mu(D^{(k)} \cap C^{(l)})}{\mu(C^{(l)})}\right).$$

We have that $H_\mu(\mathcal{D}|\mathcal{C}) = H_\mu(\mathcal{C} \vee \mathcal{D}) - H_\mu(C)$, and thus $H_\mu(\mathcal{D}|\mathcal{C}) = H_\mu(\mathcal{D}) - H_\mu(C)$ if $\mathcal{C} \prec \mathcal{D}$. Moreover, if $\mathcal{C}$ is refining a finite partition $\mathcal{C}^* \subset \mathcal{A}$ of Ω, then

$$H_\mu(\mathcal{D}|\mathcal{C}^*) \geq H_\mu(\mathcal{D}|\mathcal{C})$$

(see Walters [87, Theorem 4.3]).

Our proof of (2.8) starts with the observation that the sequence

$$\Big(H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t)\Big)_{t\in\mathbb{N}} \tag{2.9}$$

is a monotonically decreasing sequence of non-negative numbers, and therefore, has a limit. Note that the limit can also be infinite. In order to see that (2.9) is decreasing, let $(\mathcal{C})^*_{t+1}$ be the partition generated by all non-empty sets

$$C^{(a_2,a_3,\ldots,a_{t+1})} = \\ \Big\{\omega \in \Omega \,\Big|\, \Big(T(\omega), T^{\circ 2}(\omega), \ldots, T^{\circ t}(\omega)\Big) \in C^{(a_2)} \times C^{(a_3)} \times \ldots \times C^{(a_{t+1})}\Big\},$$

and $\mathcal{D}$ be the partition of all non-empty sets $C^{(a_{t+2})} = \{\omega \in \Omega \mid T^{\circ t+1}(\omega) \in C^{(a_{t+2})}\}$. Obviously, $(\mathcal{C})^*_{t+1} \prec (\mathcal{C})_{t+1}$, and by the T-invariance (see Definition 2.2) of μ, we have that

$$\mu(C^{(a_2,a_3,\ldots,a_{t+2})}) = \mu(C^{(a_1,a_2,\ldots,a_{t+1})}),$$

and thus $H_\mu(\mathcal{D}|(\mathcal{C})^*_{t+1}) = H_\mu((\mathcal{C})_{t+1}|(\mathcal{C})_t)$. It follows that

$$H_\mu((\mathcal{C})_{t+2}|(\mathcal{C})_{t+1}) = H_\mu(\mathcal{D}|(\mathcal{C})_{t+1}) \leq H_\mu(\mathcal{D}|(\mathcal{C})^*_{t+1}) = H_\mu((\mathcal{C})_{t+1}|(\mathcal{C})_t).$$

We next apply the Cesàro mean theorem (see for instance Cover and Thomas [25, Theorem 4.2.3]), which states that if a sequence $(x_t)_{t\in\mathbb{N}}$ converges to the limit x, then x is also the limit of $\frac{1}{t}\sum_{s=1}^{t} x_s$ as t approaches infinity. Hence, by defining $H_\mu((\mathcal{C})_0) := H_\mu(\Omega) = 0$ and setting $x_t := H_\mu((\mathcal{C})_t) - H_\mu((\mathcal{C})_{t-1})$, which yields that

$$\frac{1}{t}\sum_{s=1}^{t}(H_\mu((\mathcal{C})_s) - H_\mu((\mathcal{C})_{s-1})) = \frac{1}{t}H_\mu((\mathcal{C})_t),$$

we obtain the desired conclusion. Moreover,

$$\frac{1}{t}\sum_{s=1}^{t}(H_\mu((\mathcal{C})_{s+1}) - H_\mu((\mathcal{C})_s)) \geq H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t),$$

and hence

$$\liminf_{t\to\infty} \frac{1}{t}H_\mu((\mathcal{C})_t) \geq \liminf_{t\to\infty}\Big(H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t)\Big).$$ ★

Proof of Lemma 2.2 If $\mathcal{C}$ is generating under T and has the Markov property of order $s \in \mathbb{N}$, we claim, in Lemma 2.2, that

$$h_\mu^{\mathrm{KS}}(T) = H_\mu((\mathcal{C})_{s+1}) - H_\mu((\mathcal{C})_s).$$

Note that $(\mathcal{C})_s$ is the s-th element of $((\mathcal{C})_t)_{t\in\mathbb{N}}$. In the following, we take sums over words

$$(a_n, a_{n+1}, \dots, a_{n+m}) \in \{1, 2, \dots, q\}^{n+m}$$

with $n, m \in \mathbb{N}$, and write it $a_n, a_{n+1}, \dots, a_{n+m}$ for short. Moreover, we always consider elements of partitions with strictly positive measure, i.e. $\mu\left(C^{(a_1,a_2,\dots,a_{t-1})}\right) > 0$, thus any union of these elements carries a strictly positive measure, particularly, $\mu\left(C^{(a_i)}\right) > 0$ for any $i \in (1, 2, \dots, t)$. By the deliberations in Remark 2.6, we have

$$\begin{aligned} H_\mu((\mathcal{C})_{s+1}|(\mathcal{C})_s) &= H_\mu((\mathcal{C})_{s+1}) - H_\mu((\mathcal{C})_s) \\ &= -\sum_{a_1,a_2,\dots,a_{s+1}} \mu\left(C^{(a_1,a_2,\dots,a_{s+1})}\right) \ln\left(\frac{\mu\left(C^{(a_1,a_2,\dots,a_{s+1})}\right)}{\mu\left(C^{(a_1,a_2,\dots,a_s)}\right)}\right). \end{aligned}$$

The basic idea of the proof is to show that for any natural number $t > s$ the equality

$$H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t) = H_\mu((\mathcal{C})_{s+1}) - H_\mu((\mathcal{C})_s)$$

holds true. By (2.7), and since $\mathcal{C}$ is generating under T, it holds that

$$h_\mu^{\mathrm{KS}}(T) = h_\mu(T, \mathcal{C}) = \lim_{t\to\infty} \Big(H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t)\Big).$$

Before writing down $H_\mu((\mathcal{C})_{t+1})$, we begin with two observations. We utilize the T-invariance and the law of total probability since we mostly consider compound probabilities of finite partitions (see for instance Billingsley [13, Section 4]). Firstly,

$$\sum_{a_{t+1}} \frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)} = \frac{1}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)} \sum_{a_{t+1}} \mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right) = \frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)} = 1,$$

and secondly, we have

$$\begin{aligned}
&\sum_{a_1,a_2,\ldots,a_{t-s}} \mu\left(C^{(a_1,a_2,\ldots,a_{t-s},a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)\\
&= \sum_{a_1,a_2,\ldots,a_{t-s}} \mu\Big(\Big\{\omega\in\Omega \,\Big|\, \Big(\omega,T(\omega),\ldots,T^{\circ t-s-1}(\omega),T^{\circ t-s}(\omega),T^{\circ t-s+1}(\omega),\ldots,T^{\circ t-1}(\omega)\Big)\\
&\qquad\qquad \in C^{(a_1)}\times C^{(a_2)}\times\ldots\times C^{(a_{t-s})}\times C^{(a_{t-s+1})}\times C^{(a_{t-s+2})}\times\ldots\times C^{(a_t)}\Big\}\Big)\\
&= \mu\Big(\Big\{\omega\in\Omega \,\Big|\, \Big(T^{\circ t-s}(\omega),T^{\circ t-s+1}(\omega),\ldots,T^{\circ t-1}(\omega)\Big)\in C^{(a_{t-s+1})}\times C^{(a_{t-s+2})}\times\ldots\times C^{(a_t)}\Big\}\Big)\\
&\overset{T\text{-invariance}}{=} \mu\Big(\Big\{\omega\in\Omega \,\Big|\, \Big(\omega,T(\omega),\ldots,T^{\circ s-1}(\omega)\Big)\in C^{(a_{t-s+1})}\times C^{(a_{t-s+2})}\times\ldots\times C^{(a_t)}\Big\}\Big)\\
&= \mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right).
\end{aligned}$$

Now, by applying the Markov property of $\mathcal{C}$ (denoted by $(*)$), we obtain

$$\begin{aligned}
H_\mu((\mathcal{C})_{t+1}) &= -\sum_{a_1,a_2,\ldots,a_{t+1}} \mu\left(C^{(a_1,a_2,\ldots,a_{t+1})}\right)\ln\left(\mu\left(C^{(a_1,a_2,\ldots,a_{t+1})}\right)\right)\\
&\overset{(*)}{=} -\sum_{a_1,a_2,\ldots,a_{t+1}} \mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)}\ln\left(\mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)}\right)\\
&= -\sum_{a_1,a_2,\ldots,a_{t+1}} \mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)}\ln\left(\mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)\right)\\
&\quad -\sum_{a_1,a_2,\ldots,a_{t+1}} \mu\left(C^{(a_1,a_2,\ldots,a_t)}\right)\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)}\ln\left(\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\ldots,a_t)}\right)}\right).
\end{aligned}$$

By rearranging the sums and by the previous observations, we have

$$\begin{aligned}
H_\mu((\mathcal{C})_{t+1}) = &- \sum_{a_1,a_2,\dots,a_t} \mu\left(C^{(a_1,a_2,\dots,a_t)}\right) \ln\left(\mu\left(C^{(a_1,a_2,\dots,a_t)}\right)\right) \overbrace{\sum_{a_{t+1}} \frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)}}^{=1} \\
&- \sum_{a_{t-s+1},a_{t-s+2},\dots,a_{t+1}} \frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)} \ln\left(\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)}\right) \underbrace{\sum_{a_1,a_2,\dots,a_{t-s}} \mu\left(C^{(a_1,a_2,\dots,a_t)}\right)}_{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}.
\end{aligned}$$

Overall we obtain, by reindexing, the desired conclusion, i.e.

$$\begin{aligned}
H_\mu((\mathcal{C})_{t+1}) &= H_\mu((\mathcal{C})_t) - \sum_{a_{t-s+1},a_{t-s+2},\dots,a_{t+1}} \mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right) \ln\left(\frac{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_{t+1})}\right)}{\mu\left(C^{(a_{t-s+1},a_{t-s+2},\dots,a_t)}\right)}\right) \\
&= H_\mu((\mathcal{C})_t) \underbrace{- \sum_{b_1,b_2,\dots,b_{s+1}} \mu\left(C^{(b_1,b_2,\dots,b_{s+1})}\right) \ln\left(\frac{\mu\left(C^{(b_1,b_2,\dots,b_{s+1})}\right)}{\mu\left(C^{(b_1,b_2,\dots,b_s)}\right)}\right)}_{H_\mu(\mathcal{C}_{s+1})-H_\mu(\mathcal{C}_s)}. \qquad \square
\end{aligned}$$

2.2.3 Approximating the Kolmogorov-Sinai entropy

The results of Section 2.2 are summarized in Figure 2.5 as follows. Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be an increasing sequence with respect to $\prec$ and generating. Then, by Lemma 2.3, the KS entropy is the double-limit of $\frac{1}{t}H_\mu((\mathcal{C}_r)_t)$ (see Figure 2.5**(a)**) and $H_\mu((\mathcal{C}_r)_{t+1}|(\mathcal{C}_r)_t)$ (see Figure 2.5**(b)**), respectively, as r and the word length t approach infinity. The arrows in Figure 2.5 indicate the direction of convergence, i.e.

$$\frac{1}{t}H_\mu((\mathcal{C}_r)_t) \text{ and } H_\mu((\mathcal{C}_r)_{t+1}|(\mathcal{C}_r)_t)$$

decrease to the entropy rate $h_\mu(T,\mathcal{C}_r)$ for increasing $t\in\mathbb{N}$, and

$$(h_\mu(T,\mathcal{C}_r))_{r\in\mathbb{N}}$$

increases since $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is refining. Moreover, by Proposition 2.2, if $\mathcal{C}_r$ with $r\geq m$ has the Markov property of order $s\in\mathbb{N}$, the limit of t approaching infinity is obsolete for $t\geq s$ (pointed out by the dashed line; here $m=3$). If in addition $\mathcal{C}_m$ is generating under T, the KS entropy coincides with the conditional entropy of $(\mathcal{C}_m)_{s+1}$ given $(\mathcal{C}_m)_s$ (indicated by the box).

Dropping the double limit in Lemma 2.3 provides a wide range of theoretical entropy measures, for instance, it is common to work with

$$h_\mu^{(\mathcal{C}_r)_{r\in\mathbb{N}}} := \limsup_{r\to\infty}\frac{1}{\ln(|\mathcal{C}_r|)}H_\mu(\mathcal{C}_r)$$

where the prefactor is chosen in order to normalize the Shannon entropy (see Remark 2.5). In Chapter 4, we use the prefactor $\frac{1}{r}$ due to the results of Bandt and Pompe [11] and Bandt et al. [10] in ordinal dynamics. Moreover, we give concluding remarks on theoretical entropy measures directly derived from symbolic-based analysis techniques in Chapter 4. In general, these complexity measures are not only studied in their relationship to the KS entropy, but also used individually or combined in order to compare time-dependent systems, to distinguish determinism, chaos and pure randomness.

The variety of theoretical entropy measures provides a large number of practical measures of complexity, namely by dropping the respective limit. Here, we assume that the considered sequence of finite partitions $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is generating, and thus we approximate the KS entropy by fixing r and t. However, this approximation can be arbitrarily bad, particularly, if the considered system is complex, and the KS entropy is high.

Nevertheless, comparing different values of r and t can give further insight into the dynamics of the time series, and thus can be used for classification problems, detecting entropy changes etc. (see remarks and examples in Chapter 1 and Chapter 4). Note that the choice of $r\in\mathbb{N}$ and the word length $t\in\mathbb{N}$ is also a compromise between the computational costs and the information loss (see Li and Ray [63]).

$$
\xrightarrow{\lim_{t\to\infty}}
$$

$$
\lim_{r\to\infty}\Bigg\downarrow\quad
\begin{array}{ccccccccc}
H_\mu((\mathcal{C}_1)_1) & \frac{1}{2}H_\mu((\mathcal{C}_1)_2) & \frac{1}{3}H_\mu((\mathcal{C}_1)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_1)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_1) \\
H_\mu((\mathcal{C}_2)_1) & \frac{1}{2}H_\mu((\mathcal{C}_2)_2) & \frac{1}{3}H_\mu((\mathcal{C}_2)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_2)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_2) \\
H_\mu((\mathcal{C}_3)_1) & \frac{1}{2}H_\mu((\mathcal{C}_3)_2) & \frac{1}{3}H_\mu((\mathcal{C}_3)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_3)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_3) \\
H_\mu((\mathcal{C}_4)_1) & \frac{1}{2}H_\mu((\mathcal{C}_4)_2) & \frac{1}{3}H_\mu((\mathcal{C}_4)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_4)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_4) \\
H_\mu((\mathcal{C}_5)_1) & \frac{1}{2}H_\mu((\mathcal{C}_5)_2) & \frac{1}{3}H_\mu((\mathcal{C}_5)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_5)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_5) \\
\vdots & \vdots & & \dots & \vdots & \dots & & \vdots \\
H_\mu((\mathcal{C}_m)_1) & \frac{1}{2}H_\mu((\mathcal{C}_m)_2) & \frac{1}{3}H_\mu((\mathcal{C}_m)_3) & \dots & \frac{1}{s}H_\mu((\mathcal{C}_m)_s) & \dots & \searrow & h_\mu(T,\mathcal{C}_m) \\
\vdots & \vdots & & \dots & \vdots & \dots & & \vdots \\
 & & & & & & \nearrow & h_\mu^{\mathrm{KS}}(T)
\end{array}
$$

(a) Approximating the entropy rate $h_\mu(T,\mathcal{C}_m)$ for some $m \in \mathbb{N}$ by computing $\frac{1}{s}H_\mu((\mathcal{C}_m)_s)$ for a word length $s \in \mathbb{N}$, and determining the KS entropy by Lemma 2.3.

$$
\xrightarrow{\lim_{t\to\infty}}
$$

$$
\lim_{r\to\infty}\Bigg\downarrow\quad
\begin{array}{ccccc:cccc}
H_\mu((\mathcal{C}_1)_2|(\mathcal{C}_1)_1) & H_\mu((\mathcal{C}_1)_3|(\mathcal{C}_1)_2) & \dots & H_\mu((\mathcal{C}_1)_{s+1}|(\mathcal{C}_1)_s) & & \dots & \searrow & h_\mu(T,\mathcal{C}_1) \\
H_\mu((\mathcal{C}_2)_2|(\mathcal{C}_2)_1) & H_\mu((\mathcal{C}_2)_3|(\mathcal{C}_2)_2) & \dots & H_\mu((\mathcal{C}_2)_{s+1}|(\mathcal{C}_2)_s) & & \dots & \searrow & h_\mu(T,\mathcal{C}_2) \\
H_\mu((\mathcal{C}_3)_2|(\mathcal{C}_3)_1) & H_\mu((\mathcal{C}_3)_3|(\mathcal{C}_3)_2) & \dots & H_\mu((\mathcal{C}_3)_{s+1}|(\mathcal{C}_3)_s) & & \dots & \searrow & h_\mu(T,\mathcal{C}_3) \\
H_\mu((\mathcal{C}_4)_2|(\mathcal{C}_4)_1) & H_\mu((\mathcal{C}_4)_3|(\mathcal{C}_4)_2) & \dots & H_\mu((\mathcal{C}_4)_{s+1}|(\mathcal{C}_4)_s) & & \dots & \searrow & h_\mu(T,\mathcal{C}_4) \\
H_\mu((\mathcal{C}_5)_2|(\mathcal{C}_5)_1) & H_\mu((\mathcal{C}_5)_3|(\mathcal{C}_5)_2) & \dots & H_\mu((\mathcal{C}_5)_{s+1}|(\mathcal{C}_5)_s) & & \dots & \searrow & h_\mu(T,\mathcal{C}_5) \\
\vdots & \vdots & \dots & \vdots & & \dots & & \vdots \\
H_\mu((\mathcal{C}_m)_2|(\mathcal{C}_m)_1) & H_\mu((\mathcal{C}_m)_3|(\mathcal{C}_m)_2) & \dots & \boxed{H_\mu((\mathcal{C}_m)_{s+1}|(\mathcal{C}_m)_s)} & & \dots & \searrow & h_\mu(T,\mathcal{C}_m) \\
\vdots & \vdots & \dots & \vdots & & \dots & & \vdots \\
 & & & & & & \nearrow & h_\mu^{\mathrm{KS}}(T)
\end{array}
$$

(b) Approximating the entropy rate $h_\mu(T,\mathcal{C}_m)$ for some $m \in \mathbb{N}$ by computing the conditional entropy $H_\mu((\mathcal{C}_m)_{s+1}|(\mathcal{C}_m)_s) = H_\mu((\mathcal{C}_m)_{s+1}) - H_\mu((\mathcal{C}_m)_s)$ for a word length $s \in \mathbb{N}$ (see (2.7) and Remark 2.6), and determining the KS entropy by Lemma 2.3.

Figure 2.5: By Lemma 2.3, the KS entropy is the double-limit of $\frac{1}{t}H_\mu((\mathcal{C}_r)_t)$ (see **(a)**) and $H_\mu\left((\mathcal{C}_r)_{t+1}|(\mathcal{C}_r)_t\right)$ (see **(b)**), respectively, as r and the word length t approaches infinity. The arrows indicate the direction of convergence. The dashed line and the box, respectively, represent Proposition 2.2(1) and Proposition 2.2(2) (see also the notes to this figure in Section 2.2.2).

2.3 Preserving the information of the considered system

By Lemma 2.3 a basic prerequisite for estimating the complexity of a time-dependent system by a symbolic-based analysis method is that the information of the system is preserved, i.e.

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}), \tag{2.10}$$

where $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is the sequence of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ of Ω entailed by the underlying symbolic scheme. We show, in Chapter 3, that under relatively weak assumptions the search for a generating partition can be skipped if one chooses a symbolic scheme that regards a dependency between two measured values. However, in this case, one needs to assume that there is no information lost by the measuring process, i.e.

$$\mathcal{A} \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right). \tag{2.11}$$

In this section, we state sufficient conditions on the measuring process such that (2.11) holds true. Moreover, our results provide conditions on an arbitrary increasing sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions such that (2.10) is satisfied (compare to Remark 2.4). Note that it is natural, to a certain extent, to reconstruct the underlying dynamics T from a realization of the measuring process. For a deeper discussion of the topic, we refer the reader to Takens [80] and Gutman [39].

2.3.1 A metric-based analysis of measurements

Do we lose information if we perform measurements? In this section, we present sufficient conditions on the measuring process such that (2.11) holds true. Note, by setting $T = \text{id}$, these conditions are also sufficient to answer whether

$$\mathcal{A} \overset{\mu}{\subset} \sigma(\mathbf{X}) \tag{2.12}$$

is fulfilled. This is, in particular, interesting if the considered symbolic-based analysis technique preserves the information given by the observables but not the one given by the measuring process (see Chapter 3). In fact, in the case that measurements are analyzed quantitatively, there is a strong link between information preservation and a measure theoretic separation property. Our approach, in this section, is motivated by the following result.

Theorem 2.2 (Antoniouk et al. [9, Theorem 4.2.]). *Let Ω be a separable and completely metrizable topological space (see Appendix A.7 and Appendix A.9). Further, let μ be a measure on $\mathcal{B}(\Omega)$ and $\mathbf{X} = (X_i)_{i=1}^n$ with $n \in \mathbb{N}$. Then*

$$\mathcal{B}(\Omega) \overset{\mu}{\subset} \sigma(\mathbf{X})$$

if the set $\Theta^{\mathbf{X}} := \left\{\omega \in \Omega \mid \mathbf{X}^{-1}(\mathbf{X}(\omega)) \neq \{\omega\}\right\}$ lies in $\mathcal{B}(\Omega)$ and is a μ-null set.

In the following, we call $\Theta^{\mathbf{X}}$ the *set of no* $\mathbf{X}$*-separation*. Moreover, if $\Theta^{\mathbf{X}}$ lies in $\mathcal{A}$ and is a μ-null set, then we say that Ω is *almost surely* $\mathbf{X}$*-separated*, and just $\mathbf{X}$*-separated* if $\Theta^{\mathbf{X}} = \emptyset$.

Now, we restrict our attention to the measuring process $(\mathbf{X}\circ T^{\circ t})_{t\in\mathbb{N}_0}$, and investigate which states are not separated.

Definition 2.8. Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be a sequence of observables. Define the *set of no* $(\mathbf{X}, T)$*-separation* $\Theta^{\mathbf{X},T}$ by

$$\Theta^{\mathbf{X},T} := \left\{ \omega \in \Omega \;\middle|\; \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} \right)^{-1} \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} (\omega) \right) \neq \{\omega\} \right\}.$$

If $\Theta^{\mathbf{X},T} = \emptyset$, we say that Ω is $(\mathbf{X}, T)$*-separated*, and if $\Theta^{\mathbf{X},T}$ lies in $\mathcal{A}$ and is a μ-null set, we say that Ω is *almost surely* $(\mathbf{X}, T)$*-separated*.

Note that if Ω is $(\mathbf{X}, T)$-separated, then, for any two different states ω_1, $\omega_2 \in \Omega$, there exists some $i \in \mathbb{N}$ and some $t \in \mathbb{N}_0$ such that

$$X_i \left(T^{\circ t}(\omega_1)\right) \neq X_i \left(T^{\circ t}(\omega_2)\right). \tag{2.13}$$

If there exists some $i \in \mathbb{N}$ and some $t \in \mathbb{N}_0$ such that (2.13) holds for two different states ω_1, $\omega_2 \in \Omega$, we say ω_1, ω_2 are $(\mathbf{X}, T)$*-separated*. Moreover, note that the restriction of the measuring process $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}$ to any subset $B \subset \Omega \setminus \Theta^{\mathbf{X},T}$, i.e. $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} |_B$, is one-to-one.

Example: Consider the irrational rotation T_a (see Example 2.2 and Example 2.3), the Lebesgue measure λ and the continuous observable Y defined by

$$Y(\omega) = \begin{cases} 0 & \text{if } \omega \in \left[0, \frac{1}{2}\right), \\ \frac{1}{2} - \omega & \text{if } \omega \in \left[\frac{1}{2}, \frac{3}{4}\right), \\ \omega - 1 & \text{if } \omega \in \left[\frac{3}{4}, 1\right) \end{cases} \tag{2.14}$$

(see Figure 2.6). On the one hand, $\Omega = [0, 1)$ is not Y-separated. On the other hand, for any two states $\omega_1 \neq \omega_2$ of Ω, there exists a time $t \in \mathbb{N}$ such that

$$Y\left(T_a^{\circ t}(\omega_1)\right) \neq Y\left(T_a^{\circ t}(\omega_2)\right)$$

since T_a holds dense orbits.

