UNIVERSIDADE FEDERAL DE GOIÁS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Matheus Capela

Kolmogorov-Sinai entropy and dissipation in driven classical Hamiltonian systems

Goiânia
2018

Matheus Capela

Kolmogorov-Sinai entropy and dissipation in driven classical Hamiltonian systems

Dissertação apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Goiás como requisito para obtenção do título de Mestre em Física

Advisor: Lucas Chibebe Céleri

Goiânia

2018

In memory of Andrey Nikolaevich Kolmogorov.

Acknowledgements

First I would like to thank my advisor Prof. Dr. Lucas Chibebe Céleri for the opportunity of being his student. He has been an example of scientist to me, and always encouraged me to think deeply about physics.

I am thankful to Dr. Mikel Sanz and Prof. Dr. Enrique Solano for their collaboration in this project.

I also want to thank Dr. Vasileios Kiosses for his friendship and for the several discussions we had about physics. I have learned a lot from him.

I am grateful for the warm hospitality of GIQSul at the Federal University of Santa Catarina, where I was a visitor during the months of August and September 2017. I am specially grateful to Prof. Dr. Paulo Henrique Souto Ribeiro, Prof. Dr. Eduardo Inácio Duzzioni, Prof. Dr. Renné Luiz Câmara Medeiros de Araújo, Prof. Dr. Willamys Cristiano Soares Silva, B.Sc. Antônio Crispim Lourenço, M.Sc. Marcelo Felipe Zanello de Arruda.

I greatly acknowledge financial support (scholarship) from the Brazilian funding agency CNPq.

Lastly, I would like to thank my family for the encouragement and support.

Abstract

Many connections between physics and information theory have been revealed since the development of classical information theory by Shannon. A key concept in this connection is entropy, which represents the amount of information transferred to the observer who performs measurements in an experiment. Statistical mechanics is a physical theory deeply connected to information by Jaynes' Maximum Entropy principle, which defines equilibrium probability distributions as the ones that maximizes entropy under some physical constraints. In this way, these distributions are the less unbiased probabilities that can be assignment to an event. Following this path, the dissipated energy in a classical Hamiltonian process (also known as the thermodynamic entropy production) was connected to the relative entropy between the forward and backward probability densities. A recent work by Still et al. has revealed that energetic inefficiency and model inefficiency are equivalent concepts in Markovian processes, where the latter is defined as the difference in mutual information that the system's state shares with the future and past environmental variables. This raises the question whether model unpredictability and energetic inefficiency are connected in the framework of classical physics. The aim of this study is to connect the concepts of random behavior of a classical Hamiltonian system with its energetic inefficiency. The random behavior of a classical system is quantified by the Kolmogorov-Sinai entropy associated with its dynamics, an information-theoretic approach to chaos, whereas energetic inefficiency is measured by the dissipated work.

Resumo

Diversas relações entre física e teoria de informação foram estabelecidas desde o trabalho de Shannon. Entropia é um elemento essencial nesta conexão, quantificando a informação transferida em um experimento. Mecânica estatística está conectada à teoria de informação através do princípio de máxima entropia, definindo as distribuições de probabilidade de estados de equilíbrio como aquelas que maximizam a entropia sujeita as condições físicas apropriadas. A energia dissipada em um processo clássico está conectada a divergência de Kullback-Leibler. Recentemente, Still e colaboradores mostraram que a ineficiência energética em um processo estocástico Markoviano é equivalente a ineficiência do modelo, definida como a diferença em informação que o estado do sistema compartilha com as variáveis externas no futuro e passado. Isto sugere que imprevisibilidade e ineficiência energética estejam relacionadas no âmbito da física clássica. O objetivo deste trabalho é estabelecer uma relação entre o comportamento randômico de sistemas clássicos, quantificado pela entropia de Kolmogorov-Sinai, com a ineficiência energética.

Table of contents

List of figures

Nomenclature

ϕ	Time evolution map (automorphism)
F	Free energy
F_G	Generalized free energy
H	Shannon entropy
h	Kolmogorov-Sinai entropy
H_G	Generalized Shannon entropy
S	Thermodynamic entropy

Chapter 1

Introduction

Since the development of classical information theory by Shannon in 1948 [1], which is a framework for describing communication, several links with engineer and sciences has been discovered [2]. Of special interest for this study are the connections between physics and information theory which are discussed here. Originally, Shannon proposed a general scheme of communication system consisting of a source emitting messages to a receiver [1]. The amount of information transferred to the receiver in the communication process denoted as entropy is defined as the of uncertainty that the receiver had about the source before the measurement [1, 3, 2], and it is the key concept behind the connection between physics and information theory. Statistical mechanics has been connected to information theory by the Jaynes Maximum Entropy principle. According to this principle the equilibrium probability distribution maximizes information transference in the measurement process subjected to the available physical constraints[4]. Dissipation is a fundamental matter in physics, since it is related to reversibility of physical processes. In 2007 Kawai et al. first established a relation for the dissipated energy in a classical Hamiltonian process, usually called as the fluctuation dissipation theorem [5]. There, the dissipated work in a classical Hamiltonian process is proved to be equivalent to the Kullback-Leibler divergence between the probability densities in the forward and backward protocols, where in the both scenarios the system is assumed to be initially in thermal equilibrium with a heat reservoir. A great contribution to physics is the proof of second law of thermodynamics admitting thermodynamic equilibrium: it follows immediately from the fluctuation dissipation theorem, since Kullback-Leibler divergence is a non-negative quantity . The connection with information information theory could not be different in mathematics, mostly in dynamical systems theory [6, 7]. Using the mathematical formulation of entropy provided by Shannon in [1], Kolmogorov constructed a theoretical tool allowing to analyze the random behavior

of dynamical systems, nowadays called Kolmogorov-Sinai entropy. It is a parameter of the dynamical system which allows a criterion for defining chaos, since positive Kolmogorov-Sinai entropy is a feature of chaotic behavior [8].

A recent work by Still et al. has revealed that energetic inefficiency and model inefficiency are equivalent concepts in Markovian processes, where the latter is defined as the difference in mutual information that the system's state shares with the future and past environmental variables [9]. This raises the question whether model unpredictability and energetic inefficiency are connected in the framework of classical physics. The purpose of the present work is to connect the concept of random behavior of a classical Hamiltonian system, measured by the Kolmogorov-Sinai entropy associated with its dynamics, with its energetic inefficiency, measured by dissipated energy in accordance with the fluctuation dissipation theorem.

This work is organized as follows. In chapter 2 the necessary concepts regarding information measures are presented as a consequence of Shannon's axioms. Information measures concerning several random variables are also presented, for instance, joint entropy, conditional entropy and Kullback-Leibler divergence. An information-theoretic approach to dynamical systems is discussed and Kolmogorov-Sinai entropy is defined. In chapter 3 the necessary concepts about statistical mechanics are presented and discussed. The first section deals with stochastic thermodynamics and the result presented in [9]. The second section deals with classical statistical mechanics, and the fluctuation dissipation theorem [5] is presented. The Maximum Entropy principle is considered. Ultimately, the results are presented in chapter 4.

Chapter 2

Information Theory and Dynamical Systems

2.1 Classical Information Measures

The communication process becomes much more clear since the development of the Mathematical Theory of Communication by C. E. Shannon [1], which is often called by the community as Classical Information Theory.