★ **Remark 2.7** (A different representation of the set of no $(\mathbf{X}, T)$-separation). In the following, we consider the dynamical system $(\Omega, \mathcal{B}(\Omega), \mu, T)$ and a sequence $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ of observables. We say that two states $\omega_1 \neq \omega_2$ are equivalent with respect to $\mathbf{X}$ and T if

$$\left(\mathbf{X}\left(T^{\circ t}(\omega_1)\right)\right)_{t\in\mathbb{N}_0} = \left(\mathbf{X}\left(T^{\circ t}(\omega_2)\right)\right)_{t\in\mathbb{N}_0}.$$

Let $\widetilde{\Omega} \subset \Omega$ be a set which contains exactly one element of each equivalence class (we assume the axiom of choice; see for instance the remarks of Billingsley [13, A8.]). If the cardinality of $\widetilde{\Omega}$ is greater than one, then

$$\Theta^{\mathbf{X},T} = \bigcup_{\omega\in\widetilde{\Omega}} \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} \right)^{-1} \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0} (\omega) \right).$$

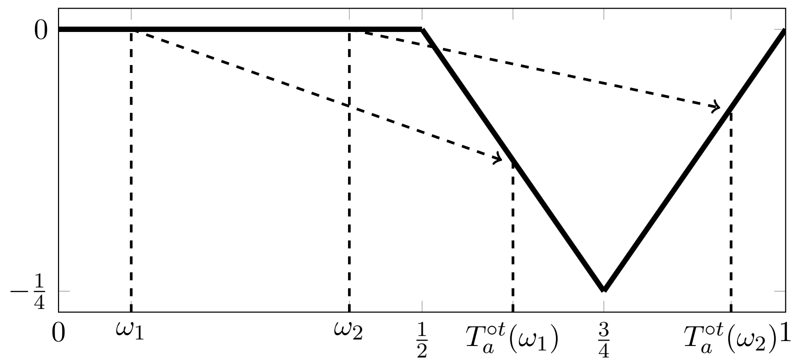


Figure 2.6: Graph of the continuous observable Y as given in (2.14). Consider the irrational rotation T_a, then $\Omega = [0,1)$ is not Y-separated, however, for any two states $\omega_1 \neq \omega_2$, there exists a time $t \in \mathbb{N}$ such that $Y\left(T_a^{\circ t}(\omega_1)\right) \neq Y\left(T_a^{\circ t}(\omega_2)\right)$ (see Example 2.2 and Example 2.3).

Since $\left(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}_0}\right)$ is Hausdorff (see Appendix A.8 and Munkres [68, Theorem 19.4.]) the singleton $\left\{\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}(\omega)\right\}$ is closed for every $\omega \in \widetilde{\Omega}$, and thus measurable with respect to $\mathcal{B}\left(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}_0}\right)$. Hence, for each $\omega \in \widetilde{\Omega}$,

$$A_\omega := \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}\right)^{-1}\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}(\omega)\right) \tag{2.15}$$

is measurable with respect to $\mathcal{B}(\Omega)$ (see also Appendix B.8). Therefore, if $\widetilde{\Omega}$ is countable, i.e. the set $\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}\left(\Theta^{\mathbf{X},T}\right)$ is countable, then $\Theta^{\mathbf{X},T} \in \mathcal{B}(\Omega)$. ★

The following theorem goes back to the work of Antoniouk et al. [9, Section 4.1.] and Keller et al. [49, Lemma 6.9].

Theorem 2.3. *Let $\mathbf{X} = (X_i)_{i \in \mathbb{N}}$ be a sequence of observables and $\Theta^{\mathbf{X},T}$ be the set of no $(\mathbf{X}, T)$-separation. If there exists some $B \in \mathcal{A}$ such that*

(i) $\Theta^{\mathbf{X},T} \subset B$,
(ii) $\mu(B) = 0$ *and*
(iii) $\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}(A \setminus B) \in \mathcal{B}\left(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}_0}\right)$ *for all* $A \in \mathcal{A}$,

then

$$\mathcal{A} \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}_0}\right).$$

Note that by slightly adapting Theorem 2.3 and considering $\Theta^{\mathbf{X}}$ one gets sufficient conditions on the observables such that (2.12) holds true.

★ **Remark 2.8** (A closer look at Theorem 2.3)**.** If Condition (i) and Condition (ii) of Theorem 2.3 are fulfilled, then Ω is *almost surely* $(\mathbf{X}, T)$*-separated.* In order to check Condition (iii) of Theorem 2.3, the following result from descriptive set theory is useful (see for instance Kechris [47, Theorem 15.1] and Kuratowski [59, Theorem 1 in Section 39 V.]). For a deeper discussion of the topic, we refer the reader to Kechris [47], Kuratowski [59] and Cantón et al. [20].

Lemma 2.4. *Let Ω be a separable and completely metrizable topological space (see Appendix A.7 and Appendix A.9). Moreover, let μ be a measure on $\mathcal{B}(\Omega)$ and $X : (\Omega, \mathcal{B}(\Omega), \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. If $A \in \mathcal{B}(\Omega)$ and the restriction $X|_A$ is one-to-one, then $X(A) \in \mathcal{B}(\Omega)$.*

Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be one-to-one. This is true, for instance, if each X_i is one-to-one or each X_i is the i-th coordinate projection (see Remark 2.2). Hence, if Ω is a separable and completely metrizable topological space and μ a measure on $\mathcal{B}(\Omega)$, then $\mathbf{X}(A) \in \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$ for all $A \in \mathcal{B}(\Omega)$. Moreover, $\Theta^{\mathbf{X}} = \emptyset$. Therefore, (2.12) holds true, and thus, in particular, (2.11) is fulfilled. In Section 2.3.2, we set another example of observables that fulfill (2.11). ★

Proof of Theorem 2.3. We need to show that for each $A \in \mathcal{A}$ there exists some

$$F \in \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)$$

such that $\mu(A \triangle F) = 0$. Now, fix $A \in \mathcal{A}$. With B as in Theorem 2.3 set

$$F := A \setminus B.$$

As $A, B \in \mathcal{A}$ also $F \in \mathcal{A}$. Since $(\mathbf{X} \circ T^{\circ t})_{t\in\mathbb{N}_0}|_F$ is one-to-one, we obtain

$$\begin{aligned} F &= \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1} \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}(F)\right) \\ &\overset{(iii)}{\in} \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1} \left(\mathcal{B}\left(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}_0}\right)\right). \end{aligned}$$

Thus $F \in \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)$ (see also Appendix B.8). Finally, by the definition of F, we obtain

$$\mu(A \triangle F) = \mu(A \cap B) \leq \mu(B) = 0,$$

which completes the proof. □

Proposition 2.3. *Let μ be a measure on the Borel σ-algebra $\mathcal{B}(\Omega)$. Suppose that all atoms (see Appendix B.2) are separated, the set $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\left(\Theta^{\mathbf{X},T}\right)$ is countable and $\mathcal{B}(\Omega) \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)$, then $\mu\left(\Theta^{\mathbf{X},T}\right) = 0$.*

Corollary 2.1. *Let μ be a measure on $\mathcal{B}(\Omega)$. Suppose that all atoms are separated by $\mathbf{X}$, the set $\mathbf{X}\left(\Theta^{\mathbf{X}}\right)$ is countable and $\mathcal{B}(\Omega) \overset{\mu}{\subset} \sigma(\mathbf{X})$. Then $\mu\left(\Theta^{\mathbf{X}}\right) = 0$.*

Proof. This follows directly from Proposition 2.3 by setting $T = \mathrm{id}$. □

Proof of Proposition 2.3. We use the results given in Remark 2.7 and conduct a proof by contradiction. Suppose that $\mu\left(\Theta^{\mathbf{X},T}\right) \neq 0$. Note that any A_ω as given in

(2.15) is measurable with respect to $\mathcal{B}(\Omega)$. Since all atoms are separated and $\widetilde{\Omega}$ is countable, there exists an element A_ω and a measurable subset $\widetilde{A}_\omega$ such that

$$\mu(A_\omega) > \mu(\widetilde{A}_\omega) > 0 \text{ and } \mu(A_\omega) > \mu(A_\omega \setminus \widetilde{A}_\omega) > 0.$$

In the case that $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}(\omega) \in B$, we have

$$\widetilde{A}_\omega \subset A_\omega \subset \left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1}(B)$$

and

$$\mu\left(\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1}(B) \bigtriangleup \widetilde{A}_\omega\right) \geq \mu(A_\omega \setminus \widetilde{A}_\omega) > 0.$$

In the case that $\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}(\omega) \notin B$, we have

$$\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1}(B) \cap \widetilde{A}_\omega = \emptyset$$

and, consequently,

$$\mu\left(\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)^{-1}(B) \bigtriangleup \widetilde{A}_\omega\right) \geq \mu(\widetilde{A}_\omega) > 0.$$

Combining the two cases, we see that $\mu(A' \bigtriangleup \widetilde{A}_\omega) > 0$ for any $A' \in \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)$, which contradicts our assumption

$$\mathcal{B}(\Omega) \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right),$$

and completes the proof (see also Appendix B.8). □

2.3.2 Symbolic dynamics

Do we lose information if we apply symbolic dynamics? In this section, we present sufficient conditions on a sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ of Ω such that (2.10) holds. Note that we pay no further attention to the origin of any finite partition in the following (see Remark 2.4).

Recall, that each finite partition of Ω entails a coarse-grained observation of the considered time-dependent system by a random variable (compare to Section 2.1.2). Therefore, we investigate initially which states are not separated by those coarse-grained observations. We consider the dynamical system $(\Omega, \mathcal{B}(\Omega), \mu, T)$ in the following. For a treatment of a more general case, we refer the reader to Rudolph [75].

Definition 2.9. Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be a sequence of finite partitions and $(Y_r)_{r\in\mathbb{N}}$ be the sequence of coarse-grained observations with respect to $(\mathcal{C}_r)_{r\in\mathbb{N}}$, i.e. any Y_r is given by

$$Y_r(\omega) := \sum_{l=1}^{|\mathcal{C}_r|} l \, \mathbf{1}_{C_r^{(l)}}(\omega) \tag{2.16}$$

for $\omega \in \Omega$. If Ω is $(Y_r)_{r\in\mathbb{N}}$-separated, we say that Ω is $((\mathcal{C}_r)_{r\in\mathbb{N}})$-*separated*. Moreover, if Ω is almost surely $((Y_r)_{r\in\mathbb{N}})$-separated, we say that Ω is *almost surely* $((\mathcal{C}_r)_{r\in\mathbb{N}})$-*separated.*

Conventionally (see for instance Rudolph [75]), $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is said to *separate two different states* $\omega_1,\ \omega_2 \in \Omega$ if there exists some $r \in \mathbb{N}$ and some $l \in \mathbb{N}_0$ such that

$$\omega_1 \in C_r^{(l)} \text{ and } \omega_2 \notin C_r^{(l)}. \tag{2.17}$$

Moreover, by convention, Ω is (almost-surely) $((\mathcal{C}_r)_{r\in\mathbb{N}})$-separated if for (almost) all distinct states $\omega_1,\ \omega_2 \in \Omega$ there exists some $i \in \mathbb{N}$ and some $t \in \mathbb{N}_0$ such that (2.17) holds. Note that this complies with the assumption that Ω is almost surely $((Y_r)_{r\in\mathbb{N}})$-separated. Hence we have two equivalent definitions of Ω being almost surely $((\mathcal{C}_r)_{r\in\mathbb{N}})$-separated.

Theorem 2.4. *Let $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be a sequence of finite partitions. If Ω is almost surely $((\mathcal{C}_r)_{r\in\mathbb{N}})$-separated, then*

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}).$$

Proof. We apply Theorem 2.3. In order to do so, let $(Y_r)_{r\in\mathbb{N}}$ be the sequence of coarse-grained observations with respect to $(\mathcal{C}_r)_{r\in\mathbb{N}}$, i.e. each Y_r is given as in (2.16). Since Ω is almost surely $((\mathcal{C}_r)_{r\in\mathbb{N}})$-separated, it is also almost surely $(Y_r)_{r\in\mathbb{N}}$-separated, i.e. $\Theta^{(Y_r)_{r\in\mathbb{N}}}$ lies in $\mathcal{B}(\Omega)$ and is a μ-null set.

Moreover, $Y_r(B) \in \mathcal{B}(\mathbb{R})$ for all $B \in \mathcal{B}(\Omega)$ since any $Y_r(B)$ is a finite union of singletons that are measurable with respect to $\mathcal{B}(\mathbb{R})$ (compare also to Remark 2.7). Hence we have

$$(Y_r)_{r\in\mathbb{N}} \in \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}} = \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$$

(see also Appendix B.8). Moreover, by Theorem 2.3,

$$\mathcal{B}(\Omega) \overset{\mu}{\subset} \sigma((Y_r)_{r\in\mathbb{N}}). \tag{2.18}$$

Further, each Y_r is both $\mathcal{B}(\Omega)$-$\mathcal{B}(\mathbb{R})$-measurable and $\sigma(\mathcal{C}_r)$-$\mathcal{B}(\mathbb{R})$-measurable (see Billingsley [13, remarks on simple real functions in Section 13]). In particular,

$$\sigma((Y_r)_{r\in\mathbb{N}}) \subset \sigma((\mathcal{C}_r)_{r\in\mathbb{N}})$$

which, together with (2.18), is the desired conclusion. □

Chapter 3

A symbolic route to the KS entropy

Does a freely chosen symbolic scheme provide a route to the KS entropy? In this chapter, we show that, under relatively weak assumptions, a symbolic scheme that regards a dependency between two measured values *provides a route to the KS entropy*, i.e. the symbolic scheme entails a sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions that is increasing and generating. This means, from a different angle, the search for a symbolic scheme that *characterizes the KS entropy directly* can be skipped, i.e. the symbolic scheme entails a partition $\mathcal{C}$ that is generating under T. A shortened version of this chapter is published in Stolz and Keller [79].

3.1 Outline of this chapter

We use our general approach of modeling a real-world system by a measure-preserving dynamical system $(\Omega, \mathcal{A}, \mu, T)$ and measurements by random variables

$$X_i : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}));\ i \in \mathbb{N}$$

(see Chapter 2). Recall, that the most ideal, however unrealistic, case is given if we know the underlying dynamics and choose a symbolic scheme that characterizes the KS entropy directly (see for instance Crutchfield and Packard [26], Bollt et al. [15], Kennel and Buhl [56] and the references given therein). Since, as discussed in Section 2.2, if a partition $\mathcal{C} \subset \mathcal{A}$ of Ω is generating under T, i.e.

$$\mathcal{A} \overset{\mu}{\subset} \sigma(((\mathcal{C})_t)_{t\in\mathbb{N}}),$$

where $((\mathcal{C})_t)_{t\in\mathbb{N}}$ is the partition sequence of $\mathcal{C}$ under T, then

$$h_\mu^{\mathrm{KS}}(T) = h_\mu(T, \mathcal{C})$$

(see Lemma 2.1).

Another possibility, discussed in Section 2.2, and of interest in this chapter, is to choose a symbolic scheme that entails a sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions $\mathcal{C}_r \subset \mathcal{A}$ of Ω that is increasing, i.e.

$$\mathcal{C}_r \prec \mathcal{C}_{r+1}$$

for all $r \in \mathbb{N}$ and generating, i.e.

$$\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}), \tag{3.1}$$

because then

$$h_{\mu}^{\mathrm{KS}}(T) = \lim_{r\to\infty} h_{\mu}(T,\mathcal{C}_r) = \sup_{r\in\mathbb{N}} h_{\mu}(T,\mathcal{C}_r) \tag{3.2}$$

(see Lemma 2.3).

We are particularly interested in symbolic schemes which entail special types of increasing sequences $(\mathcal{C}_r)_{r\in\mathbb{N}}$ of finite partitions (see Section 3.2). This enables us to study different symbolic schemes at once and to answer, in this way, whether (3.1), and thus (3.2) holds. The class of symbolic schemes that we consider arises naturally and is sufficiently general. For instance, the class encompasses threshold crossings and the ordinal approach (see Section 3.2.1 and Section 3.2.2). In fact, we obtain an optimal generalization of the ordinal idea which allows us to utilize many of the achievements made in ordinal symbolic dynamics. In the course of this, we mainly focus on results stated in Keller et al. [55], Antoniouk et al. [9] and Keller et al. [49].

Note that we study (3.1) by the real-world scenario where only measurements and symbolic-based analytical tools are at hand, and state sufficient conditions such that the information given by a measuring process is preserved, i.e.

$$\sigma\left(\left(\mathbf{X}\circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right) \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}) \tag{3.3}$$

(see Section 3.3.2). Our results include conditions which ensure that the information given by the observables is preserved, i.e.

$$\sigma(\mathbf{X}) \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}) \tag{3.4}$$

(see Section 3.3.1). Hence, if we assume in the case of (3.3) that

$$\mathcal{A} \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X}\circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right) \tag{3.5}$$

(see Section 2.3.1), and in the case of (3.4) that

$$\mathcal{A} \overset{\mu}{\subset} \sigma(\mathbf{X}), \tag{3.6}$$

then $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is generating and a route to the KS entropy is provided.

Note that if the conditions stated in this chapter are not met, the sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ can still be generating (see Section 2.3.2, Rudolph [75] and the discussion in Section 3.5). This we keep in mind, however, we draw no further attention to it in Section 3.2 and Section 3.3. Moreover, recall that (3.5) is a consequence of (3.6), hence (3.5) is a stronger assumption (see Section 2.3.1 and Figure 2.6). Moreover, if a sequence $(\mathcal{C}_r)_{r\in\mathbb{N}}$ fulfills (3.4), then (3.3) does not necessarily hold true, hence (3.4) is a weaker assumption than (3.3).

Further, we assume that T is ergodic in several statements of this chapter, however, if T is non-ergodic but Ω can be embedded into some compact metrizable space such that $\mathcal{A} = \mathcal{B}(\Omega)$, we still obtain (3.2) by applying the ergodic decomposition theorem (see Section 3.4).

We close this chapter by giving a detailed example, remarks and a visual summary of our results (see Section 3.5 and Section 3.6).

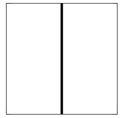
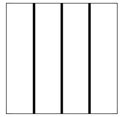
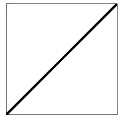
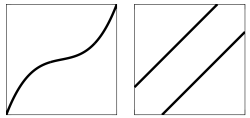
symbolic scheme			properties / description
threshold crossings	single threshold $a \in \mathbb{R}$	$\mathcal{R} = \big\{ \{(x,y) \in \mathbb{R}^2 \mid x \le a\}, \{(x,y) \in \mathbb{R}^2 \mid x > a\} \big\}$	
	subdividing $\mathbb{R}$ into $k \in \mathbb{N}$ intervals $I_1, I_2, \dots I_k$	$\mathcal{R} = \{I_1 \times \mathbb{R}, I_2 \times \mathbb{R}, \dots, I_k \times \mathbb{R}\}$	
		$E_d = \{(0,1), (1,2), (2,3), \dots, (d-1,d)\}$	timing (Def. 3.1), consistent (Def. 3.3)
ordinal	commonly used	$\mathcal{R} = \big\{ \{(x,y) \in \mathbb{R}^2 \mid y \le x\}, \{(x,y) \in \mathbb{R}^2 \mid y > x\} \big\}$	
	• strong	$E_d = \{(s,t) \mid s,t \in \{0,1,2,\dots,d\}$ with $s < t\}$ or $E_d = \{(s,t) \mid s,t \in \{0,\tau,2\tau,\dots,d\tau\}$ with $s < t\}$	full timing (Eq. (3.10)), consistent variant with a delay parameter $\tau \in \mathbb{N}$
	• weak	$E_d = \{(0,t) \mid t \in \{0,1,2,\dots,d\}\}$ or $E_d = \{(0,t) \mid t \in \{0,\tau,2\tau,\dots,d\tau\}\}$	timing, in general not consistent variant with a delay parameter $\tau \in \mathbb{N}$
	relaxed versions with $g : \mathbb{R} \hookleftarrow$	$\mathcal{R} = \{\{(x,y) \in \mathbb{R} \times \mathbb{R} \mid y \le g(x)\}, \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid y > g(x)\}\}$, or finer	for instance: 

Table 3.1: Different symbolic schemes with respective finite partition $\mathcal{R}$ of $\mathbb{R}^2$ and a set of time pairs $E_d \subset \mathbb{N}_0 \times \mathbb{N}_0$ for some natural number d.

3.2 Considering different symbolic schemes at once

We are interested in symbolic schemes that classify the mutual position of measurements at two times s and t according to a finite partition $\mathcal{R}$ of the two-dimensional Euclidean space $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. More precisely, we are interested in a symbolic scheme that entails the finite partition

$$\begin{aligned}&\left(X_i \circ T^{\circ s}, X_i \circ T^{\circ t}\right)^{-1}(\mathcal{R}) \\ &\quad = \left\{\left\{\omega \in \Omega \mid \left(X_i(T^{\circ s}(\omega)), X_i(T^{\circ t}(\omega))\right) \in R\right\} \mid R \in \mathcal{R}\right\} \subset \mathcal{A}\end{aligned} \tag{3.7}$$

of Ω for one observable, or rather the join of finite partitions

$$\mathcal{P}^{\mathcal{R},E}(T, X_i) := \bigvee_{(s,t)\in E} \left(X_i \circ T^{\circ s}, X_i \circ T^{\circ t}\right)^{-1}(\mathcal{R}) \subset \mathcal{A},$$

where $E \subset \mathbb{N}_0 \times \mathbb{N}_0$ is a finite set of time pairs.

Hence, for $n \in \mathbb{N}$ observables, the symbolic scheme entails the finite partition of Ω given by

$$\mathcal{P}^{\mathcal{R},E}\left(T, (X_i)_{i=1}^n\right) := \bigvee_{i=1}^{n} \bigvee_{(s,t)\in E} \left(X_i \circ T^{\circ s}, X_i \circ T^{\circ t}\right)^{-1}(\mathcal{R}) \subset \mathcal{A}. \tag{3.8}$$

Moreover, the symbolic scheme should involve a sequence $(E_d)_{d\in\mathbb{N}}$ with $E_d \subset \mathbb{N}_0 \times \mathbb{N}_0$ in accordance with Definition 3.1. By this,

$$(\mathcal{P}_d)_{d\in\mathbb{N}} := \left(\mathcal{P}^{\mathcal{R},E_d}(T, X_i)\right)_{d\in\mathbb{N}}$$

and

$$(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}} := \left(\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}} \tag{3.9}$$

are increasing sequences of finite partitions (see Lemma 3.1). This also means that, if we choose a finite partition $\mathcal{R}$ of $\mathbb{R}^2$ and a sequence $(E_d)_{d\in\mathbb{N}}$ consistent with Definition 3.1, we obtain symbolic schemes of the form we are interested in. In light of this, we call $\mathcal{R}$ the *basic symbolization scheme* and the tuple $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ *symbolic scheme*. Examples of $\mathcal{R}$ and E_d are listed in Table 3.1. Note that we display the two-dimensional Euclidean space $\mathbb{R}^2$ by a square for illustrative purposes. Moreover, we denote $\mathcal{P}^{\mathcal{R},E_d}(T, X_i)$ and $\mathcal{P}^{\mathcal{R},E_d}(T, (X_i)_{i=1}^n)$ briefly by $\mathcal{P}_d$ and by $\mathcal{P}_{d,n}$, respectively, when no confusion can arise, and refer to $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ as a sequence as well.

Definition 3.1. We call a sequence $(E_d)_{d\in\mathbb{N}}$ of sets E_d with

$$E_1 \subset E_2 \subset \cdots \subset \{(s,t) \mid s,t \in \mathbb{N}_0,\ s < t\}$$

a *timing* if there exists a set $\{v_0, v_1, \dots\} \subseteq \mathbb{N}_0$ with $v_0 < v_1 < v_2 < \dots$ such that:

(i) $E_d \subset \{v_0, v_1, \dots, v_d\}^2$ for each $d \in \mathbb{N}$, and
(ii) for each $s \in \{v_0, v_1, \dots, v_d\}$, there exists some $t \in \{v_0, v_1, \dots, v_d\}$ such that $(s,t) \in E_d$ or $(t,s) \in E_d$.

Condition (i) ensures that each E_d is finite, and Condition (ii) guarantees that each E_d consists of at least d and of no more than $\frac{(d+1)d}{2}$ time pairs. For that reason, the parameter d quantifies how often $\mathcal{R}$ is applied in order to assign a symbol to one state in Ω. In Remark 3.1, we consider Definition 3.1 under graph-theoretic aspects.

A timing is, for instance, given by the sets

$$E_d = \{(s,t) \mid s,t \in \{0,1,2,\dots,d\} \text{ with } s < t\}\,;\; d \in \mathbb{N}. \tag{3.10}$$

Due to the structure of (3.10), we call the associated sequence $(E_d)_{d\in\mathbb{N}}$ the *full timing* in the following.

★ **Remark 3.1** (Symbolic schemes studied by graph theory)**.** A symbolic scheme $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ can be represented by an undirected simple graph, in which the vertices represent time points that are relevant for assigning a symbol to some state in Ω and edges represent an application of $\mathcal{R}$. With that in mind, a timing can be interpreted as a sequence $(E_d)_{d\in\mathbb{N}}$ of edge-sets E_d of an infinite undirected graph (V,E) as follows. Let

$$V := \{v_0, v_1, \dots\} \subseteq \mathbb{N}_0$$

be the vertex set and

$$E := \{(s,t) \in V^2 \mid s < t\}$$

be the edge set. Each E_d is a subset of E. Condition (i) of Definition 3.1 ensures that E_d is finite and only connects vertices in $V_d := \{v_0, v_1, \dots, v_d\}$. Moreover, Condition (ii) of Definition 3.1 ensures that each vertex in V_d is at least incident to one edge in E_d, i.e. the sub-graph (V_d, E_d) of (V,E) does not have any isolated vertices. Further, by the assumption $E_1 \subset E_2 \subset \dots$, the graph (V_{d+1}, E_{d+1}) is an extension of the graph (V_d, E_d) by adding the vertex v_{d+1} to V_d and at least one new edge to E_d that is incident to v_{d+1}. In sum, we have a nested sequence of graphs, in particular, in the case of the full timing a sequence of complete graphs is given. ★

By Definition 3.1, it is indeed ensured that we consider a sequence of finite partitions that is increasing with respect to $\prec$ (see Definition 2.5) as Lemma 3.1 shows.