A communication system is defined as a collection of a source, a channel, and a receiver, see Figure 2.1. The source generates messages, hence it is characterized by an alphabet $\mathcal{X}$, which is a collection of symbols emitted by it, and its statistical structure, i.e., the probability $p(x_t)$ of the source emitting any symbol from the alphabet $\mathcal{X}$ at each instant of time t. In this study it is only considered finite alphabets. The channel is any physical medium in which the symbols $x_t \in \mathcal{X}$ can be transmitted to the receiver, the entity which acquires the symbols emitted from the source. This definition shows complete analogy with the measurement process in a physical experiment, the source playing the role of the physical system that generates the states transmitted to the physicist, which plays the role of the physical receiver, while the channel is the measurement apparatus. In what follows, the concepts concerning probability theory are defined and discussed in appendix A. The purpose of this chapter is to describe some connections between information theory and dynamical systems theory.

2.1.1 Shannon Entropy

The concept of *information* is closely related to the concept of *uncertainty* [1]. In an actual experiment one is usually not sure about the result before the measurement, for

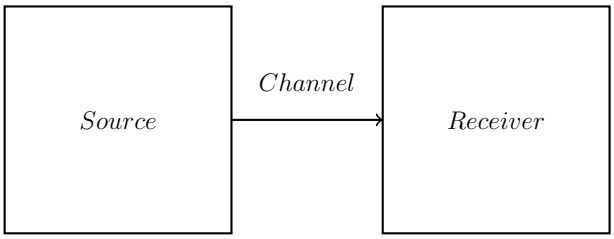

Figure 2.1 Representation of a communication process.

example one is not sure about what face will come up in the experiment of tossing a dice, and therefore there is uncertainty as to what result the measurement will produce. After the measurement process one learns the result of the experiment and information is transferred from the physical system (source) to the physicist (receiver) [10]. The informational content of some possible outcome x of a random variable X is defined as [3]:

$$\inf(x) := -\log[p(x)]. \tag{2.1}$$

It is a continuous monotonic decreasing function of the probability of the event $p(X = x)$ and it is additive, i.e., if two events, with outcomes x and x', are independent their information is the sum of the information of each individual event [1]:

$$\begin{aligned} \inf(x, x') &= -\log[p(x, x')] \\ &= -\log[p(x)p(x')] \\ &= -\log[p(x)] - \log[p(x')] \\ &= \inf(x) + \inf(x') \end{aligned}$$

From the definition in Equation (2.1) it is clear that information is not a concept regarding the meaning of a message, but it is a concept concerning all the possible outcomes of the experiment, since there is uncertainty as to the event x only due to the possibility of X taking another value from its alphabet $\mathcal{X}$. It is desirable therefore to define the amount of information contained in the random variable X. Hence, the average information of events works as a measure of uncertainty concerning the experiment [2],

[1] The logarithm is the only continuous function which transforms the product of real numbers into the sum of real numbers, $\log(a.b) = \log(a) + \log(b)$.

$$\mathbb{E}_p \inf(X) = -\sum_x p(x) \log p(x), \tag{2.2}$$

where one set $p(x) \log p(x) := 0$ if $p(x) = 0$ to $\mathbb{E}_p \inf(X)$ be a continuous function, since $\lim_{p(x)\to 0} p(x) \log p(x) = 0$.

The expected value of $\inf(x)$ in equation (2.2) is called *entropy* of the random variable X and denoted as $H(X)$. Since it is a functional of the probability p it is also denoted as $H[p]$ or $H[p(x)]$.

In information theory, uncertainty is a concept properly defined by the following properties [1, 4]:

S1. The uncertainty measure H is a continuous function of the probability distribution p, since a small variation of p do not lead to great uncertainty concerning experiment.

S2. In an experiment with equally likely outcomes, more choice implies more uncertainty. Therefore, the quantity

$$H\left[\frac{1}{n}, \ldots, \frac{1}{n}\right]$$

is a monotonic increasing function of the number n of possible outcomes.

S3.The Composition Law. The entropy H must be additive:

$$H[p(a), p(b), p(c)] = H(p(a), q) + qH\left[\frac{p(b)}{q}, \frac{p(c)}{q}\right]$$

where $q = p(b) + p(c)$.

The following theorem assures that the only information measure of a random experiment satisfying the reasonable axioms above is the Shannon entropy defined as $\mathbb{E}_p \inf(X)$.

Theorem 1. *The only function satisfying the properties S1,S2 and S3 is*

$$H(X) = k \sum_{x\in\mathcal{X}} z_c[p(x)], \tag{2.3}$$

where

$$z_c[a] = \begin{cases} -a \log_c a, & a > 0 \\ 0, & a = 0 \end{cases},$$

where the constant k and the logarithm base c are arbitrary.

The proof of Theorem 1 can be found in [4] and [11], and the proof of the next theorems and propositions may be found in [2].

Information theorists usually set $k := 1$ and $c := 2$, and entropy is measured in *bits*. The reason is that entropy measured in bits is equal to the average number of binary questions necessary to learn the outcome of a random variable (see Chapter 5 in [2]). However, in this study entropy is measured in *nats*, i.e., one set $k := 1$ and $c := e$ ($z_e := z$) due to mathematical convenience.

Definition 1. *The entropy of a probability density $\varrho : \Gamma \to \mathbb{R}_+$ is*

$$H[\varrho] = -\int \varrho(x) \ln \varrho(x) dx,$$

where the integration is performed over the support set of the probability density, $supp(\varrho) := \{x \in \Gamma | \varrho(x) \neq 0\}$.

Quite often more than one quantity is observed in an experiment and therefore it is worth defining information measures of several random variables, since a random vector is also a random variable.

Definition 2. *Let X and Y be random variables with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The joint entropy of X and Y is*

$$H(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} z[p(x, y)],$$

and if (X, Y) is a continuous random vector drawn according to ϱ its entropy is defined as

$$H(X, Y) = \int \varrho(x, y) \ln \varrho(x, y) dx dy.$$

The joint entropy of several random variables is the uncertainty about an experiment where measurements of those random variables are performed simultaneously [2].

Definition 3. *Let X and Y be discrete random variables with joint probability p, the entropy of X conditional to Y is defined as*

$$H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} z[p(x|y)],$$

and if (X, Y) is a continuous with density $\varrho : \Gamma \to \mathbb{R}_+$ it is defined as

$$H(X|Y) = \int \varrho(x, y) \ln \varrho(x|y) dx dy.$$

The entropy of X conditional to Y is the uncertainty about the result of X in an experiment that the result of Y is available [2]. The next proposition validate this interpretation of conditional entropy.

Proposition 1. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.

The definition of information measures in the case of several random variables is analogous to the one presented here, and an important result concerning it is the proposition below generalizing Proposition 1.

Proposition 2 (Chain Rule)**.** *Let $(X_1, \ldots, X_n)$ be a n-dimensional random vector with probability p. Then*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}), \quad n \in \mathbb{N}.$$

Definition 4. *The Kullback-Leibler divergence between a probability measure p and q of a random variable X with finite alphabet $\mathcal{X}$ is defined as*

$$D[p||q] = \sum_{x\in\mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}$$

and the divergence between the probability densities $\varrho : \Gamma \to \mathbb{R}_+$ and $\sigma : \Gamma \to \mathbb{R}_+$ is

$$D[\varrho||\sigma] = \int_\Gamma \varrho(x) \ln \frac{\varrho(x)}{\sigma(x)} dx.$$

The Kullback-Leibler divergence $D[p||q]$ is a measure of distinguishability of the probabilities distributions p and q in the following sense [3]: suppose that in an experiment the actual probability is p, but one believes that it is q, then the difference in the information of a single outcome x is $\ln[p(x)] - \ln[q(x)]$ and the Kullback-Leibler divergence is the average of this difference. The continuous version has the same interpretation and it is also related to irreversibility in Hamiltonian processes [5], as discussed in Chapter 3.