Lemma 3.1. *Let $(E_d)_{d\in\mathbb{N}}$ be a timing and $\mathcal{R}$ be a finite partition of $\mathbb{R}^2$. The sequence of partitions*

$$\left(\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}}$$

is an increasing sequence in n for fixed d, and for fixed n, it is an increasing sequence in d with respect to $\prec$. In fact,

$$\left(\mathcal{P}^{\mathcal{R},E_{d_j}}\left(T, (X_i)_{i=1}^{n_j}\right)\right)_{d_j,n_j\in\mathbb{N}}$$

is an increasing sequence in j if $(n_j)_{j\in\mathbb{N}}$ and $(d_j)_{j\in\mathbb{N}}$ are increasing sequences in $\mathbb{N}$.

Proof. The proof follows by the design of the partition given in Equation (3.8), which is the coarsest partition refining all

$$\left(X_i \circ T^{\circ s}, X_i \circ T^{\circ t}\right)^{-1}(\mathcal{R})$$

with $i \in \{1, 2, \ldots, n\}$ and $(s, t) \in E_d$ (see Equation (3.7)). Moreover, it holds that $E_d \subset E_{d+1}$. Thus

$$\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^{n}\right) \prec \mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^{n+1}\right)$$

and

$$\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^{n}\right) \prec \mathcal{P}^{\mathcal{R},E_{d+1}}\left(T, (X_i)_{i=1}^{n}\right),$$

which is the desired conclusion. □

★ **Remark 3.2** (Conventions)**.** For subsequent statements, we define

$$\mathcal{P}^{\mathcal{R},E_d}(T, Y_i) := \bigvee_{(s,t)\in E_d} \left(Y_i \circ T^{\circ s}, Y_i \circ T^{\circ t}\right)^{-1}(\mathcal{R}) \subset \mathcal{A}$$

and

$$\mathcal{P}^{\mathcal{R},E_d}\left(T, (Y_i)_{i=1}^{n}\right) := \bigvee_{i=1}^{n} \bigvee_{(s,t)\in E_d} \left(Y_i \circ T^{\circ s}, Y_i \circ T^{\circ t}\right)^{-1}(\mathcal{R}) \subset \mathcal{A}$$

for arbitrary random variables Y_i with $i \in \mathbb{N}$, and consider, for instance, the special case $Y_i = X_i \circ T^{\circ l}$ and $(Y_i)_{i=1}^{n} = \left(X_i \circ T^{\circ l}\right)_{i=1}^{n}$ for some $l, i, n \in \mathbb{N}$, respectively. Moreover, we say a sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ is generating if $\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}})$. ★

In the following, we demonstrate that symbolic schemes $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ are natural and unifying known symbolic approaches (compare also to Table 3.1). In particular, we discuss the determination of the KS entropy in this context (compare also to Section 2.2). Moreover, note that just the mutual position of two measured values is taken into account, entirely in the spirit of—*never forget yesterday but always live for today; you never know what tomorrow can bring or what it can take away.* This complies with the ordinal approach (see Section 3.2.2), and gives an artificial two-dimensional *blow up* of threshold crossings (see Section 3.2.1).

3.2.1 Classical symbolic analysis

Here, we discuss *classical symbolic schemes* (compare to Chapter 1). For convenience, we assume that $\Omega = \mathbb{R}$, and consider the simple case that Ω is subdivided into a finite number of intervals $I_1, I_2, \ldots, I_k$. This method is often called threshold crossings in classical symbolic analysis (see Figure 3.1).

In order to determine the entropy rate $h_\mu(T, \mathcal{C})$ with respect to the finite partition $\mathcal{C} = \{I_1, I_2, \ldots, I_k\}$, we have to generate the sequence $((\mathcal{C})_t)_{t\in\mathbb{N}}$, where

$$(\mathcal{C})_t = \left\{ C^{(a_1,a_2,\ldots,a_t)} \,\middle|\, a_1, a_2, \ldots, a_t \in \{1, 2, \ldots, k\} \right\}$$

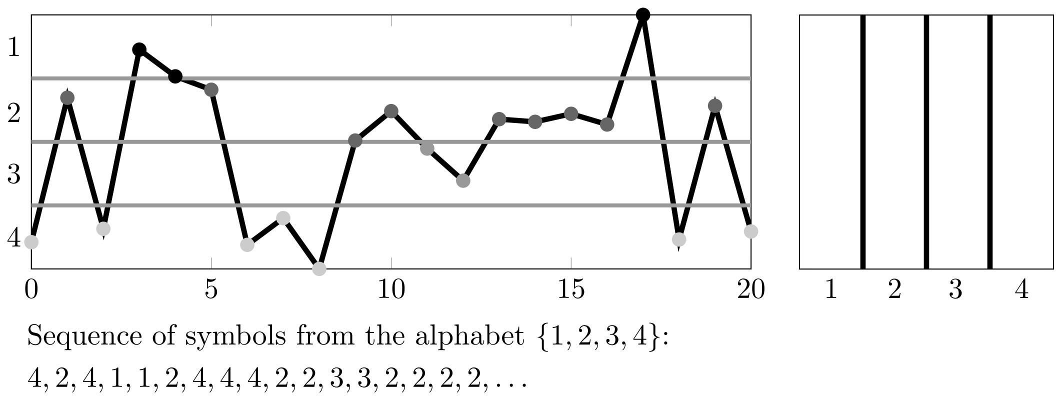


Figure 3.1: A time series is transformed into a sequence of symbols by a threshold crossings technique. Left: intervals of equal size are turned into symbols from the alphabet $\{1, 2, 3, 4\}$. Right: two-dimensional view of the basic symbolization scheme (see Section 3.2.1). This figure is published in Stolz and Keller [79].

(see Section 2.2). Recall, that $(\mathcal{C})_t \prec (\mathcal{C})_{t+1}$ for all $t \in \mathbb{N}$, and

$$C^{(a_1,a_2,\ldots,a_t)} = \left\{\omega \in \Omega \mid \omega \in I_{a_1}, T(\omega) \in I_{a_2}, \ldots, T^{\circ t-1}(\omega) \in I_{a_t}\right\}$$

consists of those states $\omega \in \Omega$ which have the symbolic itinerary $a_1, a_2, \ldots, a_t$. Note that, for all $t \in \mathbb{N}$, it holds

$$(\mathcal{C})_t = \bigvee_{s=0}^{t-1} (T^{\circ s})^{-1}(\mathcal{C}).$$

A classical symbolic scheme can be written as a tuple $(\mathcal{R}, (E_t)_{t\in\mathbb{N}})$ as proposed previously. Choose, for example,

$$\mathcal{R} := \{I_1 \times \mathbb{R}, I_2 \times \mathbb{R}, \ldots, I_k \times \mathbb{R}\} \tag{3.11}$$

(see Figure 3.1) and

$$E_t := \{(0,1), (1,2), (2,3), \ldots, (t-1,t)\};\ t \in \mathbb{N}. \tag{3.12}$$

In fact, (3.11) and (3.12) provide an artificial two-dimensional blow up of any partition $(\mathcal{C})_t$:

$$(\mathcal{C})_t = \bigvee_{s=0}^{t-1} (X \circ T^{\circ s})^{-1}((\mathcal{C})_1) = \bigvee_{(s,u)\in E_t} (X \circ T^{\circ s}, X \circ T^{\circ u})^{-1}(\mathcal{R}) = \mathcal{P}^{\mathcal{R},E_t}(T, X).$$

We deploy the single observable X (meaning $\mathbf{X} = X$ in the general framework) with $X(\omega) = \omega$ for all $\omega \in \Omega$ in accordance with our general approach (compare to Remark 2.8 where we consider coordinate projections). It follows that the partitions $\mathcal{P}_d := \mathcal{P}^{\mathcal{R},E_d}(T, X)$ with (3.11) and (3.12) coincide with the partitions $(\mathcal{C})_d$. In particular, it holds

$$h_\mu(T, \mathcal{P}_d) = h_\mu(T, (\mathcal{C})_d)$$

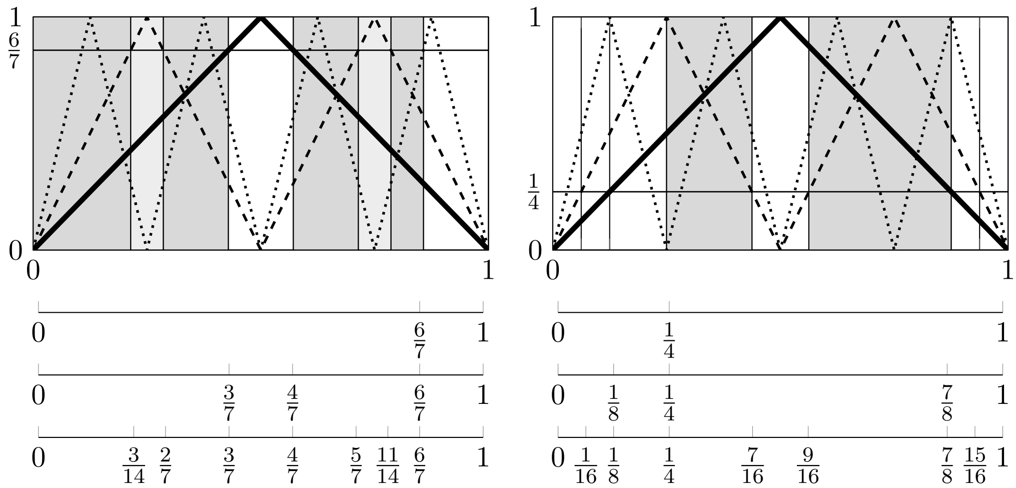


Figure 3.2: Graph of the full tent map T and the iterations $T^{\circ 2}$ and $T^{\circ 3}$ (see examples 2.4, 2.5 and 2.7). Moreover, non-generating partitions under T which are entailed by threshold crossings (see Example 3.1 for more details). Information about possible consequences if a non-generating partition is used in time series analysis is given by Bollt et al. [15]. This figure is a derivative of Stolz and Keller [79, Figure 4].

for all $d \in \mathbb{N}$. Moreover, since $h_\mu(T, (\mathcal{C})_t) = h_\mu(T, \mathcal{C})$ for all $t \in \mathbb{N}$ (see for instance Einsiedler and Schmidt [32, Theorem 3.13]), we obtain

$$h_\mu(T, \mathcal{P}_d) = h_\mu(T, \mathcal{C})$$

for all $d \in \mathbb{N}$. This means, as expected, that $(\mathcal{P}_d)_{d\in\mathbb{N}}$ is generating if and only if $\mathcal{C}$ is generating under T. In other words, $\mathcal{R}$ as given in (3.11) has no generating potential when $\mathcal{C}$ fails to be generating under T.

♣ **Example 3.1** (Threshold crossings applied to the full tent map on $[0, 1)$)**.** In Figure 3.2, we study two different initial partitions, that is on the left $\mathcal{D} = (\mathcal{D})_1 = \left\{\left[0, \frac{6}{7}\right), \left[\frac{6}{7}, 1\right)\right\}$ and on the right $\mathcal{G} = (\mathcal{G})_1 = \left\{\left[0, \frac{1}{4}\right), \left[\frac{1}{4}, 1\right)\right\}$, under the full tent map on $[0, 1)$ (see examples 2.4, 2.5 and 2.7).

Moreover, we visualize, in Figure 3.2, the partitions $(\mathcal{D})_2$, $(\mathcal{D})_3$, $(\mathcal{G})_2$ and $(\mathcal{G})_3$. Note that some elements of the partitions are unions of intervals, this is indicated for $(\mathcal{D})_3$ and $(\mathcal{G})_3$ by different grayscale values, whereas white represents non-united intervals that are elements of $(\mathcal{D})_3$ and $(\mathcal{G})_3$, respectively.

The KS entropy is $\ln(2)$ (see Example 2.7), and in fact, for the initial partition $\mathcal{C} = (\mathcal{C})_1 = \left\{\left[0, \frac{1}{2}\right), \left[\frac{1}{2}, 1\right)\right\}$, it holds

$$H_\mu(\mathcal{C}) = \tfrac{1}{2} H_\mu((\mathcal{C})_2) = \tfrac{1}{3} H_\mu((\mathcal{C})_3) = \cdots = \tfrac{1}{t} H_\mu((\mathcal{C})_t) = \cdots = \ln(2)$$

(see Figure 2.3 and Example 2.4), whereas

$$H_\mu(\mathcal{D}) = -\tfrac{6}{7}\ln\left(\tfrac{6}{7}\right) - \tfrac{1}{7}\ln\left(\tfrac{1}{7}\right) < \ln(2) \text{ and } H_\mu(\mathcal{G}) = -\tfrac{1}{4}\ln\left(\tfrac{1}{4}\right) - \tfrac{3}{4}\ln\left(\tfrac{3}{4}\right) < \ln(2)$$

(see Section 2.2.1). Since $\frac{1}{t} H_\mu((\mathcal{K})_t)$ decreases to $h_\mu(T, \mathcal{K})$ for any finite partition $\mathcal{K} \subset \mathcal{A}$ of Ω (see Walters [87, Chapter 4]), it follows, by Lemma 2.3, that $\mathcal{C}$ is a generating partition, and $\mathcal{D}$ and $\mathcal{G}$ are non-generating partitions under T. Non-generating partitions are also known as *misplaced partitions* (see Bollt et al. [15] and Steuer et al. [78] for detailed information about possible consequences if a non-generating partition is used in time series analysis). ♣

3.2.2 Ordinal symbolic analysis

In the following, we discuss the idea of Bandt and Pompe [11] (compare to Chapter 1) from our perspective. Their approach is particularly interesting since it entails a sequence of finite partitions which is generating (see Antoniouk et al. [9]). In fact, our approach unravels quite clearly the secret behind it and enables us, in turn, to list good basic symbolization schemes and timings such that the corresponding sequences of finite partition have potential to be generating.

Recall, that Bandt and Pompe had the idea to partition the state space according to ordinal patterns of order $d \in \mathbb{N}$. In our framework with a sequence $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ of random variables, this idea reads as follows. Two states ω_1 and $\omega_2 \in \Omega$ are grouped together if for each $i \in \mathbb{N}$ the orbits $\left(X_i\left(T^{\circ t}(\omega_1)\right)\right)_{t=0}^{d}$ and $\left(X_i\left(T^{\circ t}(\omega_2)\right)\right)_{t=0}^{d}$ have the same order relations, i.e. for all s, t with $0 \leq s < t \leq d$ it holds

$$X_i\left(T^{\circ s}(\omega_1)\right) \geq X_i\left(T^{\circ t}(\omega_1)\right) \text{ if and only if } X_i\left(T^{\circ s}(\omega_2)\right) \geq X_i\left(T^{\circ t}(\omega_2)\right).$$

The partition, so obtained, can be written in the form $\mathcal{P}^{\mathcal{R},E_d}(T,(X_i)_{i=1}^{n})$ with the basic symbolization scheme

$$\mathcal{R} = \left\{\left\{(x,y) \in \mathbb{R}^2 \mid y \leq x\right\}, \left\{(x,y) \in \mathbb{R}^2 \mid y > x\right\}\right\} \tag{3.13}$$

and the full timing $(E_d)_{d\in\mathbb{N}}$ as given in (3.10). Thus the sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ with (3.13) and (3.10) is increasing (see Lemma 3.1). Let ω be an element of Ω. Since the approach of Bandt and Pompe is based on the full timing, and therefore, involves to arrange

$$X_i(\omega), X_i(T(\omega)), \ldots, X_i(T^{\circ d}(\omega))$$

into an order, we call the approach *strong ordinal approach* and $d \in \mathbb{N}$ *ordinal order*.

In the literature, the time pairs are often extended by a *delay parameter* $\tau \in \mathbb{N}$, i.e. the sets

$$\left\{(s,t) \mid s,t \in \{0,\tau,2\tau,\ldots,d\tau\} \text{ with } s < t\right\};\ d \in \mathbb{N}$$

are considered (see for instance Cao et al. [21], Keller et al. [53] and Zanin et al. [88]). In this thesis, we confine ourselves to time pairs that are not extended by a delay parameter, however, the results can be adjusted to the extension. However, note here, that delay parameters are interesting for practical purposes, for instance, they allow more flexibility for later applications, and are used to detect complexity changes (see for instance Riedl et al. [73], Unakafova [85] and Keller et al. [50]).

Note that the basic symbolization scheme $\mathcal{R}$, as given in (3.13), regards a dependency between two measurements by X_i. This yields a substantial difference between

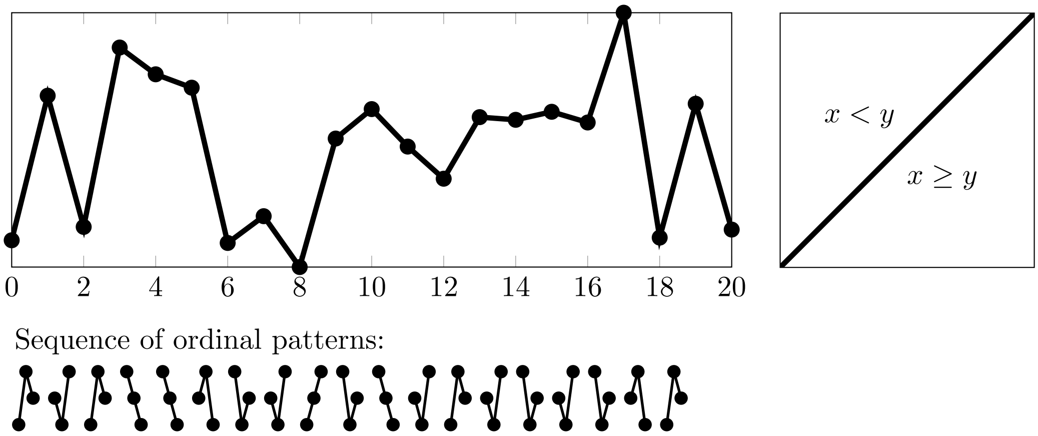


Figure 3.3: A time series is transformed into a sequence of symbols the strong ordinal way. Left: vectors of length 3 are transformed into ordinal patterns of order 2. Right: two-dimensional view of the basic symbolization scheme (see Section 3.2.2). This figure is published in Stolz and Keller [79].

classical and ordinal symbolic schemes which is shown by Antoniouk et al. [9]. The authors state, for $\mathcal{R}$ as given in (3.13) and the full timing, the following:

> If T is ergodic and $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ satisfies (3.5) or weaker (3.6), then $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ is generating, hence
>
> $$h_\mu^{\mathrm{KS}}(T) = \lim_{d,n\to\infty} h_\mu(T, \mathcal{P}_{d,n}) = \sup_{d,n\in\mathbb{N}} h_\mu(T, \mathcal{P}_{d,n}). \tag{♠}$$

Note that Antoniouk et al. [9] worked with a finite random vector $\mathbf{X} = (X_i)_{i=1}^n$, the generalization of their results to infinitely many observables is given by Keller et al. [49].

In conclusion, by choosing $\mathcal{R}$ as given in (3.13) and considering the full timing, we obtain a generating sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ (see Remark 3.2), regardless of whether

$$\mathcal{P}_{1,n} = \mathcal{P}^{\mathcal{R},E_1}\left(T, (X_i)_{i=1}^n\right)$$

is generating under T for some $n \in \mathbb{N}$ (compare to Section 3.2.1). In the next section, we discuss for which $\mathcal{R}$ and $(E_d)_{d\in\mathbb{N}}$ similar results are obtained.

3.3 Generating sequences

Does a symbolic scheme entail a generating sequence of finite partitions? In this section, we look more closely at the choices of $\mathcal{R}$ and $(E_d)_{d\in\mathbb{N}}$ for which the sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ of finite partitions preserves the information given by the observables (see Equation (3.4)) or even the information given by the measuring process (see Equation (3.3)). In fact, we show that not only the strong ordinal approach fulfills (3.4) or even (3.3), but also other symbolic-based techniques with symbolic schemes $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$.

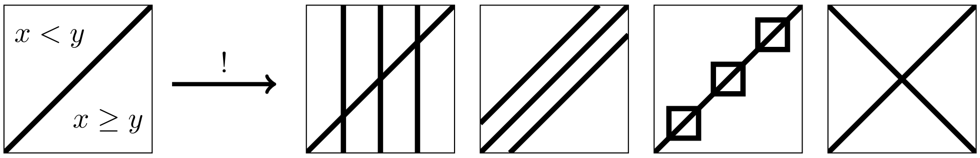


Figure 3.4: Statement (♠) remains true if $\mathcal{R}$ is substituted by a refinement of (3.13). This figure is published in Stolz and Keller [79].

The first theorem, we present here, motivates our general idea and our results. It is a special case of Theorem 3.2 (see Section 3.3.1) and an extension of ordinal symbolic analysis (compare to Figure 3.4). Since we want to include as many symbolic schemes as possible, we deploy a map $g : \mathbb{R} \hookleftarrow$ in our considerations (compare also to Section 3.3.1).

Theorem 3.1. *Let Ω be a Borel subset of $\mathbb{R}^{\mathbb{N}}$ and X_i with $i \in \mathbb{N}$ be the i-th coordinate projection. Moreover, let $(E_d)_{d\in\mathbb{N}}$ be the full timing and $\mathcal{R}$ be the basic symbolization scheme defined by*

$$\mathcal{R} = \left\{ \left\{ (x, y) \in \mathbb{R}^2 \mid y \leq g(x) \right\}, \left\{ (x, y) \in \mathbb{R}^2 \mid y > g(x) \right\} \right\}, \tag{3.14}$$

or a refinement of (3.14)*, where $g : \mathbb{R} \hookleftarrow$ is a one-to-one $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable map. If T is ergodic and $\mu(A) > 0$ for all open non-empty subsets A of Ω, then Statement (♠) is fulfilled for the tuple $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$.*

Note that (3.14) agrees with (3.13) for the special case that g coincides with the identity map, and the coordinate projections entail that (3.6) is fulfilled (see Remark 2.8). Moreover, it is worth noting that a basic symbolization scheme as given in (3.14) regards a dependency between the value measured by $g \circ X_i$ at some time $s \in \mathbb{N}$ and the value measured by X_i at some time $t \in \mathbb{N}$ with $s < t$ by construction. This is true since

$$\left\{ X_i \circ T^{\circ t} \leq g \circ X_i \circ T^{\circ s} \right\} \in \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}})$$

for all $i \in \mathbb{N}$ and $s, t \in \mathbb{N}_0$ with $s < t$. Note that the previous assertion is also fulfilled if a basic symbolization scheme is considered that refines (3.14).

The requirements on g and on the open non-empty subsets of Ω, in Theorem 3.1, are not immediately evident, however, are sufficient to avoid any information loss by considering $g \circ X_i$ and $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$, respectively (see Section 3.3.1).

3.3.1 Preserving the information given by observables

In this section, we study the question whether a symbolic-based analysis technique with underlying symbolic scheme $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ preserves the information given by the observables. In particular, we show that (3.4) is met if the considered symbolic scheme retains, for every observable X_i of the sequence $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ and $\omega \in \Omega$,

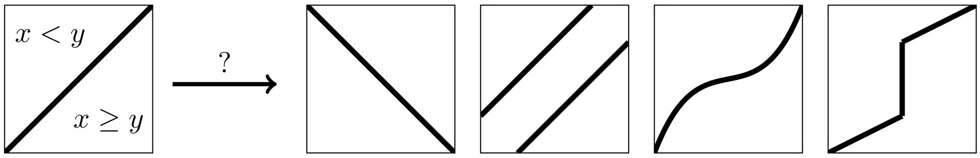


Figure 3.5: The question arises whether Statement (♠) remains true if arbitrary basic symbolization schemes are considered (see Section 3.3). This figure is published in Stolz and Keller [79].

information on whether the graph of $X_i(T^{\circ t}(\omega))$ with respect to $t \in \mathbb{N}$ passes the measured value $X_i(\omega)$. In order to do this, we utilize that, in the ergodic case, the relative number of how often a measured value is passed over time coincides almost surely with $F_{X_i} \circ X_i$. The map $F_{X_i} : \mathbb{R} \to [0, 1]$ is the distribution function of X_i given by

$$F_{X_i}(a) := \mu(\{\omega \in \Omega \mid X_i(\omega) \leq a\})$$

for all $a \in \mathbb{R}$. Moreover, we utilize that $\sigma(X_i) \overset{\mu}{\subset} \sigma(F_{X_i} \circ X_i)$.

Since we want to include as many symbolic schemes as possible, we deploy a self-map $g : \mathbb{R} \hookleftarrow$ in our considerations. Thereby, we take symbolic schemes into account which only retain information on whether the graph of $X_i(T^{\circ t}(\omega))$ passes the measured value $g \circ X_i(\omega)$. In fact, as we are going to show in this section, if $\sigma(X_i) \overset{\mu}{\subset} \sigma(g \circ X_i)$, then $\sigma(X_i) \overset{\mu}{\subset} \sigma\left((\mathcal{C}_r)_{r \in \mathbb{N}}\right)$.

Definition 3.2. Let $X : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. We call a map $\phi : \mathbb{R} \hookleftarrow$ admissible with respect to X if

$$\sigma(X) \overset{\mu}{\subset} \sigma(\phi \circ X).$$

A map $\phi : \mathbb{R} \hookleftarrow$ is, for instance, admissible with respect to a random variable X if it is a one-to-one $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable map (see Remark 3.3). More general conditions on ϕ are given in Lemma 3.4.

We present the following theorem in a very general form, i.e. we consider an arbitrary sequence $(\mathcal{C}_r)_{r \in \mathbb{N}}$ of finite partitions, in order to emphasize the essentials. When we have a certain symbolic scheme in mind, and thus a fixed sequence of finite partitions, we replace $(\mathcal{C}_r)_{r \in \mathbb{N}}$ accordingly.