Theorem 2. *The Kullback-Leibler divergence is nonnegative.*

Proof.

$$\ln a \geq 1 - \frac{1}{a} \Rightarrow D[p||q] \geq \sum_x [p(x) - q(x)] = 0. \tag{2.4}$$

The proof for continuous random variables is analogous, and it follows replacing $\sum_x$ by $\int dx\rho(x)$ in (2.4). □

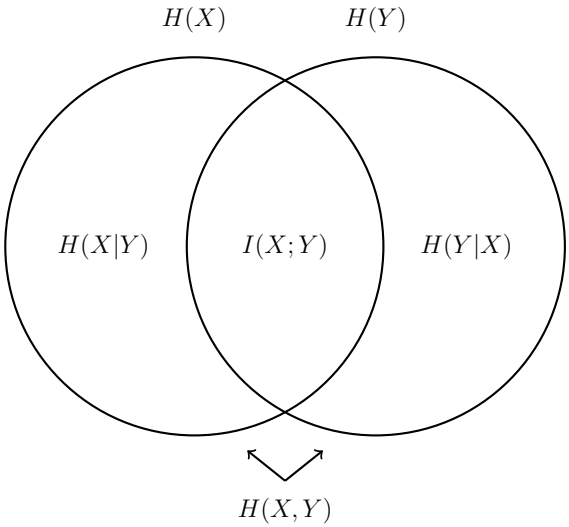

Figure 2.2 In the case of two random variables, X and Y, there is an analogy between information measures and set operations. Replacing information measures by sets, sum plays the role of set union, subtraction plays the role of set intersection, and the Venn diagrams representation still holds. Figure adapted from [2].

Definition 5. *The mutual information of X and Y is defined as*

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

if (X,Y) is discrete and distributed according to p. If (X,Y) is continuous and distributed by ϱ, then their mutual information is defined as

$$I(X;Y) = \int \varrho(x,y) \ln \frac{\varrho(x,y)}{\varrho(x)\varrho(y)} dxdy.$$

As the mutual information of two random variables is equal two the divergence between their joint probability and the distribution for independent variables, the mutual information is a measure of independence of two random variables. The Figure 2.2 shows a relation between two random variables.

2.1.2 Generalized Shannon Entropy

Entropy - the amount of information transferred in a measurement - is the unique information measure of a random variable in agreement with suitable properties. In an actual experiment the physicist collects a lot of data. Suppose that in a simple experiment the physicist learns the outcome of a random variable X performing successive measurements of such quantity, and as a result a sequence of outcomes (x_t)

is gathered, where x_t is the outcome of X at time t. One may be interested in the amount of uncertainty about the hole measurement process, i.e., the process in which (x_t) is acquired. This information measure is called as Generalized Shannon Entropy in agreement with [8], and its definition is discussed in the following.

Before the measurement performed at time t the sequence $(x_n)_{n=1}^{t-1}$ is known and it may affect the ignorance about the next result. Hence, we need to change our definition of uncertainty about an outcome. Considering the previous results $(x_n)_{n=1}^{t-1}$ the uncertainty about the next result is

$$H[p(x_t|x_0,\ldots,x_{t-1})] = \sum_{x_t} z[p(x_t|x_0,\ldots,x_{t-1})],$$

where $p(x_t|x_0,\ldots,x_{t-1})$ is the conditional probability of x_t with respect to $(x_0,\ldots,x_{t-1})$. Therefore, the uncertainty of an outcome averaged over the previous results is defined as

$$H_t(X) := \mathbb{E}_{p(x_0,\ldots,x_{t-1})} H[p(x_t|x_0,\ldots,x_{t-1})] \tag{2.5}$$

$$= \sum_{x_0} \cdots \sum_{x_{t-1}} p(x_0,\ldots,x_{t-1}) \sum_{x_t}^{n} z[p(x_t|x_0,\ldots,x_{t-1})] \tag{2.6}$$

Since there is not an outcome more privileged then other, the uncertainty of a sequence of t results is a simple average of the uncertainty of each result [8]

$$\frac{H_0(X) + \cdots + H_t(X)}{t+1}$$

and the Generalized Shannon Entropy is defined as the uncertainty of a infinite time sequence [8]

$$H_G(X) := \lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} H_n(X) \tag{2.7}$$

The nonnegative quantity $H_G(X)$ is the degree of randomness concerning the information process generated by the source. If $H_G(X) > 0$ it means that the knowledge of all previous results is not sufficient to predict the next one.

2.2 Entropy of Dynamical Systems

A dynamical system is a mathematical description of a scientific model, and it is therefore essential to understand the main properties of it. This study focuses on chaos, a feature of dynamical systems related to unpredictability of time evolution. The physical systems considered in this study are characterized as probability spaces and their time evolution functions must satisfy the conditions bellow.

Definition 6. *Let* (Γ, Σ, p_t) *be a probability space. An automorphism* $\Phi_\tau : \Gamma \to \Gamma$ *is a bijective map such that* Φ_τ *and its inverse* Φ_τ^{-1} *are measurable functions,*

$$\Phi_\tau(A), \Phi_\tau^{-1}(A) \in \Sigma, \qquad \forall A \in \Sigma,$$

and probability preserving maps,

$$p_t(A) = p_{t+\tau}[\Phi_\tau(A)] = p_{t-\tau}[\Phi_\tau^{-1}(A)], \qquad \forall A \in \Sigma.$$

Definition 7. *A Dynamical System* $(\Gamma, \Sigma, p, \phi_t)$ *is a probability space* (Γ, Σ, p) *with a one-parameter group of automorphisms* ϕ_t *depending measurably on* t*.*

In Definitions 6 and 7 the set Γ is the state space, called as phase space. The parameter $t \in \mathbb{R}$ represents time and Φ_τ represents time evolution of τ units. The automorphisms depends measurably on time, i.e.,

$$\phi_t(A) \cap B \in \Sigma, \quad \forall t \ \forall A, B \in \Sigma.$$

In this study time is considered to be discrete, $t \in \mathbb{Z}$, and therefore the time evolution is generated by a unique automorphism ϕ (see figure 2.3). Dynamical systems are denoted by $(\Gamma, \Sigma, p_t, \phi)$ and the last condition in definition 6 is written as $p_t(A) = p_{t+n}[\phi^n(A)] = p_{t-n}[\phi^{-n}(A)], \ \forall t, n \in \mathbb{N}, \forall A \in \Sigma$.

In actual experiments one usually measures average physical quantities such as internal energy, dissipated work etc, and properly defined as follows.

Definition 8. *The time average of an observable* $f : \Gamma \to \Gamma$ *with respect to the initial state* $x \in \Gamma$ *is*

$$\bar{f}(x) = \lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} f[\phi^n(x)], \qquad t \in \mathbb{Z}$$

in the discrete time case, and

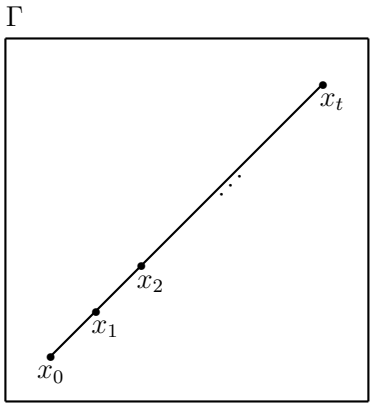

Figure 2.3 Let $(\Gamma, \Sigma, p_t, \phi_t)$ be a dynamical system. It is possible to generate a discrete time evolution as represented in the picture defining a suitable time unit $\tau := 1$. This discrete time dynamical system is denoted as $(\Gamma, \Sigma, p_t, \phi)$, $t \in \mathbb{Z}$.

$$\bar{f}(x) = \lim_{t\to\infty} \frac{1}{t} \int_0^t f[\phi_{t'}(x)]dt', \qquad t \in \mathbb{R}$$

in the continuous time case.

Definition 9. *The phase space average of an observable $f : \Gamma \to \Gamma$ with respect to a probability measure p is*

$$\langle f \rangle_p = \int f dp.$$

2.2.1 Kolmogorov-Sinai Entropy

The Kolmogorov-Sinai entropy is a quantity of dynamical systems quantifying chaos intensity, and its definition consider a coarse-grained analysis of the phase space.

Definition 10. *A finite partition $\mathcal{A}$ of the phase-space Γ is a finite collection of measurable subsets satisfying*

$$\bigcup_{\alpha\in\mathcal{A}} \alpha = \Gamma, \tag{2.8}$$

and

$$\alpha \neq \alpha' \Rightarrow \alpha \cap \alpha' = 0, \ \ \forall \alpha, \alpha' \in \mathcal{A}. \tag{2.9}$$

A finite partition of a phase space generates a collection of discrete symbols, and at each instant of time t the probability vector $[p_t(\alpha)]_{\alpha\in\mathcal{A}}$ is well-defined. The entropy of a partition, as defined bellow, quantifies the uncertainty about the element of

the partition the state of the system belongs, i.e., the information contained in the probability vector $[p_t(\alpha)]_{\alpha\in\mathcal{A}}$.

Definition 11. *The entropy of a finite partition $\mathcal{A}$ with respect to a probability measure p_t is*

$$H(\mathcal{A}) := H(\mathcal{A}, p_t) = \sum_{\alpha\in\mathcal{A}} z[p_t(\alpha)].$$

It is possible in some cases to define infinite partitions, for example if the phase space is the real line $\mathbb{R}$ one may define a finite partition $\{(-\infty, 0), [0, +\infty)\}$, and define an infinite countable partition $\{[n, n+1)|n \in \mathbb{N}\}$ as well. In this study nevertheless one only consider finite partitions, and therefore the terms "finite partition" and "partition" are used interchangeably.

Proposition 3. *If the partition $\mathcal{A}$ has $|\mathcal{A}|$ elements, then*

$$H(\mathcal{A}) \leq \ln|\mathcal{A}|.$$

Proof. Since partition $\mathcal{A}$ has $|\mathcal{A}|$ elements, follows

$$\begin{aligned} H(\mathcal{A}) &= \sum_{\alpha\in\mathcal{A}} z[p_t(\alpha)] && (2.10)\\ &= |\mathcal{A}| \sum_{\alpha} \frac{1}{|\mathcal{A}|} z[p_t(\alpha)] && (2.11)\\ &\leq |\mathcal{A}| z\left[\frac{1}{|\mathcal{A}|}\sum_{\alpha} p_t(\alpha)\right] && (2.12)\\ &= \ln(|\mathcal{A}|) && (2.13) \end{aligned}$$

where the inequality 2.12 is a consequence of the convexity of z, as assured by Jensen's theorem (see Appendix A).

□

The next definition is also in complete analogy to the information-theoretical conditional entropy.

Definition 12 (Conditional entropy)**.** *The entropy of the finite partition $\mathcal{A}$ with respect to the finite partition $\mathcal{B}$ is*

$$H(\mathcal{A}|\mathcal{B}) = \sum_{\beta\in\mathcal{B}} p_t(\beta) \sum_{\alpha\in\mathcal{A}} z[p_t(\alpha|\beta)]. \quad (2.14)$$

Definition 13. *Let $\mathcal{A}$ and $\mathcal{B}$ be finite partitions. Their refinement, denoted as $\mathcal{A} \vee \mathcal{B}$, is the partition*

$$\mathcal{A} \vee \mathcal{B} = \{\alpha \cap \beta | \alpha \in \mathcal{A}, \beta \in \mathcal{B}\}.$$

Defining the partition[2] $\phi(\mathcal{A}) := \{\phi(\alpha) | \alpha \in \mathcal{A}\}$ it is possible to study the time evolution of the partition $\mathcal{A}$, and it is possible to calculate the entropy of the partition $\mathcal{A} \vee \phi(\mathcal{A})$, as explained in the figure 2.4. The lemma 1 is analogous to the chain rule presented in Proposition 2.

Lemma 1. *Let $\mathcal{A}$ be a finite partition. Then,*

$$H[\mathcal{A} \vee \cdots \vee \phi^t(\mathcal{A})] = H[\mathcal{A}] + \sum_{j=1}^{t} H\left[\mathcal{A} \middle| \bigvee_{i=1}^{j} \phi^i(\mathcal{A})\right], \forall t \in \mathbb{Z}.$$

Proof. The first step is to observe that

$$\begin{aligned} H[\mathcal{A} \vee \phi(\mathcal{A})] &= H[\phi(\mathcal{A}) \vee \mathcal{A}] \\ &= H[\phi(\mathcal{A})] + H[\mathcal{A}|\phi(\mathcal{A})] \\ &= H[\mathcal{A}] + H[\mathcal{A}|\phi(\mathcal{A})], \end{aligned}$$

and that

$$\begin{aligned} H[\mathcal{A} \vee \cdots \vee \phi^{t+1}(\mathcal{A})] &= H[\phi(\mathcal{A}) \vee \cdots \vee \phi^{t+1}(\mathcal{A}) \vee \mathcal{A}] \\ &= H[\phi(\mathcal{A}) \vee \cdots \vee \phi^{t+1}(\mathcal{A})] + H[\mathcal{A}|\phi(\mathcal{A}) \vee \cdots \vee \phi^{t+1}(\mathcal{A})] \\ &= H[\mathcal{A} \vee \cdots \vee \phi^t(\mathcal{A})] + H[\mathcal{A}|\phi(\mathcal{A}) \vee \cdots \vee \phi^{t+1}(\mathcal{A})], \end{aligned}$$

since ϕ is measure preserving and preserves set intersection.