Theorem 3.2. *Let T be ergodic, X an observable, F_X the distribution function of X and $(\mathcal{C}_r)_{r \in \mathbb{N}}$ be a sequence of finite partitions. If there exists a map $g : \mathbb{R} \hookleftarrow$ such that*

(i) *g is admissible with respect to X,*
(ii) *F_X is admissible with respect to $g \circ X$ and*
(iii) $\left\{X \circ T^{\circ t} \leq g \circ X\right\} \in \sigma((\mathcal{C}_r)_{r \in \mathbb{N}})$ *for all $t \in \mathbb{N}_0$,*

then

$$\sigma(X) \overset{\mu}{\subset} \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}).$$

Note that Condition (i) and Condition (ii) are fulfilled if g is the identity map (see the proof of Lemma 3.4 and Remark 3.3). Hence, if a symbolic scheme yields

$$\left\{X \circ T^{\circ t} \leq X\right\} \in \sigma((\mathcal{C}_r)_{r\in\mathbb{N}}),$$

for all $t \in \mathbb{N}$, then, by Theorem 3.2, the information given by the observable X is preserved. This is for example the case if an ordinal approach with the basic symbolization scheme as given in (3.13), or a refinement of (3.13), is considered. Recall also that, in the case of classical symbolic schemes, Condition (iii) cannot be guaranteed, and therefore, Theorem 3.2 cannot be applied (compare to the deliberations and examples in Section 3.2.1 and Section 3.2.2).

In order to prove Theorem 3.2, we generalize the results of Antoniouk et al. [9, Lemmas 3.2, 3.3 and Corollary 3.4] by extending their proofs. For this purpose, let X and Y be two observables and F_X the distribution function of X. Note that the relative number of how often the graph of $X(T^{\circ t}(\omega))$ passes the measured value $Y(\omega)$ for some $\omega \in \Omega$ up to a certain time point $d \in \mathbb{N}$ is given by the *counting map* $I_d^{X,Y} : \Omega \to [0,1]$ defined by

$$I_d^{X,Y}(\omega) := \frac{1}{d}\sum_{t=1}^{d} \mathbf{1}_{\{X \circ T^{\circ t} \leq Y\}}(\omega)$$

for all $\omega \in \Omega$. We have divided the proof of Theorem 3.2 into a sequence of lemmas. In Lemma 3.2, we show the close linkage between $I_d^{X,Y}$ and $F_X \circ Y$ if T is ergodic. Note that

$$F_X(Y(\omega)) := \mu(\{\omega^* \in \Omega \mid X(\omega^*) \leq Y(\omega)\})$$

for all $\omega \in \Omega$. Further, we show in Lemma 3.3 that no information is lost if $(\mathcal{C}_r)_{r\in\mathbb{N}}$ is considered instead of $F_X \circ Y$. Finally, we state in Lemma 3.4 sufficient conditions on a map $\phi : \mathbb{R} \hookleftarrow$ such that it is admissible with respect to X.

Lemma 3.2. *Let X and Y be two observables, $F_X : \mathbb{R} \to [0,1]$ the distribution function of X and $I_d^{X,Y} : \Omega \to [0,1]$ be the counting map of X and Y. If T is ergodic, then*

$$\lim_{d\to\infty} I_d^{X,Y} = F_X \circ Y \quad \mu\text{-almost everywhere.}$$

Proof. Let $A_a = X^{-1}((-\infty, a])$ for any $a \in \mathbb{R}$. By Birkhoff's ergodic theorem (see Remark 2.1), there exists a set $N_a \subset \Omega$ such that $\mu(N_a) = 0$ and

$$F_X(a) = \mu(A_a) = \lim_{d\to\infty} \frac{1}{d}\sum_{t=1}^{d} \mathbf{1}_{\{X \circ T^{\circ t} \leq a\}}(\omega) \tag{3.15}$$

for any $a \in \mathbb{R}$ and $\omega \in \Omega \setminus N_a$. Let B be a countable dense subset of $\mathbb{R}$ such that it includes all $a \in \mathbb{R}$ for which F_X is discontinuous. Further, let $N = \bigcup_{a\in B} N_a$. Then

$\mu(N) = 0$ and (3.15) holds for each $a \in B$ and $\omega \in \Omega \setminus N$. Our next claim is that, for all $\omega \in \Omega \setminus N$, it holds

$$\lim_{d\to\infty} I_d^{X,Y}(\omega) = F_X(Y(\omega)).$$

By (3.15), this is true if $\omega \in \Omega \setminus N$ satisfies $a := Y(\omega) \in B$. It is moreover true if $\omega \in \Omega \setminus N$ and $a := Y(\omega) \in \mathbb{R} \setminus B$, which we show in the following.

Let $(b_i)_{i\in\mathbb{N}}$ and $(c_i)_{i\in\mathbb{N}}$ be two sequences converging to a with

$$b_i \in B \cap (-\infty, a) \text{ and } c_i \in B \cap (a, \infty)$$

for all $i \in \mathbb{N}$. Since $\omega \in \Omega \setminus N$, we have, for all $i \in \mathbb{N}$, that

$$F_X(b_i) = \lim_{d\to\infty} \frac{1}{d}\sum_{t=1}^{d} \mathbf{1}_{\{X\circ T^{\circ t} \le b_i\}}(\omega) \text{ and } F_X(c_i) = \lim_{d\to\infty} \frac{1}{d}\sum_{t=1}^{d} \mathbf{1}_{\{X\circ T^{\circ t} \le c_i\}}(\omega).$$

Moreover, $b_i < a < c_i$ implies

$$\sum_{t=1}^{d} \mathbf{1}_{\{X\circ T^{\circ t} \le b_i\}}(\omega) \le \sum_{t=1}^{d} \mathbf{1}_{\{X\circ T^{\circ t} \le Y\}}(\omega) \le \sum_{t=1}^{d} \mathbf{1}_{\{X\circ T^{\circ t} \le c_i\}}(\omega)$$

for all $d \in \mathbb{N}$. Furthermore, since F_X is continuous at a, we obtain

$$F_X(a) = \lim_{i\to\infty} F_X(b_i) \le \liminf_{d\to\infty} I_d^{X,Y} \le \limsup_{d\to\infty} I_d^{X,Y} \le \lim_{i\to\infty} F_X(c_i) = F_X(a).$$

Hence we can summarize that, for all $\omega \in \Omega \setminus N$, it holds $\lim_{d\to\infty} I_d^{X,Y}(\omega) = F_X(Y(\omega))$, which is the desired conclusion. □

Lemma 3.3. *Let X and Y be two observables and $(\mathcal{C}_r)_{r\in\mathbb{N}}$ be a sequence of finite partitions. If T is ergodic and*

$$\{X \circ T^{\circ t} \le Y\} \in \sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right)$$

for each $t \in \mathbb{N}_0$, then

$$\sigma\left(F_X \circ Y\right) \overset{\mu}{\subset} \sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right).$$

Proof. By assumption, $I_d^{X,Y}$ is $\sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right)$-$\mathcal{B}([0,1])$-measurable for any $d \in \mathbb{N}$ (see for instance Billingsley [13, Remarks on simple real functions in Section 13]). Hence

$$\sigma\left(\left(I_d^{X,Y}\right)_{d\in\mathbb{N}}\right) \subset \sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right).$$

Moreover, the limit of $I_d^{X,Y}$ as d approaches infinity exists for each $\omega \in \Omega$ since $I_d^{X,Y} \le I_{d+1}^{X,Y}$ and $0 \le I_d^{X,Y} \le 1$, hence

$$\sigma\left(\lim_{d\to\infty} I_d^{X,Y}\right) \subset \sigma\left(\left(I_d^{X,Y}\right)_{d\in\mathbb{N}}\right)$$

(see for instance Billingsley [13, Theorem 13.4.(ii)]). Furthermore, by Lemma 3.2, there exists a set $N \subset \Omega$ with $\mu(N) = 0$ such that

$$\lim_{d\to\infty} I_d^{X,Y}(\omega) = F_X(Y(\omega))$$

for all $\omega \in \Omega \setminus N$. Hence, for any $B \in \mathcal{B}([0,1])$, it holds

$$\mu\left((F_X \circ Y)^{-1}(B) \triangle \left(\lim_{d\to\infty} I_d^{X,Y}\right)^{-1}(B)\right) \leq \mu(N) = 0,$$

which gives

$$\sigma(F_X \circ Y) \overset{\mu}{\subset} \sigma\left(\left(I_d^{X,Y}\right)_{d\in\mathbb{N}}\right),$$

and the lemma follows. $\square$

Lemma 3.4. *Let $X : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable, $\phi : \mathbb{R} \hookleftarrow$ a $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable map and $\mathcal{G}$ be a family of subsets of $\mathbb{R}$ that generates $\mathcal{B}(\mathbb{R})$. If ϕ has the two properties*

(i) $\phi(G) \in \mathcal{B}(\mathbb{R})$ *and*
(ii) $\mu\left(X^{-1}\left(\left(\phi^{-1}\phi(G)\right) \setminus G\right)\right) = 0$

for all $G \in \mathcal{G}$, then $\sigma(X) \overset{\mu}{\subset} \sigma(\phi \circ X)$.

Proof. Since $\mathcal{G}$ generates $\mathcal{B}(\mathbb{R})$, it holds that $\sigma(X)$ is generated by the sets $X^{-1}(\mathcal{G})$ (see for instance Elstrodt [34, Chapter 1, Theorem 4.4]). Hence the lemma is proven if for any $G \in \mathcal{G}$, there exists some $G' \in \sigma(\phi \circ X)$ such that $\mu\left(X^{-1}(G) \triangle G'\right) = 0$. In order to show this, choose

$$G' = X^{-1}\left(\phi^{-1}\phi(G)\right) = (\phi \circ X)^{-1}\phi(G).$$

By (i), it holds that $G' \in \sigma(\phi \circ X)$, and by (ii), we see that

$$\mu\left(X^{-1}(G) \triangle G'\right) = \mu\left(X^{-1}\left(\left(\phi^{-1}\phi(G)\right) \setminus G\right)\right) = 0,$$

which completes the proof. $\square$

We now prove Theorem 3.2 and Theorem 3.1, and close this section with remarks to Lemma 3.4.

Proof of Theorem 3.2. By Condition (i) and Condition (ii), we obtain

$$\sigma(X) \overset{\mu}{\subset} \sigma(g \circ X) \overset{\mu}{\subset} \sigma(F_X \circ g \circ X)$$

(compare also to Lemma 3.4).

Moreover, by Condition (iii) and Lemma 3.3, we have that

$$\sigma(F_X \circ g \circ X) \overset{\mu}{\subset} \sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right).$$

Hence $\sigma(X) \overset{\mu}{\subset} \sigma\left((\mathcal{C}_r)_{r\in\mathbb{N}}\right)$, which completes the proof. $\square$

Proof of Theorem 3.1. In the following, we show that Theorem 3.1 follows by Theorem 3.2. Firstly, since Ω is a Borel subset of $\mathbb{R}^n$, X_i with $i \in \mathbb{N}$ is the i-th coordinate projection and $\mu(A) > 0$ for all open non-empty subsets A of Ω, it holds not only (3.6), but also that each F_{X_i} is one-to-one. Hence

$$\sigma(g \circ X_i) \overset{\mu}{\subset} \sigma(F_{X_i} \circ g \circ X_i)$$

for all $i \in \mathbb{N}$ which complies with Condition (ii) of Theorem 3.2 (see the proof of Lemma 3.4 and Remark 3.3).

Secondly, since $\mathcal{R}$ is given as in (3.14), or is a refinement of (3.14), with some one-to-one map $g : \mathbb{R} \hookleftarrow$, it holds

$$\left\{X_i \circ T^{\circ t} \leq g \circ X_i\right\} \in \sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}}\right)$$

for all $t \in \mathbb{N}_0$ and $i \in \mathbb{N}$ which complies with Condition (iii) of Theorem 3.2. Moreover, since g is one-to-one, it holds

$$\sigma(X_i) \overset{\mu}{\subset} \sigma(g \circ X_i)$$

for all $i \in \mathbb{N}$ which complies with Condition (i) of Theorem 3.2.

Thus, by the conditions of Theorem 3.1, we can apply Theorem 3.2 for all X_i which gives

$$\sigma(\mathbf{X}) = \bigvee_{i\in\mathbb{N}} \sigma(X_i) \overset{\mu}{\subset} \bigvee_{i\in\mathbb{N}} \sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}(T, X_i)\right)_{d\in\mathbb{N}}\right) = \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}})$$

since the sub-σ-algebras $\sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}(T, X_i)\right)_{d\in\mathbb{N}}\right)$ generate the sub-σ-algebra

$$\sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}) = \sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}}\right)$$

(see Equation (2.2)). Hence Statement (♠) is fulfilled which is the desired conclusion. □

Note that the timing $(E_d)_{d\in\mathbb{N}}$ plays no part in the previous proof, since the observables have the strong property (3.6). Hence we can replace the full timing in Theorem 3.1 by some arbitrary timing. For a fuller treatment of the case when (3.6) is not met, see Section 3.3.2.

★ **Remark 3.3** (Lemma 3.4: one-to-one measurable maps and F_X)**.** Note that since ϕ is a $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable map in Lemma 3.4, the inclusion $\sigma(X) \supset \sigma(\phi \circ X)$ holds for any random variable $X : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In order to see this, let $A \in \sigma(\phi \circ X)$, then

$$A = (\phi \circ Y)^{-1}(B) = X^{-1}\phi^{-1}(B)$$

for some $B \in \mathcal{B}(\mathbb{R})$. Hence, by $\phi^{-1}(B) \in \mathcal{B}(\mathbb{R})$, it follows that $A \in \sigma(X)$.

Figure 3.6 demonstrates that g has to be chosen reasonably with respect to the given measurements. This is similar to the choice of thresholds when threshold crossings are applied.

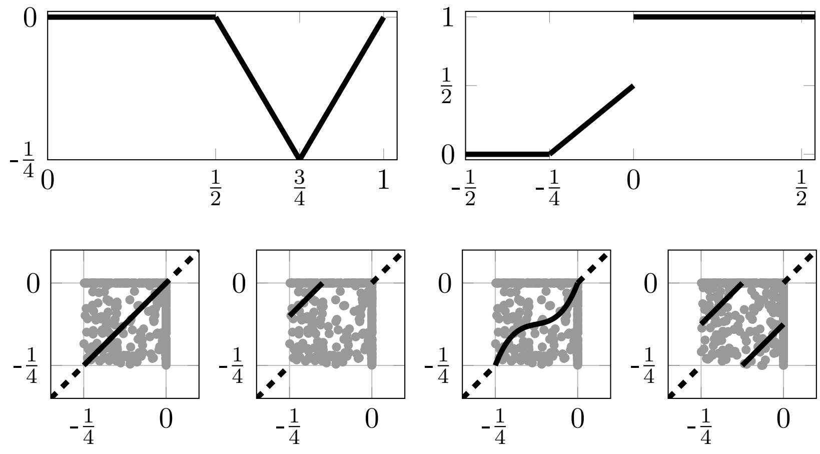


Figure 3.6: On top: Graphs of the continuous observable $Y : [0,1) \to \mathbb{R}$ given in Equation (2.14) and of the distribution function $F_Y : \mathbb{R} \to [0,1]$. On the bottom: Scatter plots of 500 pairs $(Y(\omega_1), Y(\omega_2))$ with (ω_1, ω_2) normally distributed on $[0,1)^2$ with the respective graph of a map $g : \mathbb{R} \hookleftarrow$ which is admissible with respect to F_Y if the Lebesgue measure is considered.

Lemma 3.4 is evident if ϕ is a one-to-one $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable map, because then

$$\left(\phi^{-1}\phi(B)\right) \setminus B = \emptyset$$

and $\phi(B) \in \mathcal{B}(\mathbb{R})$ for all $B \in \mathcal{B}(\mathbb{R})$ (see for instance Cantón et al. [20], compare to Lemma 2.4 as well). Nevertheless, Lemma 3.4 also includes self-maps such as the distribution function F_X of a random variable X (see Antoniouk et al. [9, Lemma 3.2]), which can be seen as follows.

Let $\mathcal{G} = \{(-\infty, a) \mid a \in \mathbb{R}\}$. Since F_X is increasing, it holds Lemma 3.4(i) for all $G \in \mathcal{G}$. Lemma 3.4(ii) is proven by Antoniouk et al. [9, Lemma 3.1(3)]. Firstly, the authors show that

$$F_X^{-1}F_X((-\infty, a)) \setminus (-\infty, a)$$

coincides either with the interval $[a, a^*]$ or with $[a, a^*)$ for any $a \in \mathbb{R}$, where

$$a^* = \sup\left(F_X^{-1}F_X(a)\right).$$

Secondly, the authors prove that $\mu\left(X^{-1}([a, a^*])\right) = 0$. Hence

$$\sigma(X) \overset{\mu}{\subset} \sigma(F_X \circ X).$$

However, this also shows that the inclusion (compare to Theorem 3.2(ii))

$$\sigma(g \circ X) \overset{\mu}{\subset} \sigma\left(F_X \circ g \circ X\right),$$

where $g : \mathbb{R} \hookleftarrow$ is a self-map, holds if we obtain

$$\mu(X^{-1}g^{-1}([a, a^*])) = 0$$

for any $a \in \mathbb{R}$. This is true if, for instance, either F_X is one-to-one or $g(\omega) = \omega$ for all $\omega \in \Omega$ where F_X is not one-to-one (compare to Figure 3.5). ★

3.3.2 Preserving the information given by the measuring process

In this section, we study the question whether a symbolic-based analysis technique with underlying symbolic scheme $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ preserves the information given by the measuring process. In fact, we show that (3.3) is met if the timing $(E_d)_{d\in\mathbb{N}}$ has the following property.

Definition 3.3. Let $(E_d)_{d\in\mathbb{N}}$ be a timing. If for each $d \in \mathbb{N}$ and time pair $(u, v) \in E_d$ it holds that $(u + 1, v + 1) \in E_{d+1}$, then we call $(E_d)_{d\in\mathbb{N}}$ *consistent.*

For instance, the full timing (see Equation (3.10)) is consistent. However, the following timing

$$E_d = \{(0, t) \mid t \in \{0, 1, 2, \ldots, d\}\}\ ;\ d \in \mathbb{N} \tag{3.16}$$

is not consistent (see Example 3.2 and compare to Section 3.5). Thus we call (3.16) the *weak timing* in the following.

Proposition 3.1. *Let X be an observable, $\mathcal{R}$ a basic symbolization scheme, $(E_d)_{d\in\mathbb{N}}$ a timing. If $(E_d)_{d\in\mathbb{N}}$ is consistent, then*

$$(\mathcal{P}_d)_t := \bigvee_{s=0}^{t-1} \mathcal{P}^{\mathcal{R},E_d}(T, X \circ T^{\circ s}) \prec \mathcal{P}^{\mathcal{R},E_{d+t-1}}(T, X) \tag{3.17}$$

for all $d \in \mathbb{N}$ and $t \in \mathbb{N}$.

Recall, that $((\mathcal{P}_d)_t)_{t\in\mathbb{N}}$ is the partition sequence of $\mathcal{P}_d$ under T. In sum, consistency guarantees that the considered symbolic-based analysis technique delivers the same result regardless where we start to observe the underlying system.

Proof of Proposition 3.1. Let us regard d as fixed. First observe that, in the case of consistency, for any $t \in \mathbb{N}$, the pairs $(s + u, s + v)$ with $s \in \{0, 1, \ldots, t - 1\}$ and $(u, v) \in E_d$ are element of E_{d+t-1} since $E_d \subset E_{d+1}$. Moreover, it holds

$$\left(X \circ T^{\circ u+1}, X \circ T^{\circ v+1}\right)^{-1}(\mathcal{R}) \prec \mathcal{P}^{\mathcal{R},E_{d+1}}(T, X)$$

(compare to Remark 3.2). Hence

$$\mathcal{P}^{\mathcal{R},E_d}(T, X \circ T) \prec \mathcal{P}^{\mathcal{R},E_{d+1}}(T, X).$$

In particular, it holds

$$\begin{aligned}\mathcal{P}^{\mathcal{R},E_d}(T, X \circ T^{\circ t-1}) &\prec \mathcal{P}^{\mathcal{R},E_{d+1}}(T, X \circ T^{\circ t-2}) \\ &\prec \mathcal{P}^{\mathcal{R},E_{d+2}}(T, X \circ T^{\circ t-3}) \prec \cdots \prec \mathcal{P}^{\mathcal{R},E_{d+t-1}}(T, X)\end{aligned}$$

for all $t \in \mathbb{N}$, which gives (3.17). □

♣ **Example 3.2** (The weak timing)**.** Note that a similar example is given in Keller et al. [49]. Let $\Omega = [0, 1)$, $T : \Omega \hookleftarrow$ be the full tent map and $\mathcal{R}$ be as given in (3.13). Define an observable Y by the staircase function

$$Y(\omega) := \sum_{l=1}^{4} l\, 1_{\left[\frac{l-1}{4}, \frac{l}{4}\right)}(\omega)$$

for all $\omega \in \Omega$. By choosing $\omega_1 = \frac{1}{4}$ and $\omega_2 = \frac{7}{12}$, we obtain

$$\left(Y\left(T^{\circ t}(\omega_1)\right)\right)_{t \in \mathbb{N}_0} = (2, 3, 1, 1, 1, 1, \ldots),$$

and

$$\left(Y\left(T^{\circ t}(\omega_2)\right)\right)_{t \in \mathbb{N}_0} = (3, 4, 2, 3, 3, 3, \ldots).$$

Hence ω_1 and ω_2 are separated by the partition $(Y \circ T^2, Y \circ T^3)^{-1}(\mathcal{R})$, and thus, by the sequence of finite partitions entailed by the strong ordinal approach. However, the time pair $(2, 3)$ is not element of any E_d as given in (3.16), i.e. for all $d \in \mathbb{N}$ and $t \in \mathbb{N}_0$, it holds

$$\mathcal{P}^{\mathcal{R}, E_d}(T, Y \circ T^{\circ 2}) \not\prec \mathcal{P}^{\mathcal{R}, E_{d+t}}(T, Y).$$

Consequently, the weak timing $(E_d)_{d \in \mathbb{N}}$ is not consistent. Note, that we call the approach based on (3.13) and the weak timing *the weak ordinal approach* in the following. ♣

Let us now state our main result with respect to (3.3).

Theorem 3.3. *Let* $\mathbf{X} = (X_i)_{i \in \mathbb{N}}$ *be a sequence of observables,* $\mathcal{R}$ *a basic symbolization scheme,* $(E_d)_{d \in \mathbb{N}}$ *a timing, and* $(\mathcal{P}_{d,n})_{d,n \in \mathbb{N}}$ *be the sequence of finite partitions constructed by* $\mathcal{R}$ *and* $(E_d)_{d \in \mathbb{N}}$*. If*

(i) $\sigma(\mathbf{X}) \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n \in \mathbb{N}})$ *and*
(ii) the timing $(E_d)_{d \in \mathbb{N}}$ *is consistent,*

then

$$\sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t \in \mathbb{N}}\right) \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n \in \mathbb{N}}).$$

Proof. Compare to Antoniouk et al. [9, Corollary 3.5]. By (ii), it holds

$$\mathcal{P}^{\mathcal{R}, E_d}(T, X_i \circ T) \prec \mathcal{P}^{\mathcal{R}, E_{d+1}}(T, X_i)$$

for all d and $i \in \mathbb{N}$. These refinements imply

$$\begin{aligned} \sigma\left(\left(\mathcal{P}^{\mathcal{R}, E_d}\left(T, \left(X_i \circ T^{\circ t}\right)_{i=1}^{n}\right)\right)_{d,n \in \mathbb{N}}\right) &\subseteq \sigma((\mathcal{P}_{d,n})_{d,n \in \mathbb{N}}) \\ &= \sigma\left(\left(\mathcal{P}^{\mathcal{R}, E_d}\left(T, (X_i)_{i=1}^{n}\right)\right)_{d,n \in \mathbb{N}}\right) \end{aligned} \tag{3.18}$$

for all $t \in \mathbb{N}_0$ (see Proposition 3.1 and Walters [87, Section 4.1]). Moreover, by (i), it holds

$$\sigma\left(\mathbf{X} \circ T^{\circ t}\right) \overset{\mu}{\subset} \sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}\left(T, \left(X_i \circ T^{\circ t}\right)_{i=1}^{n}\right)\right)_{d,n\in\mathbb{N}}\right)$$

for all $t \in \mathbb{N}_0$. Hence

$$\sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right) \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}),$$

which proves the theorem. □

If the considered timing $(E_d)_{d\in\mathbb{N}}$ is not consistent, the inclusion of Equation (3.18) is not guaranteed, and therefore, Theorem 3.3 cannot be applied (compare to Section 3.5 where we give a detailed example).

3.4 The non-ergodic case

The previous results depend on the assumption that T is ergodic, however, in the following, we show that, in order to determine the KS entropy, this assumption can be relaxed.

Theorem 3.4. *Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system where Ω can be embedded into a compact metrizable space such that $\mathcal{A} = \mathcal{B}(\Omega)$. Moreover, let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be a sequence of observables, $\mathcal{R}$ a basic symbolization scheme and $(E_d)_{d\in\mathbb{N}}$ a timing. If*

$$\mathcal{A} \overset{\mu}{\subset} \sigma(\mathbf{X})$$

and Condition (i) to Condition (iii) of Theorem 3.2 are fulfilled for all X_i; $i \in \mathbb{N}$, or

$$\mathcal{A} \overset{\mu}{\subset} \sigma\left(\left(\mathbf{X} \circ T^{\circ t}\right)_{t\in\mathbb{N}_0}\right)$$

and the conditions of Theorem 3.3 are fulfilled, then

$$h_\mu^{\mathrm{KS}}(T) = \lim_{d,n\to\infty} h_\mu(T, \mathcal{P}_{d,n}) = \sup_{d,n\in\mathbb{N}} h_\mu(T, \mathcal{P}_{d,n})$$

with $\mathcal{P}_{d,n} := \mathcal{P}^{\mathcal{R},E_d}\left(T, (X_i)_{i=1}^{n}\right)$ for $d, n \in \mathbb{N}$.

We prove Theorem 3.4 at the end of this section. Essential to the proof is the ergodic decomposition theorem (see Remark 3.4). Recall that, in the ergodic case, the assumptions of Theorem 3.4 ensure that the sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ of finite partitions is generating (see Section 3.3), and hence, by Lemma 2.3, the KS entropy can be determined.

★ **Remark 3.4** (The ergodic decomposition)**.** In the case that $(\Omega, \mathcal{A}, \mu, T)$ is not ergodic, there exists at least one element

$$A^* \in \mathcal{A}^* := \left\{A \in \mathcal{A} \mid T^{\circ -1}(A) = A\right\}$$

for which $0 < \mu(A^*) < 1$. Thus it is straightforward to study the subsystems

$$\left(A^*, \mathcal{A}|_{A^*}, \tfrac{1}{\mu(A^*)}\mu|_{A^*}, T|_{A^*}\right) \text{ and } \left(\Omega \setminus A^*, \mathcal{A}|_{\Omega\setminus A^*}, \tfrac{1}{\mu(\Omega\setminus A^*)}\mu|_{\Omega\setminus A^*}, T|_{\Omega\setminus A^*}\right),$$

where $\mathcal{A}|_{A^*}$ and $\mathcal{A}|_{\Omega\setminus A^*}$ are trace σ-algebras of A^* and $\Omega \setminus A^*$, $\mu|_{A^*}$ and $\mu|_{\Omega\setminus A^*}$ are restrictions of μ to $\mathcal{A}|_{A^*}$ and $\mathcal{A}|_{\Omega\setminus A^*}$ and $T|_{A^*}$ and $T|_{\Omega\setminus A^*}$ are restrictions of T to A^* and $\Omega \setminus A^*$ (see Walters [87] and Appendix B.7). However, these subsystems do not have to be ergodic as well.

♣ **Example 3.3** (An ergodic decomposition of a finite measure-preserving dynamical system)**.** Let Ω be a finite set with $|\Omega| = n \in \mathbb{N}$, $\mathcal{A}$ the power set of Ω and μ be the *uniform measure* on Ω, that is $\mu(\omega) = \frac{1}{n}$ for all $\omega \in \Omega$. Moreover, let $T : \Omega \hookleftarrow$ be onto such that there exists a finite partition $\mathcal{C} = \left\{C^{(1)}, C^{(2)}, \dots, C^{(q)}\right\} \subset \mathcal{A}$ with $2 \leq q \leq n$ and

$$T^{\circ -1}\left(C^{(l)}\right) = C^{(l)} \text{ for all } l \in \{1, 2, \dots q\},$$

and define $n_l := \left|C^{(l)}\right|$. Note that $\sum_{l=1}^{q} \frac{n_l}{n} = 1$. The map T is μ-preserving, however, not ergodic, since for all $l \in \{1, 2, \dots q\}$ it holds

$$0 < \mu\left(C^{(l)}\right) = \tfrac{n_l}{n} < 1.$$

Nevertheless, with

$$\mu_{C^{(l)}}(\omega) := \begin{cases} \frac{1}{n_l} & \text{if } \omega \in C^{(l)}, \\ 0 & \text{otherwise }, \end{cases}$$

the measure μ can be written as a convex combination for all $A \in \mathcal{A}$, i.e.

$$\mu(A) = \sum_{l=1}^{q} \frac{n_l}{n} \mu_{C^{(l)}}(A).$$

Furthermore, each $\mu_{C^{(l)}}$ is ergodic with respect to T. One can easily imagine a similar approach for rational rotations on the unit circle. ♣

The previous example shows that it is possible to decompose a given measure into ergodic components. In general, this is true if Ω can be embedded into some compact metrizable space such that $\mathcal{A} = \mathcal{B}(\Omega)$. For a deeper discussion and for the following theorem, we refer the reader to Einsiedler and Ward [33], Einsiedler et al. [31], Quas [72], Keller and Sinn [52] and the references therein.

Theorem 3.5 (The ergodic decomposition theorem)**.** *Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system where Ω can be embedded into some compact metrizable space such that $\mathcal{A} = \mathcal{B}(\Omega)$. Then there exists a probability space $(\Omega^*, \mathcal{A}^*, \nu)$, and each $\omega^* \in \Omega^*$ can be associated to a probability measure μ_{ω^*} on $(\Omega, \mathcal{A})$ such that the following is valid:*

Ω^* can be embedded into some compact metrizable space such that $\mathcal{A} = \mathcal{B}(\Omega^*)$, for every essentially bounded measurable function $f : \Omega \to \mathbb{R}$ the map

$$\omega^* \in \Omega^* \to \int_{\Omega^*} f \, \mathbf{d}\mu_{\omega^*}$$

is $\mathcal{A}^*$-$\mathcal{B}(\mathbb{R})$-measurable, and for ν-a.e. each $\omega^* \in \Omega^*$, the probability measure μ_{ω^*} is T-invariant and ergodic with respect to T. Moreover, μ can be decomposed such that

$$\mu = \int_{\Omega^*} \mu_{\omega^*} \, \mathbf{d}\nu(\omega^*).$$

As a consequence of the previous theorem, the KS entropy $h_\mu^{\mathrm{KS}}(T)$ and the entropy rate $h_\mu(T, \mathcal{C})$ with respect to any finite partition $\mathcal{C} \subset \mathcal{A}$ of $(\Omega, \mathcal{A})$ can be written as the integral of the entropies with respect to the ergodic decomposition (see Einsiedler et al. [31, Theorem 5.24. and Exercise 5.4.1.]).

Theorem 3.6. *Let $(\Omega, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system. If Theorem 3.5 holds, then the KS entropy $h_\mu^{\mathrm{KS}}(T)$ can be determined by*

$$h_\mu^{\mathrm{KS}}(T) = \int_{\Omega^*} h_{\mu_{\omega^*}}^{\mathrm{KS}}(T) \, \mathbf{d}\nu(\omega^*), \tag{3.19}$$

and the entropy rate by

$$h_\mu(T, \mathcal{C}) = \int_{\Omega^*} h_{\mu_{\omega^*}}(T, \mathcal{C}) \, \mathbf{d}\nu(\omega^*) \tag{3.20}$$

for each finite partition $\mathcal{C} \subset \mathcal{A}$ of $(\Omega, \mathcal{A})$. ★

Proof of Theorem 3.4. Let $(n_j)_{j\in\mathbb{N}}$ and $(d_j)_{j\in\mathbb{N}}$ be strictly increasing sequences of natural numbers, then, by the ergodic decomposition theorem (see Remark 3.4), it holds

$$\begin{aligned}
h_\mu^{\mathrm{KS}}(T) &\overset{(3.19)}{=} \int_{\Omega^*} h_{\mu_{\omega^*}}^{\mathrm{KS}}(T) \, \mathbf{d}\nu(\omega^*) \\
&= \int_{\Omega^*} \lim_{j\to\infty} h_{\mu_{\omega^*}}(T, \mathcal{P}^{\mathcal{R}, E_{d_j}}(T, (X_i)_{i=1}^{n_j})) \, \mathbf{d}\nu(\omega^*) \\
&= \lim_{j\to\infty} \int_{\Omega^*} h_{\mu_{\omega^*}}(T, \mathcal{P}^{\mathcal{R}, E_{d_j}}(T, (X_i)_{i=1}^{n_j})) \, \mathbf{d}\nu(\omega^*) \\
&\overset{(3.20)}{=} \lim_{j\to\infty} h_\mu(T, \mathcal{P}^{\mathcal{R}, E_{d_j}}(T, (X_i)_{i=1}^{n_j})) \\
&= \lim_{j\to\infty} h_\mu(T, \mathcal{P}_{d_j, n_j}).
\end{aligned}$$

We apply Lemma 2.3 in step two and the monotone convergence theorem in step three (see for instance Billingsley [13, Theorem 16.2]). □

3.5 The non-consistent case

In this section, we show that the sequence $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}} := \left(\mathcal{P}^{\mathcal{R},E_d}\left(T,(X_i)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}}$ is in general not generating if only (3.5) is fulfilled and $(E_d)_{d\in\mathbb{N}}$ is not consistent. For this purpose, let $\mathcal{R}$ be the basic symbolization scheme as given in (3.13), i.e.

$$\mathcal{R} = \left\{\left\{(x,y)\in\mathbb{R}^2 \mid y \leq x\right\}, \left\{(x,y)\in\mathbb{R}^2 \mid y > x\right\}\right\},$$

and $(E_d)_{d\in\mathbb{N}}$ be the weak timing, i.e. $E_d = \{(0,t) \mid t\in\{0,1,2,\ldots,d\}\}$ for all $d\in\mathbb{N}$ (compare to Section 3.3.2). Recall, that $(E_d)_{d\in\mathbb{N}}$ is not consistent in general (see Example 3.2), i.e.

$$\sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}\left(T,\left(X_i\circ T^{\circ t}\right)_{i=1}^n\right)\right)_{d,n\in\mathbb{N}}\right) \subset \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}})$$

cannot be guaranteed. Further, note that Theorem 3.2 holds for all observables of the sequence $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ with $g : \mathbb{R} \hookleftarrow$ being the identity map, hence

$$\sigma(\mathbf{X}) \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}})$$

(see Remark 3.3.1). Moreover, if (3.6) holds, then $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ is generating.

Further, consider the dynamical system $([0,1),\mathcal{B}(([0,1)),\lambda,T_a)$ with the irrational rotation T_a, and equip $[0,1)$ with the metric d (see Appendix A.3) given by

$$d(\omega,\omega^*) = \min(|\omega-\omega^*|,\, 1-|\omega-\omega^*|)$$

for all $\omega,\omega^*\in\Omega$. Recall, that the irrational rotation T_a is ergodic and has dense orbits (see Example 2.3). Moreover, T_a is an *isometry*, i.e. the map is injective and *distance preserving*, that is

$$d(T_a(\omega),T_a(\omega^*)) = d(\omega,\omega^*)$$

for all ω, $\omega^*\in\Omega$.

Definition 3.4. Let $[a,b)\subset[0,1)$ with $a<b$. We say an observable $X:\Omega\to\mathbb{R}$ has an *increasing tendency on* $[a,b)$ if $X(a) < X(\omega)$ for all $\omega\in(a,b)$ and a *decreasing tendency on* $[a,b)$ if $X(a) > X(\omega)$ for all $\omega\in(a,b)$, respectively. Moreover, we say an observable $X:\Omega\to\mathbb{R}$ is *nice on* Ω if for every $\omega\in\Omega$ there exists some $\varepsilon>0$ such that either X has an increasing or decreasing tendency on $[\omega,\omega+\varepsilon)$.

From a geometrical point of view, a continuous observable is, for example, nice if its graph has a non-vertical tangent line at each $\omega\in\Omega$. So, all differentiable nowhere constant observables are nice.

Proposition 3.2. *Let X be a continuous and nice observable on the dynamical system $([0,1),\mathcal{B}([0,1)),\lambda,T_a)$ with $\lim_{\omega\to1} X(\omega) = X(0)$. Further, let $\mathcal{R}$ be as given in* (3.13)*, or an refinement of* (3.13)*, and $(E_d)_{d\in\mathbb{N}}$ be the weak timing, or a timing that includes the time pairs of the weak timing. Moreover, let $[\omega_1,\omega_2)\subset[0,1)$ with $\omega_1<\omega_2$, and either X has an increasing or decreasing tendency on $[\omega_1,\omega_2)$, then $(\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}$ separates ω_1 and ω_2 (see Definition 2.9).*

Proof. We give the proof only for the case that X has an increasing tendency on $[\omega_1, \omega_2)$, the decreasing case can be handled in much the same way. We start by defining a set $\Omega_{\omega_2} \subset \Omega$ by

$$\Omega_{\omega_2} := \{\omega \in \Omega \mid X(\omega) = X(\omega_2) \text{ and there exists some } \varepsilon > 0 \text{ such that } X \text{ has a decreasing tendency on } [\omega, \omega + \varepsilon)\}.$$

We claim that $\Omega_{\omega_2} \neq \emptyset$. This is true since $\lim_{\omega \to 1} X(\omega) = X(0)$, X is nice and has an increasing tendency on $[\omega_1, \omega_2)$. Thus X yields intervals of decreasing tendencies, which cover at least the range $[X(\omega_2), X(\omega_1))$. Especially, if there exists some ε such that X has a decreasing tendency on $[\omega_2, \omega_2 + \varepsilon)$, then $\omega_2 \in \Omega_{\omega_2}$. This is entirely in the spirit of—*what goes up must come down.*

Now, for the proof it is natural to distinguish between the case that the intersection $\Omega_{\omega_2} \cap [\omega_2, 1)$ is not empty and the case that it is empty.

In the case that $\Omega_{\omega_2} \cap [\omega_2, 1) \neq \emptyset$, we first choose the element $\widehat{\omega} \in \Omega_{\omega_2}$ such that $\widehat{\omega} \geq \omega_2$ and $d(\widehat{\omega}, \omega_2)$ is minimal for all elements of $\Omega_{\omega_2} \cap [\omega_2, 1)$. Then $X(\omega) \geq X(\omega_2)$ for all $\omega \in [\omega_2, \widehat{\omega})$. Secondly, we choose a $\delta > 0$ such that for a point $t \in \mathbb{N}$ it holds

$$T_a^{\text{ot}}(\omega_2) \in (\widehat{\omega}, \widehat{\omega} + \delta) \text{ and } T_a^{\text{ot}}(\omega_1) \in (\omega_1, \widehat{\omega}) = (\omega_1, \omega_2) \cup [\omega_2, \widehat{\omega}).$$

Clearly, $\delta < d(\omega_1, \omega_2)$. Note that, for all $\omega \in (\omega_1, \omega_2)$, it holds $X(\omega) > X(\omega_1)$, and on $[\omega_2, \widehat{\omega})$ the observable X is bounded from below by $X(\omega_2)$. So, $X(\omega_1) < X(T_a^{\text{ot}}(\omega_1))$ and $X(\omega_2) > X(T_a^{\text{ot}}(\omega_2))$.

In the case that $\Omega_{\omega_2} \cap [\omega_2, 1) = \emptyset$, it holds $\omega < \omega_2$ for all $\omega \in \Omega_{\omega_2}$. We choose the element $\widehat{\omega} = \min(\Omega_{\omega_2})$. Then $X(\omega) \geq X(\omega_2)$ for all $\omega \in [0, \widehat{\omega}) \cup [\omega_2, 1)$. Moreover, we choose a $\delta > 0$ such that for a point $t \in \mathbb{N}$ it holds

$$T_a^{\text{ot}}(\omega_2) \in (\widehat{\omega}, \widehat{\omega} + \delta) \text{ and } T_a^{\text{ot}}(\omega_1) \in [0, \widehat{\omega}) \cup (\omega_1, \omega_2) \cup [\omega_2, 1).$$

Note as above that $\delta < d(\omega_1, \omega_2)$, and, for all $\omega \in (\omega_1, \omega_2)$, it holds $X(\omega) > X(\omega_1)$, and on $[0, \widehat{\omega}) \cup [\omega_2, 1)$ the observable X is bounded from below by $X(\omega_2)$. So we have again, $X(\omega_1) < X(T_a^{\text{ot}}(\omega_1))$ and $X(\omega_2) > X(T_a^{\text{ot}}(\omega_2))$.

This completes the proof since for each time point $t \in \mathbb{N}$ there exists some $d \in \mathbb{N}$ such that $(0, t) \in E_d$. □

Up to now, we have omitted observables which are locally constant. In fact, these constant parts play a key role to show the difference between a symbolic-based analysis technique which is based on (3.13) and the weak timing and, for example, an approach which entails the same timing but the basic symbolization scheme given by

$$\mathcal{R} := \big\{ \{(x, y) \in \mathbb{R}^2 \mid y < x\}, \{(x, y) \in \mathbb{R}^2 \mid y > x\}, \{(x, y) \in \mathbb{R}^2 \mid x = y\} \big\} \tag{3.21}$$

(see Example 3.4).

♣ **Example 3.4** (A nice observable). By Definition 3.4 the observable as given in (2.14), i.e.

$$Y(\omega) = \begin{cases} 0 & \text{if } \omega \in \left[0, \frac{1}{2}\right), \\ \frac{1}{2} - \omega & \text{if } \omega \in \left[\frac{1}{2}, \frac{3}{4}\right), \\ \omega - 1 & \text{if } \omega \in \left[\frac{3}{4}, 1\right), \end{cases}$$

is nice on $\left[\frac{1}{2}, 1\right)$ (see Figure 2.6 and Figure 3.6).

(i) We first examine the basic symbolization scheme (3.13) together with the weak timing. We claim that $(\mathcal{P}^{\mathcal{R},E_d}(T,Y))_{d\in\mathbb{N}}$ separates any two states $\omega_1 \neq \omega_2$ of the interval $\left[\frac{1}{2}, 1\right)$, however, fails to separate any two different elements of the interval $\left[0, \frac{1}{2}\right)$.

In order to show the former case, let $\omega_1 < \omega_2$. If $\omega_1 < \frac{3}{4}$ and $\omega_2 \geq \frac{3}{4}$, pick a point $t \in \mathbb{N}$ such that $T_a^{\circ t}(\omega_1)$ is close to and greater than ω_1 and $T_a^{\circ t}(\omega_2)$ is close to and greater than ω_2. Then

$$Y(\omega_1) \geq Y(T_a^{\circ t}(\omega_1)) \text{ and } Y(\omega_2) < Y(T_a^{\circ t}(\omega_2)).$$

If $\omega_1 \neq \omega_2$ are elements of $\left[\frac{1}{2}, \frac{3}{4}\right)$ or of $\left[\frac{3}{4}, 1\right)$, the separation follows by applying Proposition 3.2. However, there is no separation of the elements $\omega \in \left[0, \frac{1}{2}\right)$, since for any $t \in \mathbb{N}$ it holds

$$\{\omega \in \Omega \mid Y(\omega) \geq Y(T_a^{\circ t}(\omega))\} = [0, \tfrac{1}{2}) \cup \{\omega \in [\tfrac{1}{2}, 1) \mid Y(\omega) \geq Y(T_a^{\circ t}(\omega))\}.$$

(ii) We now turn to the case where we consider the same timing but a different partition $\mathcal{R}$, namely the one given in (3.21). In this setting, $(\mathcal{P}^{\mathcal{R},E_d}(T,Y))_{d\in\mathbb{N}}$ separates any two states $\omega_1 \neq \omega_2$ of Ω, which can be seen as follows.
The separation of points $\omega \in \left[\frac{1}{2}, 1\right)$ follows from the deliberations in (i). Moreover, in order to separate the states $\omega_1 < \omega_2$ of the interval $\left[0, \frac{1}{2}\right)$, we pick a point $t \in \mathbb{N}$ such that $T_a^{\circ t}(\omega_1) \in \left[0, \frac{1}{2}\right)$ and $T_a^{\circ t}(\omega_2) \in \left(\frac{1}{2}, 1\right)$. Then $Y(\omega_1) = Y(T_a^{\circ s}(\omega_1))$ and either $Y(\omega_2) > Y(T_a^{\circ s}(\omega_2))$ or $Y(\omega_2) < Y(T_a^{\circ s}(\omega_2))$. ♣

Note that if an observable X is constant on an interval $[a,b) \subset [0,1)$ with $a < b$, then $(\mathcal{P}^{\mathcal{R},E_d}(T,X))_{d\in\mathbb{N}}$ with $\mathcal{R}$ as given in (3.21) separates any two different states $\omega_1, \omega_2 \in [a,b)$ as long as there exists at least one interval $[c,d) \subset [0,1)$ with $c < d$ where X has an increasing or a decreasing tendency.

Also, it is obvious that there exists neither a separation with respect to the measuring process $(X \circ T^{\circ t})_{t\in\mathbb{N}_0}$ nor to $(\mathcal{P}^{\mathcal{R},E_d}(T,X))_{d\in\mathbb{N}}$ with $\mathcal{R}$ and $(E_d)_{d\in\mathbb{N}}$ freely chosen if X is a constant observable on $[0,1)$.

3.6 Remarks

In this chapter, we discussed symbolic routes to the KS entropy by studying different symbolic schemes at once (see Figure 3.7 for an overview). We included commonly known approaches (see Section 3.2.1, Section 3.2.2 and Section 3.5). At first glance our approach seems rather involved. However, through it, we show that under relatively weak assumptions, symbolic schemes that regard a dependency between two measured values, in contrast to, for instance, threshold crossings, provide a route to the KS entropy. This is, in particular, engaging if the KS entropy cannot directly be characterized.

Both from the theoretical and the practical point of view, it is interesting to study the speed of convergence in (3.2), for instance, in order to generally compare the efficiency of basic symbolization schemes and to estimate entropies, respectively. One of the main objectives is to find a symbolic-based analysis technique that preserves, in finitely many steps, as much information of the original system as possible.

Moreover, a careful weighing of the advantages and disadvantages is necessary; including questions such as which schemes allow simple interpretations and efficient algorithms. In the next chapter, we give an overview over the topic, however, entrust details to further studies.

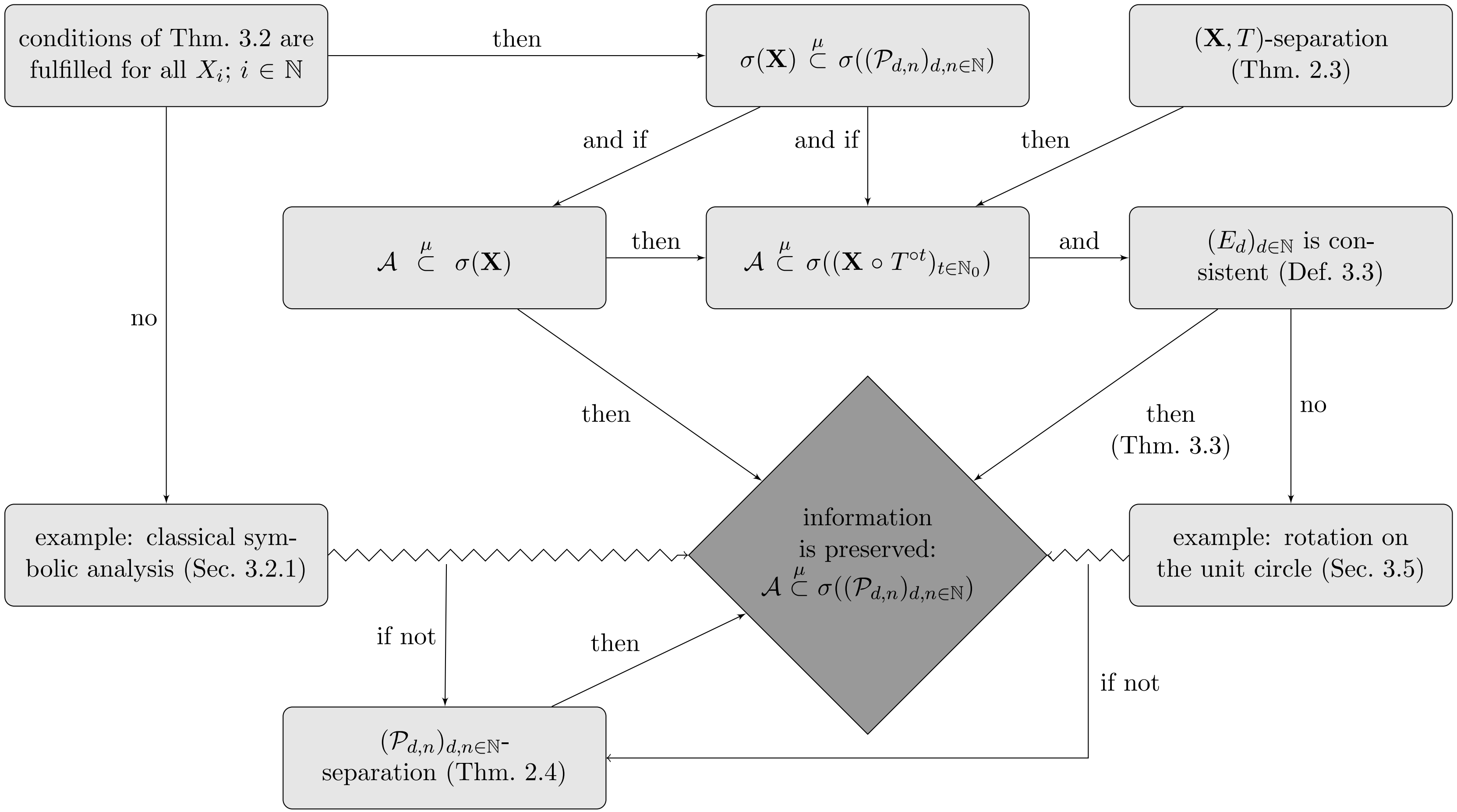


Figure 3.7: Summary of Chapter 3 with a focus on our results of Section 3.3.2 and Section 3.5. Let $\mathcal{R}$ be a basic symbolization scheme and $(E_d)_{d\in\mathbb{N}}$ be a timing (see Section 3.2). The labeled arrows represent the important relations and the jagged line indicates that $\mathcal{A} \overset{\mu}{\subset} \sigma((\mathcal{P}_{d,n})_{d,n\in\mathbb{N}}) := \sigma\left(\left(\mathcal{P}^{\mathcal{R},E_d}(T,(X_i)_{i=1}^n)\right)_{d,n\in\mathbb{N}}\right)$ cannot be guaranteed.