From this, it follows the relation:

[2] Since ϕ is an automorphism,

$$\phi(\alpha) \in \Sigma, \qquad \forall \alpha \in \mathcal{A}, \tag{2.15}$$

and

$$\alpha \neq \alpha' \Rightarrow \phi(\alpha) \neq \phi(\alpha') \Rightarrow \phi(\alpha) \cap \phi(\alpha') = \phi(\alpha \cap \alpha') = \phi(\emptyset) = \emptyset, \qquad \alpha, \alpha' \in \mathcal{A}. \tag{2.16}$$

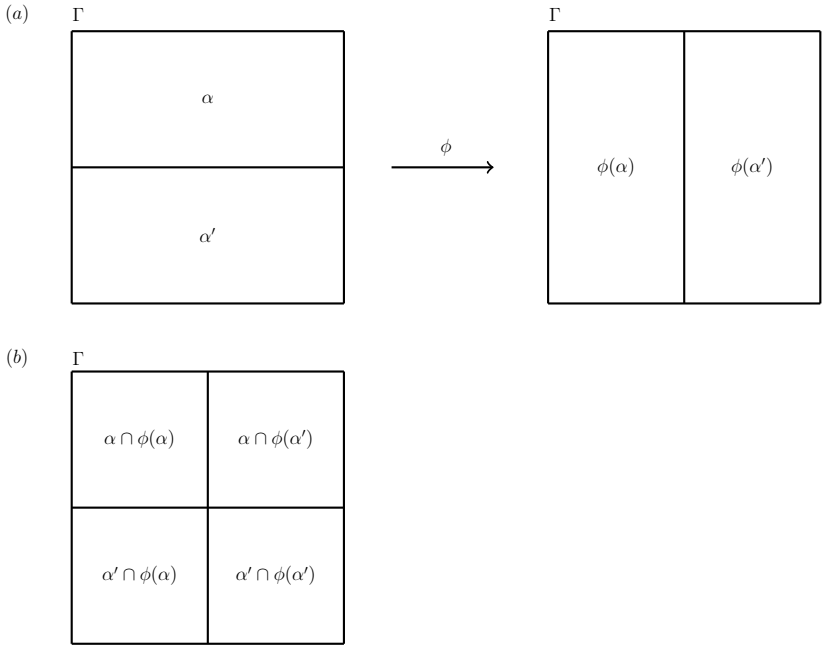

Figure 2.4 Time evolution in the coarse-grained phase space scenario: (a) considering a partition $\mathcal{A} = \{\alpha, \alpha'\}$ of the phase space Γ, the partition $\phi(\mathcal{A}) = \{\phi(\alpha), \phi(\alpha')\}$ shows how the partition α behaves under time evolution; (b) It is possible to determine the transitions between cells taking the refinement $\mathcal{A} \vee \phi(\mathcal{A})$, for example $\alpha' \cap \phi(\alpha)$ is the set of states which transitioned from α to α' . It is always possible to calculate the entropy of $\mathcal{A} \vee \phi(\mathcal{A})$ since ϕ depends measurably in time. Figure adapted from [8].

$$H[\mathcal{A} \vee \cdots \vee \phi^t(\mathcal{A})] = H[\mathcal{A}] + \\ + \sum_{j=1}^{t} H\left[\mathcal{A}\middle|\bigvee_{i=1}^{j}\phi^i(\mathcal{A})\right] \Rightarrow H[\mathcal{A} \vee \cdots \vee \phi^{t+1}(\mathcal{A})] = H[\mathcal{A}] + \sum_{j=1}^{t+1} H\left[\mathcal{A}\middle|\bigvee_{i=1}^{j}\phi^i(\mathcal{A})\right].$$

□

Definition 14. *The entropy rate of a finite partition $\mathcal{A}$ with respect to an automorphism ϕ is*

$$h(\mathcal{A}, \phi) = \lim_{t\to\infty} \frac{H(\mathcal{A} \vee \phi(\mathcal{A}) \vee \cdots \vee \phi^{t-1}(\mathcal{A}), p_t)}{t}. \tag{2.17}$$

Definition 15. *The entropy of an automorphism ϕ is [6, 7]*

$$h(\phi) = \sup_{\mathcal{A}\in\mathcal{F}} h(\mathcal{A}, \phi) \tag{2.18}$$

where $\mathcal{F}$ is the set of all finite partitions of the phase space.

The entropy of an automorphism $h(\phi)$ is usually called the Kolmogorov-Sinai entropy of the dynamical system $(\Gamma, \Sigma, p_t, \phi)$. As discussed before, $h(\phi)$ is a parameter of chaos intensity, from which follows the definition 16 that will be justified by the Frigg's equivalence theorem.

Definition 16. *The dynamical system $(\Gamma, \Sigma, p_t, \phi)$ is chaotic if and only if*

$$h(\phi) > 0.$$

2.2.2 Generalized Shannon Entropy of Dynamical Systems

Unpredictability is a concept in dynamical systems theory quantified by Kolmogorov-Sinai entropy. This is the statement of Frigg's Equivalence Theorem [8]: Kolmogorov-Sinai entropy is a measure of randomness in the sense that it is equal to the Generalized Shannon Entropy, a result presented in detail here.

Outcomes of a source have been defined as elements of an alphabet, which is a finite collection of symbols. The equivalent concept to alphabets in dynamical systems theory are the partitions, since the phase space may be a continuous space, as example the phase space of classical systems presented in Chapter 3. The trajectory of an initial state generates a sequence of elements of the partition, as explained in Figure 2.5(a). The sequence (α_t) is equivalent to a sequence of outcomes of a source (x_t).

The entropy of a partition $H(\mathcal{A}, p_t)$ is the amount of information transferred as a consequence of a measurement with precision defined by $\mathcal{A}$ at time t. As before, one needs to consider the past measurements, since it affects the ignorance about the result. The procedure is exactly the same as done in the last section, and therefore one needs to calculate the probabilities of sequences (α_t). It is discussed in Figure 2.5(b).

The probability of finding the state in α_t, at time t, conditional to the sequence $(\alpha_0, \ldots, \alpha_{t-1})$ is

$$\begin{aligned} p(\alpha_t|\alpha_0, \ldots, \alpha_{t-1}) &:= \frac{p(\alpha_0, \ldots, \alpha_{t-1}, \alpha_t)}{p(\alpha_0, \ldots, \alpha_{t-1})} \\ &= \frac{p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1}) \cap \alpha_t]}{p_{t-1}[\phi^{t-1}(\alpha_0) \cap \cdots \cap \alpha_{t-1}]}. \end{aligned} \tag{2.19}$$

Following the same ideas as before one can calculate the averaged uncertainty over the previous results of a result, defined as

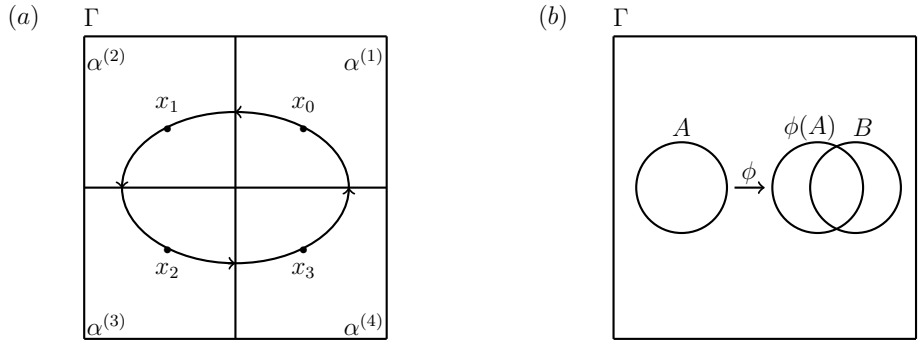

Figure 2.5 The sequence (x_t) defines the path in the coarse-grained phase space: (a) considering the partition $\mathcal{A} = \{\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)}\}$ and a initial state $x_0 \in \alpha^{(1)}$, at time $t = 1$ the state of the system is $x_1 \in \alpha^{(2)}$, at time $t = 2$ it is $x_2 \in \alpha^{(3)}$, and at time $t = 3$ it is $x_3 \in \alpha^{(4)}$. The sequence (x_0, x_1, x_2, x_3) generates the path $(\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)})$ in the coarse-grained phase space. (b) The probability of a path (A, B) in the coarse-grained phase space, $p(A, B)$, is equal to the probability measure of the intersection of $\phi(A)$ and B at time $t = 1$, $p_t[B \cap \phi(A)]$. Figure adapted from [8].