Chapter 4

Entropy measures based on symbolic schemes

From our results in Chapter 3 we know that symbolic-based analysis techniques with symbolic schemes $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ exploit the information of the original system with varying degrees of success depending on the properties of $\mathcal{R}$ and $(E_d)_{d\in\mathbb{N}}$, respectively.

In this chapter, we extend our results by evaluating the information content of symbolic-based analysis techniques. In order to do this, we study and apply entropy measures that are based on a basic symbolization scheme $\mathcal{R}$ and a timing $(E_d)_{d\in\mathbb{N}}$. This chapter is organized as follows. Firstly, we examine the relationship of the entropy measures based on $\mathcal{R}$ and $(E_d)_{d\in\mathbb{N}}$ to the KS entropy. Secondly, we estimate different entropy measures and study their asymptotic behavior in dependence on the orbit length. Note that some of the results presented in this chapter are published in Keller et al. [50].

4.1 Relationship to the Kolmogorov-Sinai entropy

There are many possibilities to define an entropy measure based on a basic symbolization scheme $\mathcal{R}$ and a timing $(E_d)_{d\in\mathbb{N}}$ (compare to Section 2.2.3). Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be a sequence of observables. Here, we study entropy measures $h_\mu^{\mathcal{R},E_d}(T, \mathbf{X})$ defined by

$$h_\mu^{\mathcal{R},E_d}(T, \mathbf{X}) := \lim_{n\to\infty} \limsup_{d\to\infty} \frac{1}{d} H_\mu\left(\mathcal{P}^{\mathcal{R},E_d}(T, (X_i)_{i=1}^n)\right).$$

We consider $h_\mu^{\mathcal{R},E_d}(T, \mathbf{X})$ since it includes the popular Permutation entropy that is based on the ordinal idea and has a close relationship to the KS entropy $h_\mu^{\mathrm{KS}}(T)$ (see Bandt and Pompe [11], Bandt et al. [10] and Example 4.1). It is an open question for which basic symbolization schemes $\mathcal{R}$ and timings $(E_d)_{d\in\mathbb{N}}$ the entropy $h_\mu^{\mathcal{R},E_d}(T, \mathbf{X})$ coincides with the KS entropy; also for the Permutation entropy this question is not yet answered (see for instance Keller et al. [55, 49] and Antoniouk et al. [9]). However, the following general statements provide further insight into the relationship of $h_\mu^{\mathcal{R},E_d}(T, \mathbf{X})$ and the KS entropy.

Firstly, let $\mathcal{R}$ and $\mathcal{R}'$ be two different basic symbolization schemes and $(E_d)_{d\in\mathbb{N}}$ and $(E'_d)_{d\in\mathbb{N}}$ be two different timings. If $\mathcal{P}^{\mathcal{R}',E'_d}(T, (X_i)_{i=1}^n)$ is a refinement of

$\mathcal{P}^{\mathcal{R},E_d}(T,(X_i)_{i=1}^n)$ for all $d \in \mathbb{N}$ and $n \in \mathbb{N}$, then

$$H_\mu(\mathcal{P}^{\mathcal{R}',E'_d}(T,(X_i)_{i=1}^n)) \geq H_\mu(\mathcal{P}^{\mathcal{R},E_d}(T,(X_i)_{i=1}^n))$$

for all $d \in \mathbb{N}$ and $n \in \mathbb{N}$, which gives

$$h_\mu^{\mathcal{R}',E'_d}(T,\mathbf{X}) \geq h_\mu^{\mathcal{R},E_d}(T,\mathbf{X})$$

(see Remark 2.5). Hence, in particular, if

$$h_\mu^{\mathrm{KS}}(T) = h_\mu^{\mathcal{R},E_d}(T,\mathbf{X}), \text{ then } h_\mu^{\mathrm{KS}}(T) = h_\mu^{\mathcal{R}',E'_d}(T,\mathbf{X}).$$

Secondly, Keller et al. show in [55] that

$$\limsup_{d\to\infty} \frac{1}{d} H_\mu(\mathcal{C}_d) \geq \lim_{d\to\infty} h_\mu(T,\mathcal{C}_d)$$

if $(\mathcal{C}_d)_{d\in\mathbb{N}}$ is an increasing sequence of finite partitions and $\mathcal{C}_{d+t-1}$ is finer than $(\mathcal{C}_d)_t$ for all $d \in \mathbb{N}$ and $t \in \mathbb{N}$ with $t > 2$. Recall that $((\mathcal{C}_d)_t)_{t\in\mathbb{N}}$ is the partition sequence of $\mathcal{C}_d$ under T (see Section 2.2.1). In terms of our formalization of symbolic-based analysis techniques the result of Keller et al. [55] reads as follows.

Proposition 4.1. *Let X be an observable, $\mathcal{R}$ a basic symbolization scheme, $(E_d)_{d\in\mathbb{N}}$ a consistent timing and $\mathcal{P}_d := \mathcal{P}^{\mathcal{R},E_d}(T,X)$. Then*

$$h_\mu^{\mathcal{R},E_d}(T,X) := \limsup_{d\to\infty} \frac{1}{d} H_\mu(\mathcal{P}_d) \geq \lim_{d\to\infty} h_\mu(T,\mathcal{P}_d). \tag{4.1}$$

In particular, if $h_\mu^{\mathrm{KS}}(T) = \lim_{d\to\infty} h_\mu(T,\mathcal{P}_d)$, then

$$h_\mu^{\mathcal{R},E_d}(T,X) \geq h_\mu^{\mathrm{KS}}(T).$$

We repeat the proof of Keller et al. [55, Lemma 9] for the sake of completeness in the following.

Proof of Proposition 4.1. In order to shorten notation, let $h := \lim\limits_{d\to\infty} h_\mu(T,\mathcal{P}_d)$. Our proof starts with the following two observations. Firstly, since $(E_d)_{d\in\mathbb{N}}$ is consistent, it holds that

$$(\mathcal{P}_d)_t := \bigvee_{s=0}^{t-1} \mathcal{P}^{\mathcal{R},E_d}(T, X \circ T^{\circ s}) \prec \mathcal{P}_{d+t-1}$$

for all $d \in \mathbb{N}$ and $t \in \mathbb{N}$ (see Section 3.3.2), which gives

$$H_\mu(\mathcal{P}_{d+t-1}) \geq H_\mu((\mathcal{P}_d)_t)$$

for all $d \in \mathbb{N}$ and $t \in \mathbb{N}$. Secondly, if $h > c$ for some $c > 0$, then there exists some $d \in \mathbb{N}$ and some $t_d \in \mathbb{N}$ with $\frac{1}{t} H_\mu((\mathcal{P}_d)_t) > c$ for all $t \geq t_d$, since

$$h = \lim_{d\to\infty} h_\mu(T,\mathcal{P}_d) = \lim_{d\to\infty}\lim_{t\to\infty} \frac{1}{t} H_\mu((\mathcal{P}_d)_t) = \lim_{d\to\infty}\lim_{t\to\infty} \frac{1}{t+d-1} H_\mu((\mathcal{P}_d)_t).$$

By the definition of the Shannon entropy, it is evident that (4.1) is true if $h = 0$. Therefore, assume that $h > 0$. The main idea of the proof is to show that $h_\mu^{\mathcal{R},E_d}(T,X) \geq h - \varepsilon$ for every $\varepsilon > 0$ which gives (4.1). For this purpose, regard $\varepsilon > 0$ as fixed, and choose $c > 1$ such that $h > c(h - \varepsilon)$ still holds. By our previous deliberation, there exists some $d \in \mathbb{N}$ and some $t_d \in \mathbb{N}$ with $\frac{1}{t} H_\mu\left((\mathcal{P}_d)_t\right) > c(h-\varepsilon)$ for all $t \geq t_d$. It follows that

$$\begin{aligned}\frac{1}{d+t-1} H_\mu\left(\mathcal{P}_{d+t-1}\right) &\geq \frac{1}{d+t-1} H_\mu\left((\mathcal{P}_d)_t\right) \\ &\geq \frac{1}{(c-1)t+t-1} H_\mu\left((\mathcal{P}_d)_t\right) > \frac{1}{ct} H_\mu\left((\mathcal{P}_d)_t\right) > h - \varepsilon\end{aligned}$$

for all $t \geq \max\left\{t_d, \frac{d}{c-1}\right\}$. Hence,

$$h_\mu^{\mathcal{R},E_d}(T,X) = \limsup_{d\to\infty} \frac{1}{d+t-1} H_\mu(\mathcal{P}_{d+t-1}) \geq h - \varepsilon,$$

which completes the proof. □

In the next corollary, we study the case of infinitely many observables.

Corollary 4.1. *Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be a sequence of observables, $\mathcal{R}$ a basic symbolization scheme, $(E_d)_{d\in\mathbb{N}}$ a consistent timing and $\mathcal{P}_{d,n} := \mathcal{P}^{\mathcal{R},E_d}(T,(X_i)_{i=1}^n)$. Then*

$$\lim_{n\to\infty} \limsup_{d\to\infty} \frac{H(\mathcal{P}_{d,n})}{d} \geq \lim_{d,n\to\infty} h_\mu(T,\mathcal{P}_{d,n}).$$

In particular, if $h_\mu^{\mathrm{KS}}(T) = \lim_{d,n\to\infty} h_\mu\left(T,\mathcal{P}_{d,n}\right)$, then

$$h_\mu^{\mathcal{R},E_d}(T,\mathbf{X}) \geq h_\mu^{\mathrm{KS}}(T).$$

Proof. This follows directly from Proposition 4.1 (compare to Keller et al. [49]). □

♣ **Example 4.1** (The Permutation entropy)**.** The ordinal symbolic approach provides interesting entropy measures that are used for data analysis in various applications (see Amigó et al. [7], Zanin et al. [88], Keller et al. [50], the special topic [6], the special issue [65] and the references given therein). A popular entropy measure that is based on the strong ordinal approach is the *Permutation entropy* $h_\mu^{\mathcal{R},E_d}(T,\mathbf{X})$, i.e.

$$\mathcal{R} = \left\{\left\{(x,y) \in \mathbb{R}^2 \mid y \leq x\right\}, \left\{(x,y) \in \mathbb{R}^2 \mid y > x\right\}\right\} \tag{4.2}$$

and

$$E_d = \{(s,t) \mid s,t \in \{0,1,2,\ldots,d\} \text{ with } s < t\}\,;\; d \in \mathbb{N}$$

(see Section 3.2.2). The definition goes back to Bandt and Pompe [11] and Bandt et al. [10], and is justified by their discovery that the Permutation entropy has a close relationship to the KS entropy. In fact, the authors show, in the case that Ω is a real interval, $\mathcal{A} = \mathcal{B}(\Omega)$, $T : \Omega \hookleftarrow$ is a piecewise continuous and monotone map and $X(\omega) = \omega$ for all $\omega \in \Omega$, that

$$h_\mu^{\mathrm{KS}}(T) = h_\mu^{\mathcal{R},E_d}(T, \mathbf{X}). \tag{4.3}$$

Therefore, in the same setting, if a symbolic-based analysis technique is considered that entails the full timing and a refinement of (4.2), then (4.3) is also fulfilled. However, as we already mentioned, it is an open question whether the KS entropy $h_\mu^{\mathrm{KS}}(T)$ and the Permutation entropy $h_\mu^{\mathcal{R},E_d}(T, \mathbf{X})$ generally coincide.

Note that there exist many modifications of the Permutation entropy, including conditional variants and variants using some additional metric information. Moreover, sometimes the Shannon entropy is replaced by other entropies. For a deeper discussion about different variants of the Permutation entropy, we refer the reader to Keller et al. [50] and the references given therein. ♣

4.2 Empirical entropy measures in dependence on the orbit length

The results of the previous chapters provide an interesting insight into the theory behind symbolic-based techniques in time series and system analysis. In closing this thesis, we compare different basic symbolization schemes $\mathcal{R}$ and timings $(E_d)_{d\in\mathbb{N}}$ by analyzing simulated orbits of ergodic systems. In difference to Chapter 1 where we emphasized possible applications of symbolic-based analysis techniques, for instance, in order to detect complexity changes in a time series and to classify or to cluster data sets, we now study the asymptotic behavior of symbolic-based analysis techniques and their potential to approximate the KS entropy. We also discuss statistical problems that are common for any kind of time series analysis. In this section, we proceed as follows.

- In Section 4.2.1, we compare the entropies

  $$H_\mu((\mathcal{P}_d)_{t+1}) - H_\mu((\mathcal{P}_d)_t) \text{ and } \frac{1}{d} H_\mu((\mathcal{P}_d)_t)$$

  for different word lengths $t \in \mathbb{N}$ and ordinal orders $d \in \mathbb{N}$ by applying the strong ordinal approach to a simulated orbit of the logistic map T_4. In particular, this section is interesting in order to understand Figure 2.5**(b)** from a practical point of view, i.e. approximating the KS entropy by a conditional entropy.

- Section 4.2.2 is devoted to the study of different basic symbolization schemes. We compare threshold crossings with the strong ordinal idea by analyzing a simulated orbit of the logistic map T_4 in dependence on the orbit length.

- In difference to Section 4.2.2, we consider, in Section 4.2.3, the basic symbolization scheme entailed by the strong ordinal approach but with varying timings. We analyze a simulated orbit of Arnold's cat map (see for instance Andries et al. [8] and Remark 4.3) in dependence on the orbit length.

Note that, in the following, we estimate approximations of the KS entropy since we fix $d \in \mathbb{N}$ and $t \in \mathbb{N}$. These estimates are consistent by Birkhoff's ergodic theorem (see Section 2.1) and show similar behavior in dependence on the orbit length. All curves level off for long orbits, i.e. the relative distribution of symbol words is stable. However, for a short orbit length the estimates are quite bad since not all symbol words, that are substantial to understand the underlying dynamics, can be observed sufficiently. These *undersampling* problems increase the interest in symbolic-based analysis techniques that entail a small number of symbols for large values of $d \in \mathbb{N}$ and $t \in \mathbb{N}$ since, in general, it is a compromise between the computational costs and the information loss (see Li and Ray [63]).

Our algorithms to decode simulated orbits into sequences of symbols are implemented straightforwardly and not efficiently, i.e. we neglect statements about time and memory. For a deeper discussion of efficient algorithms that transform a time series into a sequence of ordinal patterns, and to compute the empirical Permutation entropy of an ordinal order $d \in \mathbb{N}$ and variants of it, consult Riedl et al. [73], Unakafova and Keller [84] and Unakafova [85]. Here, we restrict ourselves mainly to a description of the results, and touch on only a few aspects of the algorithms applied. For a thorough treatment, we refer the reader to Kurths et al. [60] and Keller et al. [53].

★ **Remark 4.1** (Simulated data)**.** In this section, we assume that measurements directly provide the state of the considered system (see Remark 2.2), and hence we analyze orbits of different dynamics T directly.

We are aware that numerical simulations actually require an analysis of stability and round-off errors, in particular, if dynamics T on a metric space (Ω, d) are considered that are sensitive to initial conditions, i.e. there exists some $\delta > 0$ such that for each $\omega \in \Omega$ and each $\varepsilon > 0$ there exists some $\omega' \in \Omega$ and some $t \in \mathbb{N}_0$ with

$$d(\omega, \omega') < \varepsilon \text{ and } d\left(T^{\circ t}(\omega), T^{\circ t}(\omega')\right) > \delta.$$

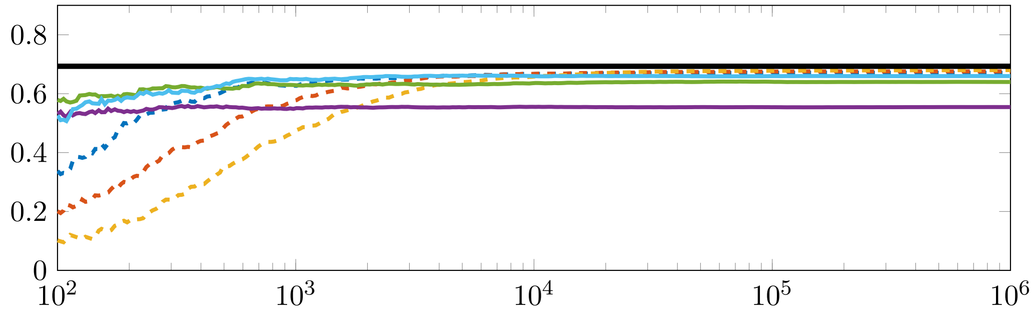
Since we study, in this section, different symbolic schemes, and consider systems with known KS entropies, we neglect the study of quality or reliability of our simulated data in the following. For a deeper discussion and further ideas on the topic see, for instance, Hammel [41], Peitgen et al. [71], Sprott [77] and Buzzi [19].

Moreover, note that we choose the non-causal perspective in our algorithms in order to be consistent with the preceding chapters and for convenience of simpler notation. However, in truth, just glances at the past are possible, and therefore, most implementations are the other way around (compare to Unakafova and Keller [84] and Unakafova [85]). ★

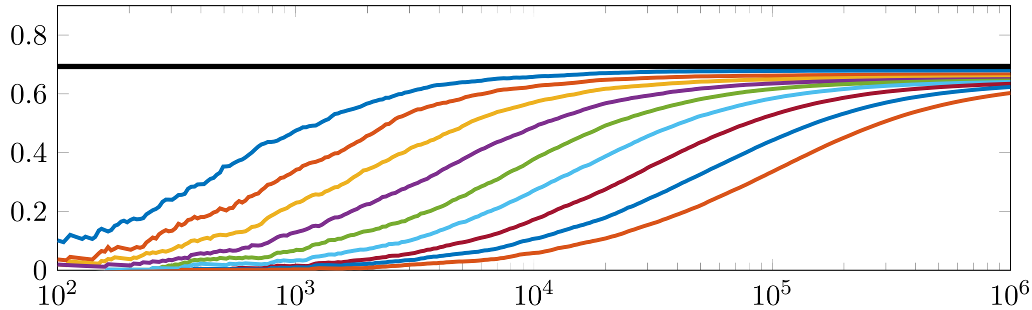
4.2.1 Empirical conditional entropy measures

Let $\mathcal{C} \subset \mathcal{A}$ be a finite partition of Ω. In Section 2.2, we introduced in (2.4) and (2.7) two equivalent definitions of the entropy rate $h_\mu(T, \mathcal{C})$, whereby the conditional entropy is less or equal to the Shannon entropy divided by $t \in \mathbb{N}$, i.e.

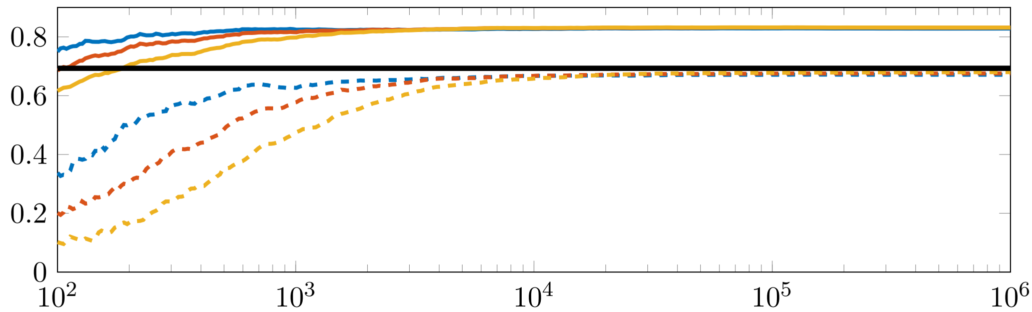
$$H_\mu((\mathcal{C})_{t+1}) - H_\mu((\mathcal{C})_t) \leq \frac{1}{t} H_\mu((\mathcal{C})_t).$$



(a) Estimates of the ordinal Conditional entropy for $d \in \{2, 3, \dots, 7\}$ (purple, green, light blue, blue dotted, red dotted, yellow dotted).



(b) Estimates of $H_\mu((\mathcal{P}_7)_{t+1}) - H_\mu((\mathcal{P}_7)_t)$ for $t \in \{1, 2, \dots, 9\}$ (from left to right) based on the strong ordinal approach.



(c) Estimates of the ordinal Conditional entropy (dotted lines) and of the Permutation entropy (blue, red, yellow) for $d \in \{5, 6, 7\}$.

Figure 4.1: Different estimates of the KS entropy (black line: $\ln(2)$) of the logistic map $T_4(\omega) = 4\omega(1 - \omega)$ with $\omega \in [0, 1)$ in dependence on the orbit length.

The inequality is, in particular, interesting for applications, and motivates the ordinal Conditional entropy (see Unakafov and Keller [83] and Unakafov [82]) as an alternative to the Permutation entropy. These entropies are defined as follows.

Let $\mathcal{R}$ be the basic symbolization scheme as given in (4.2) and $(E_d)_{d \in \mathbb{N}}$ be the full timing (see Example 4.1). We call $\frac{1}{d} H_\mu(\mathcal{P}_d)$ and $H_\mu((\mathcal{P}_d)_2) - H_\mu((\mathcal{P}_d)_1)$, respectively, the *Permutation entropy* and the *ordinal Conditional entropy* for some order $d \in \mathbb{N}$. In fact, the ordinal Conditional entropy often shows better results than the Permutation entropy in practice, as we also underline in the following (compare to the results of Unakafov and Keller [83] and Unakafov [82]). In order to give a perspective on the matter, as a first step we simulated an orbit $(x_s)_{s=0}^{\mathcal{T}}$ with $\mathcal{T} \in \mathbb{N}$ of the logistic map T_4 (see Example 2.1), i.e.

$$x_{s+1} = 4x_s(1 - x_s)$$

for all $s \in \{0, 1, \ldots, \mathcal{T} - 1\}$ and some randomly chosen x_0. As a second step, we decoded $(x_s)_{s=0}^{\mathcal{T}}$ into a sequence of symbols by applying the strong ordinal approach for different orders $d \in \mathbb{N}$. Subsequently, we fixed a word length $t \in \mathbb{N}$, and naively estimated $\frac{1}{d}H_\mu(\mathcal{P}_d)$ and $H_\mu((\mathcal{P}_d)_{t+1}) - H_\mu((\mathcal{P}_d)_t)$ in dependence on the orbit length $\mathcal{T}$ by replacing the probabilities by relative frequencies of symbol word occurrences.

Note that the logistic map T_4 is ergodic with respect to μ_4 (see for instance Chan and Tong [22]). Further, the KS entropy of T_4 is $\ln(2)$ (see for instance Grassberger [37] and Unakafov [82] and the references given therein) which we indicate by the horizontal black line in Figure 4.1. Logistic maps are often considered in order to test and to compare ordinal complexity measures (see for instance Amigó [5], the special topic [6], Amigó et al. [7] and the references given therein). Moreover, note that the following elaboration, in particular Figure 4.1, are derivatives from Keller et al. [50].

As we already remarked in Section 2.2.3, asymptotics can be very slow, in particular, if the considered system is complex. In such a case, one should pick high values of the order $d \in \mathbb{N}$ and the word length $t \in \mathbb{N}$. However, it is not surprising that in this case one needs long-term measurements in order to obtain a reliable estimation, as shown in Figure 4.1.

Estimates of $H_\mu((\mathcal{P}_d)_2) - H_\mu((\mathcal{P}_d)_1)$ for orders $d \in \{2, 3, \ldots, 7\}$ in dependence on the orbit length $\mathcal{T}$ between 10^2 and 10^6 are presented in Figure 4.1**(a)**. It seems that the approximation of the ordinal Conditional entropy, or rather the KS entropy, is reasonable for $d \in \{5, 6, 7\}$ once the orbit is long enough. However, in general, the higher the order $d \in \mathbb{N}$, the more time is needed to stabilize the estimation. The same problem is also apparent in Figure 4.1**(b)**. Here, we present naive estimates of $H_\mu((\mathcal{P}_7)_{t+1}) - H_\mu((\mathcal{P}_7)_t)$ for $t \in \{1, 2, \ldots, 9\}$. Observe, that the approximation of the KS entropy is still rather bad if, for instance, $t \in \{7, 8, 9\}$ and the orbit length is $\mathcal{T} = 10^5$.

Figure 4.1**(c)** shows a comparison of the naive estimates of

$$\frac{1}{d}H_\mu(\mathcal{P}_d) \text{ and } H_\mu((\mathcal{P}_d)_2) - H_\mu((\mathcal{P}_d)_1)$$

for orders $d \in \{5, 6, 7\}$. By the result of Bandt et al. [10], estimates of the Permutation entropy for very high ordinal orders $d \in \mathbb{N}$ must be close to the KS entropy. However, for a fixed order $d \in \mathbb{N}$, the ordinal Conditional entropy seems to be a better choice than the Permutation entropy. This coincides with the results of Unakafov and Keller [83] and Unakafov [82].

★ **Remark 4.2** (Estimating the Shannon entropy)**.** Estimating the Shannon entropy is not as straightforward as it may seem. In fact, for each data set, one has to find the most suitable compromise between bias reduction and statistical error. For a good overview and deeper discussion, we refer to Paninski [69] and Bonachela et al. [16] and the references therein. The literature provides many proposals and remarks on how to correct the naive estimator, or whether a different estimator is more reasonable (see for instance Roulston [74], Grassberger [38], Schürmann [76] and the references given therein). ★

4.2.2 Different basic symbolization schemes

In this section, we compare different basic symbolization schemes. In particular, we highlight the problem of threshold crossings if a partition is considered that is not generating under the dynamics. We decoded an orbit of the logistic map T_4 into a sequence of symbols. Subsequently, we fixed a word length $t \in \mathbb{N}$, and naively estimated the difference $H_\mu((\mathcal{P}_d)_{t+1}) - H_\mu((\mathcal{P}_d)_t)$ by replacing the probabilities by relative frequencies of symbol word occurrences.