$$
\begin{aligned}
H_t(\mathcal{A}) &:= \sum_{\alpha_0} \cdots \sum_{\alpha_{t-1}} p(\alpha_0, \ldots, \alpha_{t-1}) \sum_{\alpha_t} z[p(\alpha_t | \alpha_0, \ldots, \alpha_{t-1})] \\
&= \sum_{\alpha_0} \cdots \sum_{\alpha_{t-1}} p_{t-1}[\phi^{t-1}(\alpha_0) \cap \cdots \cap \alpha_{t-1}] \sum_{\alpha_t} z \left[\frac{p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1}) \cap \alpha_t]}{p_{t-1}[\phi^{t-1}(\alpha_0) \cap \cdots \cap \alpha_{t-1}]} \right].
\end{aligned}
$$

Lemma 2.

$$H_t(\mathcal{A}) = H[\mathcal{A} | \phi(\mathcal{A}) \vee \cdots \vee \phi^t(\mathcal{A})].$$

Proof. Since the map ϕ is measure preserving and a one-to-one map preserves intersection, i.e., $\phi(A \cap B) = \phi(A) \cap \phi(B)$ $(A, B \subset \Gamma)$, it follows that

$$p_{t-1}[\phi^{t-1}(\alpha_0) \cap \cdots \cap \alpha_{t-1}] = p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1})],$$

and therefore

$$H_t(\mathcal{A}) = \sum_{\alpha_0} \cdots \sum_{\alpha_{t-1}} p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1})] \sum_{\alpha_t} z \left[\frac{p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1}) \cap \alpha_t]}{p_t[\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1})]} \right].$$

The sets $\phi^t(\alpha_0) \cap \cdots \cap \phi(\alpha_{t-1})$ are the elements of the partition $\phi(\mathcal{A}) \vee \cdots \vee \phi^t(\mathcal{A})$, and considering the definition of conditional entropy finishes the demonstration.

□

The uncertainty of a sequence $(\alpha_0, \ldots, \alpha_{t-1})$ is defined as

$$\frac{H_0(\mathcal{A}) + \cdots + H_{t-1}(\mathcal{A})}{t},$$

and the Generalized Shannon entropy of the dynamical system with respect to the partition $\mathcal{A}$ is

$$H_G(\phi, \mathcal{A}) := \lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} H_n(\mathcal{A}). \tag{2.20}$$

Definition 17. *The Generalized Shannon entropy of the dynamical system is defined as*

$$H_G(\phi) = \sup_{\alpha\in\mathcal{F}} H_G(\phi, \mathcal{A}) \tag{2.21}$$

where $\mathcal{F}$ is the set of all finite partitions of the phase space.

The Frigg's equivalence theorem follows directly from lemmas 1 and 2.

Theorem 3 (Frigg's equivalence theorem [8])**.** *The Kolmogorov-Sinai entropy of a dynamical system is equal to its Generalized Shannon Entropy:*

$$h(\phi) = H_G(\phi).$$

Chapter 3

Information theory and thermodynamics

3.1 Stochastic processes

Systems that time evolution drives randomly initial states to final states are called stochastic systems. Stochastic processes are ubiquitous in nature (see [9] and references therein), hence, its study is crucial. Chapter 2 highlighted an example: partitioning the phase space of a dynamical system generates a stochastic system, since starting from a particular α_0, time evolution could drives it to several possible final regions α_t. Therefore, the states of the system must be represented as random variables.

Definition 18. *Let (Γ, Σ, p) be a probability space. A stochastic process is a sequence of random variables defined on the sample space Γ, and denoted as $(X_t)_{t\in T}$.*

In definition 18, the index t represents time. The process is called continuous time stochastic process if the set T, representing the duration of the process, is a interval of $\mathbb{R}$, and it is called discrete time stochastic process if T is a subset of $\mathbb{Z}$. The random variable X_t represents the state of the system at time t.

Given a stochastic process $(X_t)_{t\in\mathbb{Z}_+}$, and denoting the alphabet of its random variables as $\mathcal{X}$, one call it a Markov process if the probability of the immediate future state X_{t+1} depends conditionally only on the current state X_t, i.e.,

$$(X_t)_{t\in\mathbb{N}} \ is \ a \ Markov \ process \ . = . \ p(x_{t+1}|x_0, \ldots, x_t) = p(x_{t+1}|x_t), \quad \forall t, \ \forall x_t.$$

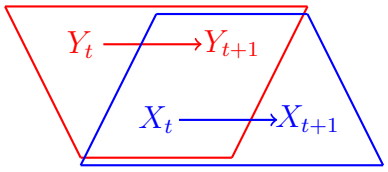

Figure 3.1 During the work step $Y_t \to Y_{t+1}$ the system is enclosed by adiabatic walls, and the difference in energy is equal to the work done by the external agent. The system is then allowed to a relaxation step, enclosed by diathermal walls, and remaining the external parameter fixed. Figure adapted from [9].

Stochastic thermodynamics is the branch of physics dealing with thermodynamic properties of stochastic processes such as fluctuation relations and entropy production (see [12] and references therein). It is the study of stochastic systems with well-defined physical quantities such as energy, and therefore physical laws are usually considered, for example the conservation of energy in closed systems.

This section discusses the connection between unpredictability and thermodynamic inefficiency presented by Still et al. [9], and the considered stochastic process is described as follows.

A stochastic system with state-space $\mathcal{X}$ and a heat reservoir with inverse temperature $\beta := 1/k_B T$ are initially in equilibrium. The system's state at time t is a random variable X_t taking values in $\mathcal{X}$. An external agent drives the system stochastically and represented at each instant of time t by a random variable Y_t, called external variable, taking values in the alphabet $\mathcal{Y}$. Energy, which is also a random variable, is a well-defined quantity of the model expressed as $E(X_t, Y_t)$.

At each moment of the discrete time stochastic process $(X_t, Y_t)_{t\in\mathbb{Z}_+}$ there is two relevant steps called *work step* and *relaxation step*. In the work step, denoted as $Y_t \to Y_{t+1}$, the action of the external agent transfers an amount of energy to the system equals to $W[y_t \mapsto y_{t+1}|x_t] := E(x_t, y_{t+1}) - E(x_t, y_t)$, where $E(x_t, y_t)$ represents the energy of the system with state X_t taking the value x_t and external variable Y_t taking the value y_t. The external parameter change, $y_t \mapsto y_{t+1}$, drives the system to a new state in a relaxation step, denoted as $X_t \to X_{t+1}$, in which an amount of energy $Q[x_{i_t} \mapsto x_{i_{t+1}}|y_{t+1}] := E(x_{t+1}, y_{t+1}) - E(x_t, y_{t+1})$ is transferred to the reservoir. See figure 3.1 for a representation of $(X_t, Y_t)_{t\in\mathbb{Z}_+}$. The dynamics is assumed to be Markovian, i.e.,

$$p(x_{t+1}|y_0, x_0, \ldots, y_t, x_t, y_{t+1}) = p(x_{t+1}|x_t, y_{t+1}), \quad \forall t \in \mathbb{Z}_+, \ \forall x_t \in \mathcal{X}, \ \forall y_t \in \mathcal{Y}.$$

There is no physical law which determines the initial probability distribution of a stochastic process, it depends on the problem to be solved. Information theory nevertheless establishes a criterion for assigning a probability to a random variable X called Maximum-Entropy method [4], and described as follows. The probability $p_{X|Y=y}$ must be consistent with the normalization condition,

$$\sum_{x\in\mathcal{X}} p(x|y) = 1, \qquad \forall y \in \mathcal{Y}, \tag{3.1}$$

and since the system and the reservoir are in equilibrium with internal energy U, the probability distribution must be consistent with

$$\sum_{x\in\mathcal{X}} p(x|y)E(x,y) = U, \qquad \forall y \in \mathcal{Y}, \tag{3.2}$$

Conditions (3.1) and (3.2) generally do not determine a unique probability distribution. The Maximum-Entropy method is concerned with the maximization of information theory entropy H constrained to all prior information [4]; conditions (3.1) and (3.2) must be considered in this case as the system is in equilibrium with the heat reservoir. Let $\lambda - 1$ and β be Lagrangian multipliers. Necessary conditions to this variational problem can be stated as

$$\frac{d}{dp(x|y)}\left\{H[p_{X|y}] - (\lambda-1)\left[\sum_{x'\in\mathcal{X}} p(x'|y) - 1\right] - \beta\left[\sum_{x'\in\mathcal{X}} p(x'|y)E(x',y) - U\right]\right\} = 0.$$

Taking the derivatives, it follows that

$$p_{eq}(x|y) = e^{-\lambda-\beta E(x,y)}. \tag{3.3}$$

Substituting (3.3) into (3.1) we find the value of the Lagrange multiplier λ,

$$\sum_x p_{eq}(x|y) = e^{-\lambda}\sum_x e^{-\beta E(x,y)} = 1.$$

Defining the partition function as $Z(\beta, y) := \sum_x e^{-\beta E(x,y)}$, it follows that

$$\lambda = \ln Z. \tag{3.4}$$

From (3.3) and (3.2) follows the equation which allows to find β,

$$\langle E(X,y)\rangle = \frac{1}{Z}\sum_x E(x,y)e^{-\beta E(x,y)} = -\frac{\partial}{\partial\beta}\ln Z. \tag{3.5}$$

Proposition 4. *The equilibrium distribution (3.3) is the unique distribution which maximizes information theory entropy constrained to (3.1) and (3.2) [4].*

Proof. Let p and p_{eq} be probability distributions of the random variable X. It follows,

$$D[p||p_{eq}] = -H[p] - \sum_x p(x|y)\ln p_{eq}(x|y) \geq 0, \tag{3.6}$$

$$\therefore H[p] \leq \ln Z + \beta\langle E(X,y)\rangle_p \tag{3.7}$$

The equality holds iff $p = p_{eq}$. □

Defining the thermodynamic entropy as $S(U) := k_B H[p_{eq}]$, it follows from inequality (3.7) that the Lagrange multiplier β is identified as the inverse temperature,

$$\beta = \frac{\partial S}{\partial U} := \frac{1}{k_B T}. \tag{3.8}$$

Therefore, equilibrium probability distributions are the less unbiased assignments to random variables, since it is the probability which maximizes uncertainty.

Describing the system by an equilibrium distribution at time $t = 0$ implies that it has well defined thermodynamic quantities, such as thermodynamic potentials. The free energy is defined as a function of an equilibrium probability distribution,

$$F(y) := \langle E(x,y)\rangle_{p_{eq}(x|y)} + k_B T H[p_{eq}(y|x)]. \tag{3.9}$$

In general the system is driven out of equilibrium by the external agent, and it is worthy to define a mathematical quantity characterizing non-equilibrium probability distributions. In this spirit, the generalized free energy is defined as a function of general probability distributions remaining fixed the mathematical form of free energy.

Definition 19. *The generalized free energy is*

$$F_G[p(x|y)] = \langle E(x,y)\rangle_{p(x|y)} + k_B T H[p(x|y)]. \tag{3.10}$$

Note that $F(y) = F_G[p_{eq}(x|y)]$. A feature of definition 19 is the proposition bellow.

Proposition 5. *The generalized free energy is equal to the free energy plus an additional term proportional to the Kullback-Leibler divergence between the actual probability distribution and the equilibrium probability distribution,*

$$F_G[p(x_t|y_t)] = F(y_t) + F_t^{add}[p(x_t|y_t)],$$

where $\beta F_t^{add}[p(x_t|y_t)] = D[p(x_t|y_t)||p_{eq}(x_t|y_t)]$.

Proof.

$$\begin{aligned} D[p(x_t|y_t)||p_{eq}(x_t|y_t)] &= \sum_{x_t \in \mathcal{X}} p(x_t|y_t) \ln p(x_t|y_t) - \sum_{x_t \in \mathcal{X}} p(x_t|y_t) \ln p_{eq}(x_t|y_t) \\ &= -H[p(x_t|y_t)] + \beta \langle E(x_t, y_t) \rangle_{p(x_t|y_t)} - \beta F[y_t] \\ &= \beta F_G[p(x_t|y_t)] - \beta F[y_t]. \end{aligned}$$

□

Energetic inefficiency of a process is defined as the amount of work performed on the system in excess to the difference in generalized free energy, $W_{diss} := W - \Delta F_G$, and called as generalized dissipated work [9]. This is not a usual definition, since it is more common defining it as the amount of work performed on the system in excess to the difference in equilibrium free energy [13, 5]. The instantaneous memory kept by system's states about the driving signal at time t is $I_t^{mem} := I(S_t; X_t)$, and its instantaneous predictive information is $I_t^{pred} := I(S_t; X_{t+1})$. Hence, the amount of instantaneous non-predictive information $I_t^{np} := I_t^{mem} - I_t^{pred}$, is defined as model inefficiency. The lemma 3 is a statement of the equivalence between energetic inefficiency and model inefficiency during a work step.

Lemma 3. *The average dissipated work is proportional to the instantaneous non-predictive information in each work step* $Y_t \to Y_{t+1}$*,*

$$\beta \langle W_{diss}[x_t \mapsto x_{t+1}|s_t] \rangle = I_t^{np}.$$

Proof.

$$\begin{aligned} I_t^{np} &:= I_t^{mem} - I_t^{pred} \\ &= I(S_t; X_t) - I(S_t; X_{t+1}) \\ &= H(S_t|X_{t+1}) - H(S_t|X_t) \\ &= \beta[\langle E(s_t, x_{t+1}) \rangle_{p(s_t, x_{t+1})} - \langle F_G[p(s|x_{t+1})] \rangle_{p(x_{t+1})}] \\ &- \beta[\langle E(s_t, x_t) \rangle_{p(s_t, x_t)} - \langle F_G[p(s|x_t)] \rangle_{p(x_t)}] \\ &= \beta \langle W[x_t \mapsto x_{t+1}|s_t] \rangle - \beta \langle \Delta F_G[x_t \mapsto x_{t+1}] \rangle \\ &= \beta \langle W_{diss}[x_t \mapsto x_{t+1}|s_t] \rangle. \end{aligned}$$

□

Theorem 4. *The total nostalgia defined as the total non-predictive information,* $I^{np} = \sum_t I_t^{np}$*, is a lower bound for the total average dissipation,*

$$\beta\langle W_{diss}\rangle \geq I^{np}.$$

The proof of theorem 4 can be found in [9].

3.2 Classical statistical mechanics

The physical state of a classical system at time t is well-defined by its generalized coordinates, $q_t^1, \ldots, q_t^f$, and conjugated momenta, $p_t^1, \ldots, p_1^f$ [14], the vector

$$x_t = (q_t^1, \ldots, q_t^f, p_t^1, \ldots, p_1^f), \tag{3.11}$$

where $f \in \mathbb{N}$ is the number of degrees of freedom of the system.