The results are shown in Figure 4.2**(a)** for different word lengths $t \in \mathbb{N}$ and different basic symbolization schemes in dependence on the orbit length $\mathcal{T}$ between 10^2 and 10^6. The blue, red and yellow curves are due to threshold crossings with the partition $\{[0, 0.5), [0.5, 1)\}$ that is generating under T_4 and with the misplaced partitions $\{[0, 0.9), [0.9, 1)\}$, $\{[0, 0.4), [0.4, 1)\}$ for $d = 1$ and $t = 8$, i.e. they are non-generating under T_4 (compare to Example 3.1 where we considered the full tent map). Moreover, the purple, green and light blue curves are due to threshold crossings also non-generating under T_4 with

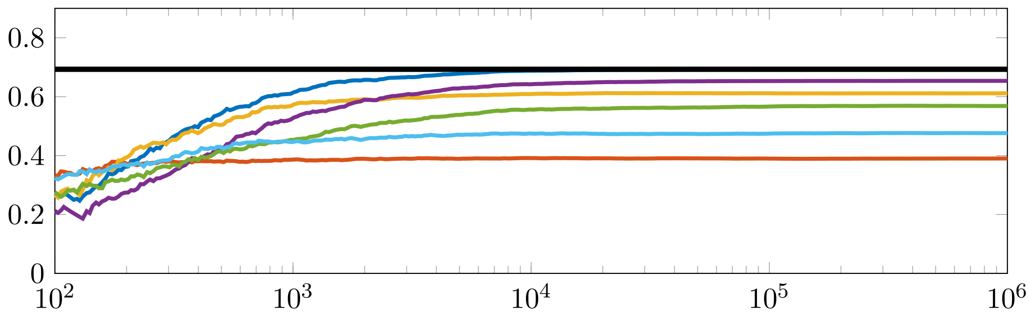
- $\{[0, 0.45), [0.45, 0.55), [0.55, 1)\}$,
- $\{[0, 0.05), [0.05, 0.1), [0.1, 0.15), [0.15, 0.2), [0.2, 1)\}$, and
- $\{[0, 0.05), [0.05, 0.95), [0.95, 1)\}$.

The curves underline not only our assertions of Section 4.2.1, i.e. the parameters have to be chosen reasonably in regard to the orbit length, but also confirm that the threshold crossings technique based on the partition $\{[0, 0.5), [0.5, 1)\}$ delivers, in contrast to misplaced partitions, the best results. Unfortunately, as we discussed in detail in Chapter 2 and Chapter 3, it is very unrealistic, if not impossible, to choose such a partition for a threshold crossings technique if the dynamics are unknown.

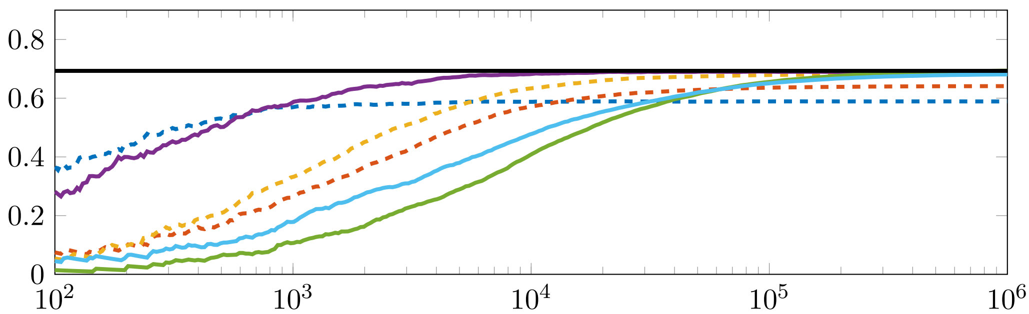
In Figure 4.2**(b)**, the dashed curves are due to the strong ordinal approach for $(d, t) \in \{(3, 4), (6, 5), (8, 2)\}$ (blue, red and yellow). By contrast, we also applied a symbolic-based analysis technique that entails the full timing and the following basic symbolization scheme for some $\varepsilon > 0$:

$$\begin{aligned} \mathcal{R} = \{ \, &\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x - \varepsilon \leq y \leq x + \varepsilon\} \, , \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y > x + \varepsilon\} \\ &\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y < x - \varepsilon\} \, \} \, . \end{aligned} \tag{4.4}$$

The results are shown in Figure 4.2**(b)**, whereby the purple curve is due to the parameters $\varepsilon = 0.05$, $d = 3$ and $t = 4$, the green curve is due to $\varepsilon = 0.05$, $d = 6$ and $t = 5$ and the light blue curve is based on $\varepsilon = 0.01$, $d = 6$ and $t = 5$. At this point, we do not want to discuss which choices of ε are reasonable, but want to call attention to the diversity of possible complexity measures brought together by our unifying way to formalize symbolic schemes. Note that, for a fixed $d \in \mathbb{N}$ and $t \in \mathbb{N}$, symbolic schemes with (4.4) entail, in general, more symbols and symbol words than symbolic schemes based on the strong ordinal approach. Thus they have



(a) The classical symbolic scheme with different initial partitions for $d = 1$ and $t = 8$ (for more details on the thresholds, we refer the reader to Section 4.2.2).



(b) Strong ordinal approach for $(d, t) \in \{(3, 4), (6, 5), (8, 2)\}$ (dashed curves: blue, red and yellow) and a symbolic-based technique where $(E_d)_{d\in\mathbb{N}}$ is the full timing and $\mathcal{R}$ is characterized by some ε-tube for $(d, t, \varepsilon) \in \{(3, 4, 0.05), (6, 5, 0.05), (6, 5, 0.01)\}$ (solid curves: purple, green, light blue).

Figure 4.2: Different estimates of the KS entropy (black line: $\ln(2)$) of the logistic map T_4 (see Example 2.1) in dependence on the orbit length by naively estimating $H_\mu((\mathcal{P}_d)_{t+1}) - H_\mu((\mathcal{P}_d)_t)$.

a greater potential of disclosing encapsulated information. However, the larger the number of symbols or symbol words are, the greater the risk of undersampling. Both phenomena, the accurate evaluation of encapsulated information and the problem of undersampling, are observable in Figure 4.2**(b)**.

4.2.3 Different timings

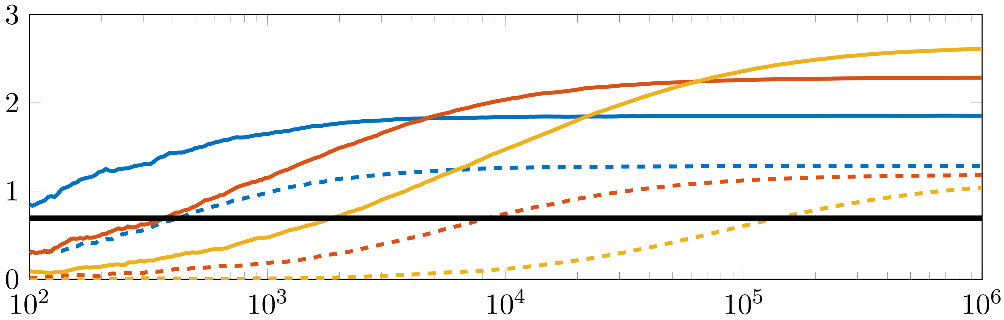
In this section, we compare the weak ordinal approach and the strong ordinal approach, or rather the weak timing, i.e.

$$E_d = \{(0, t) \mid t \in \{0, 1, 2, \ldots, d\}\}\,;\; d \in \mathbb{N},$$

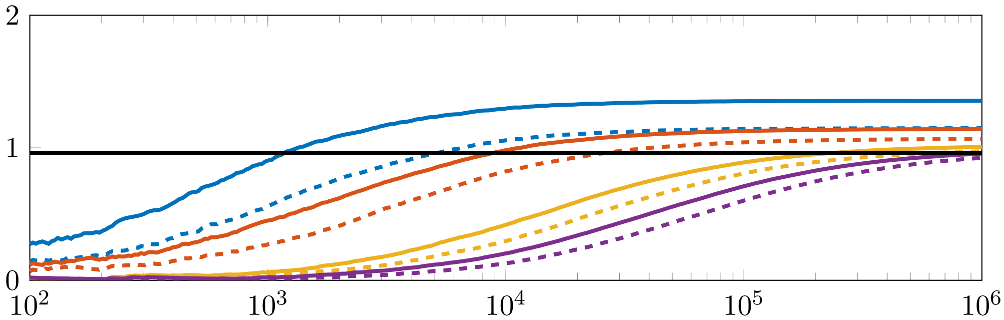
and the full timing, i.e.

$$E_d = \{(s, t) \mid s, t \in \{0, 1, 2, \ldots, d\} \text{ with } s < t\}\,;\; d \in \mathbb{N}$$

(see Section 3.2.2 and Section 3.3.2) with respect to the ordinal idea. Hence, in the following, let $\mathcal{R}$ be the basic symbolization scheme as given in (4.2). In order to



(a) The ordinal approach for word length $t = 2$ and $d \in \{3, 5, 7\}$ (blue, red, yellow).



(b) The ordinal approach for $d = 3$ and word length $t \in \{3, 4, 6, 7\}$ (blue, red, yellow, purple).

Figure 4.3: Comparing the weak (solid lines) and the full timing (dashed lines) by naively estimating $H_\mu((\mathcal{P}_{d,2})_{t+1}) - H_\mu((\mathcal{P}_{d,2})_t)$ for orbits of Arnold's cat map in dependence on the orbit length. The KS entropy is $\ln\left(\frac{3+\sqrt{5}}{2}\right)$ (black line).

compare the two timings, we simulated an orbit $(x_s, y_s)_{s\in\mathbb{N}_0}$ of Arnold's cat map T (see Remark 4.3), i.e.

$$(x_{s+1}, y_{s+1}) = (2x_s + y_s, x_s + y_s) \mod 1$$

for all $s \in \mathbb{N}_0$ and some randomly chosen (x_0, y_0). We decoded $(x_s)_{s\in\mathbb{N}}$ and $(y_s)_{s\in\mathbb{N}}$ into sequences of symbols $(r_s)_{s\in\mathbb{N}_0}$ and $(u_s)_{s\in\mathbb{N}_0}$ using the strong as well as the weak ordinal approach (see Keller and Sinn [51]), and subsequently, assigned a new unique symbol to all pairs $(r_s, u_s)_{s\in\mathbb{N}_0}$ with the same values. In this way, we obtained a new sequence of symbols in accordance with the two-dimensional setting. Again, we fixed a word length $t \in \mathbb{N}$, and naively estimated the difference

$$H_\mu((\mathcal{P}_{d,2})_{t+1}) - H_\mu((\mathcal{P}_{d,2})_t) \tag{4.5}$$

by replacing the probabilities by relative frequencies of symbol word occurrences. The results are shown in Figure 4.3 for different word lengths $t \in \mathbb{N}$ and orders $d \in \mathbb{N}$ in dependence on the orbit length $\mathcal{T}$ between 10^2 and 10^6. The curves show that (4.5) is not yet a good approximation of the KS entropy if the word length $t \in \mathbb{N}$ is small. Moreover, the curves reflect that (4.5) decreases to the entropy rate $h_\mu(T, \mathcal{P}_{d,2})$ for an increasing word length $t \in \mathbb{N}$, and that $(h_\mu(T, \mathcal{P}_{d,2}))_{d\in\mathbb{N}}$ increases. Further, since the full timing is consistent and entails all time pairs of the weak timing, it is evident that the strong ordinal approach provides better estimates of (4.5) as long as the

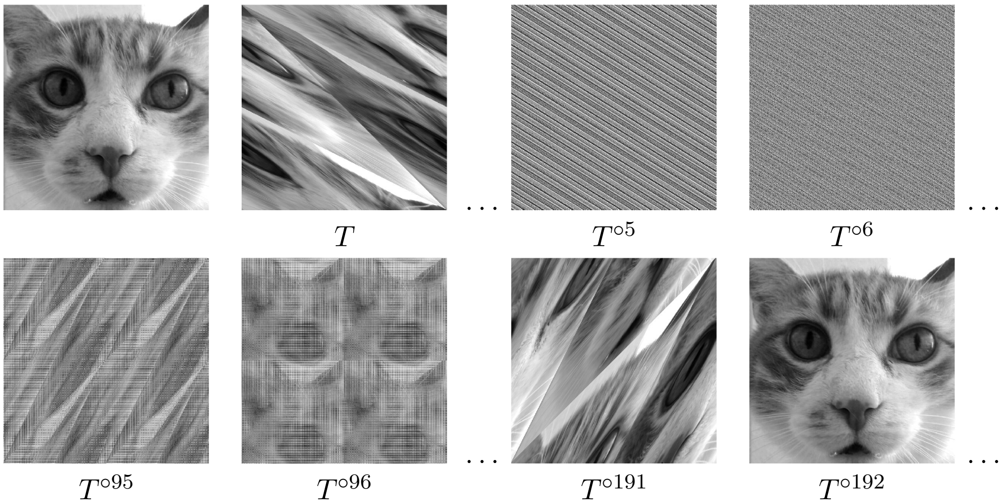


Figure 4.4: Visualization of Arnold's cat map T with the help of a 768×768 image of Bubi, i.e. we partition Ω into 589824 pixels and determine the t-th iterate of each pixel under T. It holds that for $t = 192$ each pixel is back at its place of origin. The picture of Bubi was taken by Dagmar Weigl in October 2017 and kindly provided for this thesis.

orbit length is high. However, for small orbit lengths the undersampling problem is somewhat less pronounced for the weak ordinal approach. This is an advantage of symbolic schemes $(\mathcal{R}, (E_d)_{d\in\mathbb{N}})$ where $(E_d)_{d\in\mathbb{N}}$ is the weak and not the full timing. Since less comparison steps are needed and thus less symbols and symbol words, the weak ordinal approach has, in general, better statistical properties. Particularly, if $d \in \mathbb{N}$ and $t \in \mathbb{N}$ are large, i.e. the symbols and symbol words are observed sufficiently. In fact, we have

$$|(\mathcal{P}_{d,n})_t| \leq \Big((d+1)!\,(d+1)^{t-1}\Big)^n \text{ and } |\mathcal{P}_{d+t-1,n}| \leq ((d+t)!)^n$$

if the full timing is entailed (see Unakafova [85, Lemma 6]), and

$$|(\mathcal{P}_{d,n})_t| \leq \left(\left(2^d\right)^t\right)^n \text{ and } |\mathcal{P}_{d+t-1,n}| \leq \left(2^{d+t-1}\right)^n$$

if the weak timing is entailed. Note that, in the case of the strong ordinal approach, successive words contain almost the same information. This overlapping gives the formula for the upper bound of $|(\mathcal{P}_{d,n})_t|$, and is exploited by the fast algorithms for the computation of empirical Permutation and ordinal Conditional entropy (see for instance Unakafova and Keller [84] and Unakafova [85]). Moreover, in Figure 4.3**(b)**, it is recognizable that the curves, in particular, the purple curve obtained by applying the weak ordinal approach, get closer to the ones obtained by applying the strong ordinal approach for an increasing word length $t \in \mathbb{N}$. Therefore, it is worth to consider the weak ordinal approach since, in general, less memory is needed, and hence larger values of $d \in \mathbb{N}$ and $t \in \mathbb{N}$ are possible.

★ **Remark 4.3** (Arnold's cat map)**.** Let $\Omega = [0,1) \times [0,1)$ and consider the transformation $T_a : ([0,1) \times [0,1)) \hookleftarrow$ defined by

$$T(\omega) = T(\omega_1, \omega_2) = (2\omega_1 + \omega_2, \omega_1 + \omega_1) \mod 1$$

for all $\omega = (\omega_1, \omega_2) \in \Omega$. In literature, T is called *Arnold's cat map.* If the opposite sides of the unit square are identified with each other (see Brin and Stuck [17]), then T is an automorphism on the two-dimensional torus induced by

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

(see Denker et al. [29, Section 24]). Moreover, the eigenvalues of A are $\frac{3\pm\sqrt{5}}{2}$, hence T is ergodic and the KS entropy is $\ln\left(\frac{3+\sqrt{5}}{2}\right)$ (see Denker et al. [29, Section 24]). For a deeper discussion of Arnold's cat map, we refer also to Andries et al. [8], Katok and Hasselblatt [46], Collet and Eckemann [24] and the references given therein. A visualization of T is given in Figure 4.4. ★

4.2.4 Remarks

In this chapter, we not only shed new light on the different symbolic schemes used in literature (see Amigó et al. [7], Daw et al. [27], Kurths et al. [60] and Zanin et al. [88]), but also on the use of the conditional entropy (compare to Keller et al. [50]). In doing so, we emphasized that it is interesting and worth comparing different schemes due to their different asymptotic behavior.

Clearly, common problems of time series analyses have to be faced. These problems are due to the finite nature of the considered time series and due to the approximate calculation of a processing system. In general, it is a trade-off between computational capacity and accuracy that includes undersampling problems, the choice of parameters, stationarity assumptions, and so forth (see for instance Keller et al. [50]). We are convinced that the relatively new ordinal approach is going to benefit from results achieved in the analysis of measured data with classical symbolic techniques, for instance, in order to estimate a good basic symbolization scheme (see Steuer et al. [78], Letellier [62], Li and Ray [63] and the references given therein).

Appendix

In the following, we recall some facts and properties on topology (A.) and measure theory (B.) that are relevant in our discussions. For a deeper discussion of the theory, we refer the reader to Billingsley [13], Fremlin [35], Munkres [68], Brin and Struck [17], Bogachev [14], and the references given therein.

A.1 A non-empty set Ω is *topological* if there exists a family Υ of subsets of Ω satisfying the following axioms:
(i) The empty set $\emptyset$ and Ω belong to Υ.
(ii) If $U_1, U_2, \ldots \in \Upsilon$, then $\bigcup_{n=1}^{\infty} U_n \in \Upsilon$ (*closed under countable unions*).
(iii) If U_1 and $U_2 \in \Upsilon$, then $U_1 \cap U_2 \in \Upsilon$ (*closed under finite intersections*).
We call (Ω, Υ) a *topological space*, the elements of Υ *open subsets* and Υ a *topology*.

A.2 Let (Ω, Υ) and (Ω^*, Υ^*) be two topological spaces. A map $\phi : \Omega \to \Omega^*$ is *continuous* if $\phi^{-1}(U) \in \Upsilon$ holds for all sets $U \in \Upsilon^*$.

A.3 A non-empty set Ω is said to be *metrical* if there exists a non-negative map $d : \Omega \times \Omega \to [0, \infty)$ satisfying the following axioms for any ω_1, ω_2 and $\omega_3 \in \Omega$:
(i) $d(\omega_1, \omega_2) = 0 \Leftrightarrow \omega_1 = \omega_2$ (*identity of indiscernibles*),
(ii) $d(\omega_1, \omega_2) = d(\omega_2, \omega_1)$ (*symmetry*) and
(iii) $d(\omega_1, \omega_2) \leq d(\omega_1, \omega_3) + d(\omega_3, \omega_2)$ (*triangle inequality*).
In this case, d is called *metric* and (Ω, d) a *metric space*.

A.4 Let (Ω, d) be a metric space. The metric d yields for any $\varepsilon > 0$ and $\omega \in \Omega$ an *open neighborhood* in Ω given by

$$U_{\omega,\varepsilon} := \{\omega^* \in \Omega \mid d(\omega, \omega^*) < \varepsilon\}.$$

A subset $V \subseteq \Omega$ is *open* if every $\omega \in V$ has an open neighborhood in V. A subset $V \subseteq \Omega$ is *closed* if its complement $\Omega \setminus V$ is open. In fact, the family of open subsets of (Ω, d) is a topology.

A.5 A topological space (Ω, Υ) is *compact* if each *open cover* of Ω, i.e. $\Omega = \bigcup_{i \in I} U_i$ with $U_i \in \Upsilon$ for some index set I, has a finite *subcover*, i.e.

$$\Omega = \bigcup_{k=1}^{n} U_{i_k} \text{ with } i_k \in I.$$

A.6 A topological space (Ω, Υ) is *homeomorphic* to a topological space (Ω^*, Υ^*) if there exists a function $\phi : \Omega \to \Omega^*$ with the following properties:
(i) ϕ is a bijection,
(ii) ϕ and ϕ^{-1} are continuous.
The map ϕ is called a *homeomorphism* between Ω and Ω^*. If ϕ is a map between Ω and Ω^* and a homeomorphism between Ω and $\phi(\Omega)$, then ϕ is called an *embedding*, i.e. Ω *can be embedded into* Ω^*.

A.7 Let (Ω, Υ) be a topological space. If there exists a sequence $(\omega_i)_{i\in\mathbb{N}}$ with $\omega_i \in \Omega$ such that every open non-empty subset U of Ω contains at least one ω_i, then (Ω, Υ) is called *separable*.

A.8 A topological space (Ω, Υ) is called a *Hausdorff space* if for any two states $\omega_1 \neq \omega_2$ of Ω there exist two disjoint open subsets U, $V \in \Upsilon$ with $\omega_1 \in U$ and $\omega_2 \in V$.

A.9 Let (Ω, d) be a metric space and $(\omega_i)_{i\in\mathbb{N}}$ with $\omega_i \in \Omega$ a *Cauchy sequence*, i.e. for every $\varepsilon > 0$ there exists some $N_\varepsilon \in \mathbb{N}$ such that $d(\omega_n, \omega_m) < \varepsilon$ for n, $m \geq N_\varepsilon$. The space (Ω, d) is *complete* if every Cauchy sequence converges in Ω. A topological space (Ω, Υ) is *metrizable* if it is homeomorphic to a metric space, and *completly metrizable* if it is homeomorphic to a complete metric space.

We consider the *Cartesian product* $\Omega^{\mathbb{N}} := \{(\omega_1, \omega_2, \dots) \mid \omega_i \in \Omega \text{ for all } i \in \mathbb{N}\}$ endowed with the *product topology*, i.e. the coarsest topology such that all canonical projections

$$p_i : \Omega^{\mathbb{N}} \to \Omega$$

are continuous. In particular, we are interested in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}_0} = \mathbb{R}^{\mathbb{N}\times\mathbb{N}_0}$. Note that by the countability of $\mathbb{N} \times \mathbb{N}_0$, we restrict theoretical discussions to $\mathbb{R}^{\mathbb{N}}$.

B.1 A *measurable space* $(\Omega, \mathcal{A})$ is a non-empty set Ω equipped with a *σ-algebra* $\mathcal{A}$, i.e. a collection of subsets of Ω satisfying the following axioms:
(i) The empty set $\emptyset$ belongs to $\mathcal{A}$.
(ii) If $A \in \mathcal{A}$, then $\Omega \setminus A \in \mathcal{A}$ (*closed under complements*).
(iii) If $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ (*closed under countable unions*).
A set $A \subseteq \Omega$ is called *measurable* if $A \in \mathcal{A}$. A σ-algebra $\mathcal{F} \subset \mathcal{A}$ is called a *sub-σ-algebra* of $\mathcal{A}$.

B.2 Let $(\Omega, \mathcal{A})$ be a measurable space. A map $\mu : \mathcal{A} \to \mathbb{R}$ is *countable additive* if

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{i=1}^{\infty} \mu(A_n)$$

for all pairwise disjoint sets $A_1, A_2, \dots \in \mathcal{A}$. A *measure space* $(\Omega, \mathcal{A}, \mu)$ is a measurable space $(\Omega, \mathcal{A})$ equipped with a *measure* $\mu : \mathcal{A} \to \mathbb{R}$, i.e. μ is non-negative and countable additive. A set $A \in \mathcal{A}$ is a *null set* if $\mu(A) = 0$. An *atom* is an one-point subset with positive measure. A *measure space* $(\Omega, \mathcal{A}, \mu)$ is *complete* if every subset B of any null set $A \in \mathcal{A}$ is measurable. In this case, μ is called *complete*.

B.3 A *probability space* $(\Omega, \mathcal{A}, \mu)$ is a measure space for which $\mu : \mathcal{A} \to [0, 1]$ with $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$. In this case, μ is called a *probability measure*.

B.4 Let $(\Omega, \mathcal{A})$ and $(\Omega^*, \mathcal{A}^*)$ be two measurable spaces. A map $\phi : \Omega \to \Omega^*$ is $\mathcal{A}$-$\mathcal{A}^*$ *measurable* if

$$\phi^{-1}(A^*) := \{\omega \in \Omega \mid \phi(\omega) \in A^*\} \in \mathcal{A}$$

for all $A^* \in \mathcal{A}^*$. In the case of measurability, we write: $\phi : (\Omega, \mathcal{A}) \to (\Omega^*, \mathcal{A}^*)$. If we consider a probability space as a domain, then a function

$$X : (\Omega, \mathcal{A}, \mu) \to (\Omega^*, \mathcal{A}^*)$$

is called *random variable*.

B.5 Let $(\Omega, \mathcal{A})$ be a non-empty measurable space and $\mathcal{M}$ an arbitrary family of subsets of Ω. The sub-σ-algebra generated by $\mathcal{M}$ is given by the smallest σ-algebra of Ω containing $\mathcal{M}$, that is

$$\sigma(\mathcal{M}) := \bigcap_{\substack{\mathcal{F} \subset \mathcal{A} \text{ is } \sigma\text{-algebra} \\ \text{and } \mathcal{M} \subset \mathcal{F}}} \mathcal{F}.$$

The family $\mathcal{M}$ *generates* $\mathcal{A}$ if $\sigma(\mathcal{M}) = \mathcal{A}$. Let $(\mathcal{F}_r)_{r\in\mathbb{N}}$ be a *sequence of σ-algebras* of Ω. The *join*

$$\bigvee_{r\in\mathbb{N}} \mathcal{F}_r := \sigma\left(\bigcup_{r\in\mathbb{N}} \mathcal{F}_r\right) \subset \mathcal{A}$$

is the minimal σ-algebra containing all $\mathcal{F}_r$. In the case of two σ-algebras, we write $\mathcal{F}_1 \vee \mathcal{F}_1 := \bigvee_{r=1}^{2} \mathcal{F}_r$.

B.6 Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and let $\mu^* : 2^\Omega \to \mathbb{R}$ (where 2^Ω denotes the power set of Ω) be defined by

$$\mu^*(B) := \inf\{\mu(A) \mid A \in \mathcal{A}, B \subseteq A\}$$

for all $B \subseteq \Omega$. Then

(i) for any $B \subseteq \Omega$ there exists a measurable set A such that $B \subseteq A$ and $\mu^*(B) = \mu(A)$ (A is called a *measurable envelope of B*),
(ii) $\mu^*(B) \le \mu^*(C)$ for any $B \subseteq C \subseteq \Omega$,
(iii) μ^* is *countable subadditive*, i.e. $\mu^*(\bigcup_{i\in\mathbb{N}}) \le \sum_{i\in\mathbb{N}} \mu^*(B_i)$ for any family $(B_i)_{i\in\mathbb{N}}$ with $B_i \subseteq \Omega$,
(iv) $\mu^*(A) = \mu(A)$ for any $A \in \mathcal{A}$, i.e. $0 = \mu^*(\emptyset) \le \mu^*(B) \le \mu^*(\Omega) = 1$

(see for instance Fremlin [35, Chapter 13]). The map μ^* is called the *outer measure with respect to μ*.

B.7 Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and $B \subset \Omega$. The *trace σ-algebra $\mathcal{A}|_B$ of the space B* is given by

$$\mathcal{A}|_B := \{A \cap B \mid A \in \mathcal{A}\}.$$

The *restriction of μ to $\mathcal{A}|_B$* is defined by

$$\mu|_B(A \cap B) = \mu(A \cap \widetilde{B});\ A \in \mathcal{A},$$

where $\widetilde{B}$ is an arbitrary measurable envelope of B. Let $\mathcal{M}$ be an arbitrary family of subsets of Ω such that $\sigma(\mathcal{M}) = \mathcal{A}$, then $\sigma(\mathcal{M} \cap B) = \mathcal{A}|_B$ (see Billingsley [13, Theorem 10.1.]).

B.8 Let (Ω, Υ) be a topological space. The σ-algebra $\mathcal{B}(\Omega) := \sigma(\Upsilon)$ is called the *Borel σ-algebra* of Ω. The elements $B \in \mathcal{B}(\Omega)$ are called *Borel sets*, and a measure μ on $\mathcal{B}(\Omega)$ is called a *Borel measure* on Ω. Mostly, we are interested in the Borel σ-algebra $\mathcal{B}(\mathbb{R}^\mathbb{N})$, for which $\mathcal{B}\left(\mathbb{R}^\mathbb{N}\right) = \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}} := \sigma(\mathcal{Z})$ (see for instance Bogachev [14, Lemma 6.4.2]), where $\mathcal{Z} = (Z_n)_{n\in\mathbb{N}}$ is the family of *cylinder sets*

$$Z_n(B_1, B_2, \ldots, B_n) = \left\{(x_t)_{t\in\mathbb{N}} \in \mathbb{R}^\mathbb{N} \mid x_i \in B_i, i \in \{1, 2, \ldots, n\}\right\}$$

with $n \in \mathbb{N}$ and $B_i \in \mathcal{B}(\mathbb{R})$ for all $i \in \{1, 2, \ldots, n\}$.

Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}} : (\Omega, \mathcal{A}, \mu) \to \left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})\right)$ be a sequence of random variables $X_i : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\sigma(\mathbf{X})$ be the sub-σ-algebra of $\mathbf{X}$ given by

$$\sigma(\mathbf{X}) := \bigvee_{i\in\mathbb{N}} \sigma(X_i) = \bigvee_{i\in\mathbb{N}} \left\{X_i^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\right\} \subset \mathcal{A}.$$

Since $\mathcal{Z}$ generates $\mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$, it holds that $\sigma\left(\mathbf{X}^{-1}(\mathcal{Z})\right) = \mathbf{X}^{-1}(\sigma(\mathcal{Z}))$ (see for instance Elstrodt [34, Chapter 1, Theorem 4.4]), i.e. $\mathbf{X}^{-1}(F) \in \sigma\left(\mathbf{X}^{-1}(\mathcal{Z})\right)$ for all $F \in \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$. Further,

$$\mathbf{X}^{-1}(Z_n(B_1, B_2, \ldots, B_n)) = \bigcap_{i=1}^{n} X_i^{-1}(B_i) \in \bigvee_{i=1}^{n} \sigma(X_i)$$

for all $n \in \mathbb{N}$ and $B_i \in \mathcal{B}(\mathbb{R})$ with $i \in \{1, 2, \ldots, n\}$, i.e. $\mathbf{X}^{-1}(F) \in \sigma(\mathbf{X})$ for all $F \in \mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right)$.

Bibliography

[1] Adami, C. What is complexity? *BioEssays 24*, 12 (2002), 1085–1094. https://doi.org/10.1002/bies.10192.

[2] Allen, R. M., Nolet, G., Morgan, W. J., Vogfjörd, K., Bergsson, B. H., Erlendsson, P., Foulger, G. R., Jakobsdóttir, S., Julian, B. R., Pritchard, M., Ragnarsson, S., and Stefánsson, R. The thin hot plume beneath Iceland. *Geophysical Journal International 137*, 1 (1999), 51–63. https://doi.org/10.1046/j.1365-246x.1999.00753.x.

[3] Allen, R. M., Nolet, G., Morgan, W. J., Vogfjörd, K., Bergsson, B. H., Erlendsson, P., Foulger, G. R., Jakobsdóttir, S., Julian, B. R., Pritchard, M., Ragnarsson, S., and Stefánsson, R. Imaging the mantle beneath Iceland using integrated seismological techniques. *Journal of Geophysical Research: Solid Earth 107*, B12 (2002), ESE 3–1–ESE 3–16. 2325, https://doi.org/10.1029/2001JB000595.

[4] Allen, R. M., Nolet, G., Morgan, W. J., Vogfjörd, K., Nettles, M., Ekström, G., Bergsson, B. H., Erlendsson, P., Foulger, G. R., Jakobsdóttir, S., Julian, B. R., Pritchard, M., Ragnarsson, S., and Stefánsson, R. Plume-driven plumbing and crustal formation in Iceland. *Journal of Geophysical Research: Solid Earth 107*, B8 (2002), ESE 4–1–ESE 4–19. https://doi.org/10.1029/2001JB000584.

[5] Amigó, J. *Permutation complexity in dynamical systems. Ordinal patterns, permutation entropy and all that.* Springer Series in Synergetics, Springer Complexity. Springer-Verlag, Berlin, 2010. https://doi.org/10.1007%2F978-3-642-04084-9.

[6] Amigó, J. M., Keller, K., and Kurths, J., Eds. *Recent progress in symbolic dynamics and permutation complexity. Ten years of permutation entropy* (2013), vol. 222 of *2*, The European Physical Journal Special Topics. https://doi.org/10.1140/epjst/e2013-01839-6.

[7] Amigó, J. M., Keller, K., and Unakafova, V. A. Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences 373*, 2034 (2015), 20140091. https://doi.org/10.1098/rsta.2014.0091.

[8] Andries, J., Fannes, M., Tuyls, P., and Alicki, R. The dynamical entropy of the quantum Arnold cat map. *Letters in Mathematical Physics 35*, 4 (1995), 375–383. https://doi.org/10.1007%2Fbf00750844.

[9] Antoniouk, A., Keller, K., and Maksymenko, S. Kolmogorov-Sinai entropy via separation properties of order-generated sigma-algebras. *Discrete and Continuous Dynamical Systems - A 34*, 5 (2014), 1793–1809. https://doi.org/10.3934%2Fdcds.2014.34.1793.

[10] Bandt, C., Keller, G., and Pompe, B. Entropy of interval maps via permutations. *Nonlinearity 15*, 5 (2002), 1595–1602. https://doi.org/10.1088%2F0951-7715%2F15%2F5%2F312.

[11] Bandt, C., and Pompe, B. Permutation entropy. A natural complexity measure for time series. *Physical Review Letters 88*, 17 (2002), 174102. https://doi.org/10.1103%2Fphysrevlett.88.174102.

[12] Berger, S., Schneider, G., Kochs, E., and Jordan, D. Permutation entropy: Too complex a measure for EEG time series? *Entropy 19*, 12 (2017), 692. https://doi.org/10.3390%2Fe19120692.

[13] Billingsley, P. *Probability and measure*, 3 ed. Wiley series in probability and mathematical statistics. Wiley, New York, 1995.

[14] Bogachev, V. I. *Measure Theory Volume I and II*. Springer-Verlag, Berlin, 2007. https://doi.org/10.1007%2F978-3-540-34514-5.

[15] Bollt, E. M., Stanford, T., Lai, Y.-C., and Życzkowski, K. What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossings analysis of chaotic time-series. *Physica D: Nonlinear Phenomena 154*, 3-4 (2001), 259–286. https://doi.org/10.1016%2Fs0167-2789%2801%2900242-1.

[16] Bonachela, J. A., Hinrichsen, H., and Muñoz, M. A. Entropy estimates of small data sets. *Journal of Physics A: Mathematical and Theoretical 41*, 20 (2008), 202001. https://doi.org/10.1088%2F1751-8113%2F41%2F20%2F202001.

[17] Brin, M., and Stuck, G. *Introduction to Dynamical Systems*. Cambridge University Press, Cambridge, 2002. https://doi.org/10.1017%2Fcbo9780511755316.

[18] Bronstein, I. N., and Semendjajew, K. A. *Taschenbuch der Mathematik*. G. Musiol and H. Mühlig, Eds., Verlag Europa-Lehrmittel, Harri Deutsch. [In German].

[19] Buzzi, J. Chaos and ergodic theory. In *Mathematics of Complexity and Dynamical Systems*, R. A. Meyers, Ed. Springer-Verlag, New York, 2012, pp. 63–87. https://doi.org/10.1007%2F978-1-4614-1806-1_6.

[20] Cantón, A., Granados, A., and Pommerenke, C. Borel images and analytic functions. *Michigan Mathematical Journal 52*, 2 (2004), 279–287. https://doi.org/10.1307%2Fmmj%2F1091112076.

[21] Cao, Y., Tung, W.-W., Gao, J. B., Protopopescu, V. A., and Hively, L. M. Detecting dynamical changes in time series using the permutation entropy. *Physical Review E 70*, 4 (2004). https://doi.org/10.1103%2Fphysreve.70.046217.

[22] Chan, K.-S., and Tong, H. *Chaos: A Statistical Perspective.* Springer Series in Statistics. Springer-Verlag, New York, 2001. `https://doi.org/10.1007%2F978-1-4757-3464-5`.

[23] Choe, G. H. *Computational Ergodic Theory*, vol. 13 of *Algorithms and Computation in Mathematics.* Springer-Verlag, Berlin, 2005. `https://doi.org/10.1007%2Fb138894`.

[24] Collet, P., and Eckmann, J.-P. *Concepts and Results in Chaotic Dynamics. A Short Course.* Theoretical and Mathematical Physics. Springer-Verlag, Berlin, 2006. `https://doi.org/10.1007%2F978-3-540-34706-4`.

[25] Cover, T. M., and Thomas, J. A. *Elements of Information Theory*, 2 ed. John Wiley & Sons, Inc., 2006. `https://doi.org/10.1002%2F047174882x`.

[26] Crutchfield, J. P., and Packard, N. H. Symbolic dynamics of noisy chaos. *Physica D: Nonlinear Phenomena 7*, 1-3 (1983), 201–223. `https://doi.org/10.1016%2F0167-2789%2883%2990127-6`.

[27] Daw, C. S., Finney, C. E. A., and Tracy, E. R. A review of symbolic analysis of experimental data. *Review of Scientific Instruments 74*, 2 (2003), 915–930. `https://doi.org/10.1063%2F1.1531823`.

[28] Denker, M. *Einführung in die Analysis dynamischer Systeme.* Springer-Verlag, Berlin, 2005. `https://doi.org/10.1007%2Fb137966` [In German].

[29] Denker, M., Grillenberger, C., and Sigmund, K. *Ergodic Theory on Compact Spaces*, vol. 527 of *Lecture Notes in Mathematics.* Springer-Verlag, 1976. `https://doi.org/10.1007%2Fbfb0082364`.

[30] Einarsson, P., Brandsdóttir, B., Gudmundsson, M. T., Björnsson, H., Grínvold, K., and Sigmundsson, F. Center of the Iceland hotspot experiences volcanic unrest. *Eos, Transactions American Geophysical Union 78*, 35 (1997), 369–375. `https://doi.org/10.1029%2F97eo00237`.

[31] Einsiedler, M., Lindenstrauss, E., and Ward, T. Entropy in ergodic theory and homogeneous dynamics. Unpublished manuscript. Available from: `http://www.personal.leeds.ac.uk/~mattbw/entropy/welcome.html` [access: June 10, 2017], 2015.

[32] Einsiedler, M., and Schmidt, K. *Dynamische Systeme. Ergodentheorie und topologische Dynamik.* Mathematik kompakt. Birkhäuser Basel, Basel, 2014. `https://doi.org/10.1007%2F978-3-0348-0634-3` [In German].

[33] Einsiedler, M., and Ward, T. *Ergodic Theory: with a view towards Number Theory*, vol. 259 of *Graduate Texts in Mathematics.* Springer-Verlag, London, 2011. `https://doi.org/10.1007%2F978-0-85729-021-2`.

[34] Elstrodt, J. *Maß- und Integrationstheorie*, 4 ed. Springer Lehrbuch. Springer-Verlag, Berlin, 2005. `https://doi.org/10.1007%2F978-3-642-17905-1` [In German].

[35] Fremlin, D. H. *Measure theory. Volume 1. The irreducible minimum.* Torres Fremlin, Colchester, 2000. Corrected second printing of the 2000 original.

[36] Glynn, C. C., and Konstantinou, K. I. Reduction of randomness in seismic noise as a short-term precursor to a volcanic eruption. *Scientific Reports 6*, 1 (2016). https://doi.org/10.1038%2Fsrep37733.

[37] Grassberger, P. Information flow and maximum entropy measures for 1-d maps. *Physica D: Nonlinear Phenomena 14*, 3 (1985), 365–373. https://doi.org/10.1016%2F0167-2789%2885%2990095-8.

[38] Grassberger, P. Entropy estimates from insufficient samplings. *ArXiv Physics e-prints* (2003). Available from: https://arxiv.org/abs/physics/0307138, version 2 [access: January 10, 2017].

[39] Gutman, Y. Takens' embedding theorem with a continuous observable. In *Ergodic Theory. Advances in dynamical systems*, I. Assani, Ed. De Gruyter, 2016, pp. 134–141. https://doi.org/10.1515%2F9783110461510-006.

[40] Guust, N. *Seismic study of the Iceland hotspot.* International Federation of Digital Seismograph Networks. Other/Seismic Network. 10.7914\SN\XD_1996, 1996. FDSN entry: http://www.fdsn.org/networks/detail/XD_1996/.

[41] Hammel, S. M., Yorke, J. A., and Grebogi, C. Do numerical orbits of chaotic dynamical processes represent true orbits? *Journal of Complexity 3*, 2 (1987), 136–145. https://doi.org/10.1016%2F0885-064x%2887%2990024-0.

[42] Hasselblatt, B., and Katok, A. *A First Course in Dynamics. With a Panorama of Recent Developments.* Cambridge University Press, Cambridge, 2003. https://doi.org/10.1017/CBO9780511998188.

[43] Hively, L. M., Protopopescu, V. A., and Gailey, P. C. Timely detection of dynamical change in scalp EEG signals. *Chaos: An Interdisciplinary Journal of Nonlinear Science 10*, 4 (2000), 864–875.

[44] Isermann, R. *Identifikation dynamischer Systeme 1*, 2 ed. Springer-Verlag, Berlin, 1992. https://doi.org/10.1007%2F978-3-642-84679-3 [In German].

[45] Itô, K. *Introduction to Probability Theory.* Cambridge University Press, Cambridge, 1984.

[46] Katok, A., and Hasselblatt, B. *Introduction to the Modern Theory of Dynamical Systems*, 4 ed. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1995. https://doi.org/10.1017%2Fcbo9780511809187.

[47] Kechris, A. S. *Classical Descriptive Set Theory*, 156 ed. Graduate Texts in Mathematics. Springer-Verlag, New York, 1995. https://doi.org/10.1007%2F978-1-4612-4190-4.

[48] Keller, K., and Lauffer, H. Symbolic analysis of high-dimensional time series. *International Journal of Bifurcation and Chaos 13*, 9 (2003), 2657–2668. https://doi.org/10.1142%2Fs0218127403008168.

[49] Keller, K., Maksymenko, S., and Stolz, I. Entropy determination based on the ordinal structure of a dynamical system. *Discrete and Continuous Dynamical Systems - Series B 20*, 10 (2015), 3507–3524. https://doi.org/10.3934%2Fdcdsb.2015.20.3507.

[50] Keller, K., Mangold, T., Stolz, I., and Werner, J. Permutation entropy. New ideas and challenges. *Entropy 19*, 12 (2017), 134. https://doi.org/10.3390%2Fe19030134, licensed under CC BY 4.0.

[51] Keller, K., and Sinn, M. A standardized approach to the Kolmogorov–Sinai entropy. *Nonlinearity 22*, 10 (2009), 2417–2422. https://doi.org/10.1088%2F0951-7715%2F22%2F10%2F006.

[52] Keller, K., and Sinn, M. Kolmogorov–Sinai entropy from the ordinal viewpoint. *Physica D: Nonlinear Phenomena 239*, 12 (2010), 997–1000. https://doi.org/10.1016%2Fj.physd.2010.02.006.

[53] Keller, K., Sinn, M., and Emonds, J. Time series from the ordinal viewpoint. *Stochastics and Dynamics 7*, 2 (2007), 247–272. https://doi.org/10.1142%2Fs0219493707002025.

[54] Keller, K., Unakafov, A., and Unakafova, V. Ordinal patterns, entropy, and EEG. *Entropy 16*, 12 (2014), 6212–6239. https://doi.org/10.3390%2Fe16126212.

[55] Keller, K., Unakafov, A. M., and Unakafova, V. A. On the relation of KS entropy and permutation entropy. *Physica D: Nonlinear Phenomena 241*, 18 (2012), 1477–1481. https://doi.org/10.1016%2Fj.physd.2012.05.010.

[56] Kennel, M. B., and Buhl, M. Estimating good discrete partitions from observed data. Symbolic false nearest neighbors. *Physical Review Letters 91*, 8 (2003). https://doi.org/10.1103%2Fphysrevlett.91.084102.

[57] Kitchens, B. P. *Symbolic Dynamics. One-sided, Two-sided and Countable State Markov Shifts*. Universitext. Springer-Verlag, Berlin, 1998. https://doi.org/10.1007%2F978-3-642-58822-8.

[58] Krengel, U. *Ergodic Theorems*, vol. 6 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 1985. https://doi.org/10.1515%2F9783110844641.

[59] Kuratowski, K. *Topology Volume 1*. Academic Press New York, PWN-Polish Scientific Publishers, Warsaw, 1966. https://doi.org/10.1016%2Fc2013-0-11022-7.

[60] Kurths, J., Schwarz, U., Witt, A., Krampe, R. T., and Abel, M. Measures of complexity in signal analysis. In *AIP Conference Proceedings* (New York, 1996), R. Katz, Ed., vol. 375, AIP, pp. 33–54. Paper presented at Chaotic Fractal and Nonlinear Signal Processing, 3rd Technical Conference on Nonlinear Dynamics (Chaos) and Full Spectrum Processing, New London, CT, USA, 10–13 July 1995, https://doi.org/10.1063%2F1.51037.

[61] Ladyman, J., Lambert, J., and Wiesner, K. What is a complex system? *European Journal for Philosophy of Science 3*, 1 (2013), 33–67. https://doi.org/10.1007%2Fs13194-012-0056-8.

[62] Letellier, C. Symbolic sequence analysis using approximated partition. *Chaos, Solitons & Fractals 36*, 1 (2008), 32–41. https://doi.org/10.1016%2Fj.chaos.2006.06.025.

[63] Li, Y., and Ray, A. Unsupervised symbolization of signal time series for extraction of the embedded information. *Entropy 19*, 12 (2017), 148. https://doi.org/10.3390%2Fe19040148.

[64] Lind, D., and Marcus, B. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995. https://doi.org/10.1017%2Fcbo9780511626302.

[65] Martínez, R. A., Ed. *Symbolic Entropy Analysis and Its Applications*. Entropy, 2017 and 2018. [Special Issue].

[66] Marx, B., and Vogt, W. *Dynamische Systeme. Theorie und Numerik*. Spektrum Akademischer Verlag, Heidelberg, 2011. https://doi.org/10.1007%2F978-3-8274-2448-8 [In German].

[67] MATLAB Release R2017b. The MathWorks, Inc., Natick, Massachusetts, United States.

[68] Munkres, J. R. *Topology*, 2 ed. Prentice Hall, Upper Saddle River, 2000.

[69] Paninski, L. Estimation of entropy and mutual information. *Neural Computation 15*, 6 (2003), 1191–1253. https://doi.org/10.1162%2F089976603321780272.

[70] Parry, W., and Williams, R. F. Block coding and a zeta function for finite markov chains. *Proceedings of the London Mathematical Society s3-35*, 3 (1977), 483–495. https://doi.org/10.1112%2Fplms%2Fs3-35.3.483.

[71] Peitgen, H.-O., Jürgens, H., and Saupe, D. *Fractals for the Classroom. Part Two: Complex Systems and Mandelbrot*. Springer-Verlag, New York, 1992. https://doi.org/10.1007/978-1-4612-4406-6.

[72] Quas, A. Ergodicity and mixing properties. In *Mathematics of Complexity and Dynamical Systems*, R. A. Meyers, Ed. Springer-Verlag, New York, 2012, pp. 225–240. https://doi.org/10.1007%2F978-1-4614-1806-1_15.

[73] Riedl, M., Müller, A., and Wessel, N. Practical considerations of permutation entropy. A tutorial review. *The European Physical Journal Special Topics 222*, 2 (2013), 249–262. https://doi.org/10.1140%2Fepjst%2Fe2013-01862-7.

[74] Roulston, M. S. Estimating the errors on measured entropy and mutual information. *Physica D: Nonlinear Phenomena 125*, 3-4 (1999), 285–294. https://doi.org/10.1016%2Fs0167-2789%2898%2900269-3.

[75] Rudolph, D. J. *Fundamentals of Measurable Dynamics. Ergodic Theory on Lebesgue Spaces*. Oxford University Press, New York, 1990.

[76] Schürmann, T. Bias analysis in entropy estimation. *Journal of Physics A: Mathematical and General 37*, 27 (2004), L295–L301. https://doi.org/10.1088%2F0305-4470%2F37%2F27%2Fl02.

[77] Sprott, J. C. *Chaos and time-series analysis*. Oxford University Press, New York, 2003.

[78] Steuer, R., Molgedey, L., Ebeling, W., and Jiménez-Montaño, M. Entropy and optimal partition for data analysis. *The European Physical Journal B - Condensed Matter and Complex Systems 19*, 2 (2001), 265–269. https://doi.org/10.1007%2Fs100510170335.

[79] Stolz, I., and Keller, K. A general symbolic approach to Kolmogorov-Sinai entropy. *Entropy 19*, 12 (2017), 675. https://doi.org/10.3390%2Fe19120675, licensed under CC BY 4.0.

[80] Takens, F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L. Young, Eds., Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1981, pp. 366–381. https://doi.org/10.1007%2Fbfb0091924.

[81] Trzeciak, J. *Writing Mathematical Papers in English. A practical guide.* European Mathematical Society Publishing House, Zürich, 2005. https://doi.org/10.4171%2F014 [Revised Edition].

[82] Unakafov, A. M. *Ordinal-patterns-based segmentation and discrimination of time series with applications to EEG data.* Phd thesis, Universität zu Lübeck, Lübeck, 2015. Available from: http://www.zhb.uni-luebeck.de/epubs/ediss1524.pdf [access: June 08, 2017].

[83] Unakafov, A. M., and Keller, K. Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena 269* (2013), 94–102. https://doi.org/10.1016%2Fj.physd.2013.11.015.

[84] Unakafova, V., and Keller, K. Efficiently measuring complexity on the basis of real-world data. *Entropy 15*, 12 (2013), 4392–4415. https://doi.org/10.3390%2Fe15104392.

[85] Unakafova, V. A. *Investigating Measures of Complexity for Dynamical Systems and for Time Series.* Phd thesis, Universität zu Lübeck, Lübeck, 2015. Available from: http://www.zhb.uni-luebeck.de/epubs/ediss1525.pdf [access: June 08, 2017].

[86] Unakafova, V. A. Ordinal-patterns-based analysis (beta-version). *MATLAB Central File Exchange* (2017). Available from: http://www.mathworks.com/matlabcentral/fileexchange/63782-ordinal-patterns-based-analysis--beta-version- [access: February 27, 2018].

[87] Walters, P. *An Introduction to Ergodic Theory*, vol. 458 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 1975.

[88] Zanin, M., Zunino, L., Rosso, O. A., and Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy 14*, 8 (2012), 1553–1577. https://doi.org/10.3390%2Fe14081553.