As the generalized coordinates and conjugated momenta are real valued functions of the time,

$$q^i, p^i : \mathbb{R} \to \mathbb{R}, \qquad i = 1, \ldots, f,$$

the phase space of a classical system is a subset of the $2f$ dimensional euclidean space, $\Gamma \subset \mathbb{R}^{2f}$.

The dynamics in the phase space is *deterministic*, which means that starting from a initial state, time evolution could only drives it to a unique final state. It is determined by the Hamiltonian map of the system, a function of the phase space vector x and time. In this study, the Hamiltonian depends on time only due to an external parameter y, and it is also restricted to the case of where the Hamiltonian is the energy of the system, denoting it as $E(x, y)$. The equations of motion take the form [14]

$$\dot{q^i} = \frac{\partial E(x,y)}{\partial p^i}, \tag{3.12}$$

$$\dot{p^i} = -\frac{\partial E(x,y)}{\partial q^i}. \tag{3.13}$$

The time evolution of every initial state x_0 is determined by the solution of the *Hamilton's equations* (3.13) and (3.13) , $\phi : \Gamma \times \mathbb{R} \to \Gamma$. The mapping $x_t = \phi_t(x_0)$

is called trajectory of the initial state x^0, and as the dynamics is deterministic the trajectory of distinct initial states cannot cross,

$$x_0 \neq z_0 \Rightarrow x_t \neq z_{t'}, \qquad \forall x_0, z_0 \in \Gamma, \forall t, t' \in \mathbb{R}. \tag{3.14}$$

This exact description of an initial state x_0 of a classical system with Hamiltonian map E is called *classical mechanics*. The initial state of a classical system, nevertheless, may not be available to the physicist, for example as a consequence of experimental precision of its apparatus. In this cases, it is well worth considering an average description of the system provided by *classical statistical mechanics*. The method consists of consider a large collection of distinct systems with same physical structure called as *ensemble* [15], a collection of distinct initial states of a classical system characterized by a Hamiltonian map and subject to the physical constraints of the problem. The quantity of points in the ensemble must be so large such that it reaches the continuum limit with a continuous density of points, $n : \Gamma \times \mathbb{R} \to \mathbb{R}$. The quantity of points in the ensemble at each instant of time t must be finite,

$$N = \int_\Gamma n_t(x) dx, \tag{3.15}$$

where dx is the infinitesimal phase space volume measure, $dx = dq_1 \dots dq_f dp_1 \dots dp_f$.

The ensemble points are distributed in the phase space such that it is possible to directly relate its density n with a probability density $\varrho : \Gamma \times \mathbb{R} \to \mathbb{R}$:

$$\varrho_t(x) := \frac{n_t(x)}{N} \tag{3.16}$$

The probability of finding the system's state in a region A of the phase space can be calculated from the statistical method,

$$p_t[A] = \int_A dx \varrho_t(x), \qquad A \subset \Gamma. \tag{3.17}$$

Definition 20. *A classical dynamical system is the collection $(\Gamma, \Sigma, p_t, \phi_t)$, where Γ is a subset of the $2f$-dimensional euclidean space, p_t is a probability measure defined by a continuous density, and ϕ_t is a one-parameter group defined by the solution of the Hamilton's equations.*

An important property of classical dynamical systems is announced by the Liouville's theorem [15, 5, 13].

Theorem 5. *Liouville's theorem. Classical dynamical systems preserve the volume measure,*

$$v[A] = v[\phi_t(A)], \qquad \forall A \subset \Gamma, \forall t \in \mathbb{R},$$

and probability density along any trajectory,

$$\varrho_t(x_t) = \varrho_0(x_0).$$

An important class of states are called equilibrium states, since once reached it the system never changes its state spontaneously. The theory describing this special feature of classical systems is called classical statistical mechanics of equilibrium. The ensemble method as described above equipped with the postulate 1 is the Maximum-Entropy method applied to classical physics and reproduces the usual description of statistical mechanics of equilibrium [4].

Postulate 1. *The equilibrium probability density ϱ^{eq} of a classical statistical system maximizes its information theory entropy subject to the physical constraints.*

The Maximum-Entropy approach to equilibrium statistical mechanics is analogous to the one described in section 3.1, since it is necessary to maximize the information theory entropy $H[\rho]$ constrained to all available information; in the case of a system and a heat reservoir with temperature T in equilibrium, its internal energy U is fixed. The solution to this variational problem is given by

$$\frac{\delta}{\delta\varrho(x)}\left\{H[\varrho] - (\lambda - 1)\left[\int_\Gamma \varrho(x')dx' - 1\right] - \beta\left[\int_\Gamma E(x', y)\varrho(x')dx' - U\right]\right\} = 0, \quad (3.18)$$

where $\lambda - 1$ and β are Lagrangian multipliers.

Taking the functional derivative [1], it follows,

[1] Note that $\frac{\delta\varrho(x')}{\delta\varrho(x)} = \delta(x - x')$, where $\delta(x - x')$ is the Dirac delta distribution. To more detail on functionals [14] is recommended.

$$\begin{aligned}
\frac{\delta}{\delta\varrho(x)}\left\{H[\varrho]-(\lambda-1)\left[\int_\Gamma \varrho(x')dx'-1\right]-\beta\left[\int_\Gamma E(x',y)\varrho(x')dx'-U\right]\right\} &= \\
&= \frac{\delta H[\varrho]}{\delta\varrho(x)}-(\lambda-1)\frac{\delta}{\delta\varrho(x)}\int_\Gamma \varrho(x')dx'-\beta\frac{\delta}{\delta\varrho(x)}\int_\Gamma E(x',y)\varrho(x')dx \\
&= -\int_\Gamma\left[\frac{\delta\varrho(x')}{\delta\varrho(x)}\ln\varrho(x')+\frac{\delta\varrho(x')}{\delta\varrho(x)}\right]dx'-(\lambda-1)\int_\Gamma\frac{\delta\varrho(x')}{\delta\varrho(x)}dx'+ \\
&\quad -\beta\int_\Gamma E(x',y)\frac{\delta\varrho(x')}{\delta\varrho(x)}dx \\
&= -\ln\varrho(x)-\lambda-\beta E(x,y) \\
&= 0,
\end{aligned}$$

$$\therefore \varrho^{eq}(x) = e^{-\lambda-\beta E(x,y)}. \tag{3.19}$$

The λ parameter is determined by the normalization condition,

$$\int_\Gamma \varrho^{eq}(x)dx = e^{-\lambda}\int_\Gamma e^{-\beta E(x,y)}dx = 1.$$

Defining the partition function, $Z(\beta,y) := \int_\Gamma e^{-\beta E(x,y)}dx$, it follows that

$$\lambda = \ln Z. \tag{3.20}$$

The value of β can be determined from the internal energy condition,

$$U = \int_\Gamma E(x,y)\varrho(x)dx = -\frac{\partial}{\partial\beta}\ln Z(\beta,y). \tag{3.21}$$

Proposition 6. *The procedure in agreement with postulate 1 determines a unique solution.*

Proof.

$$D[\varrho||\varrho^{eq}] \geq 0 \Rightarrow H[\rho] \leq \ln Z + \beta\langle E(x,y)\rangle_\rho.$$

□

This study deals with chaos in a very specific process on classical systems, in accordance with [5, 13] and defined in the following. In our model, the system is initially in canonical thermodynamical equilibrium with a heat reservoir with inverse

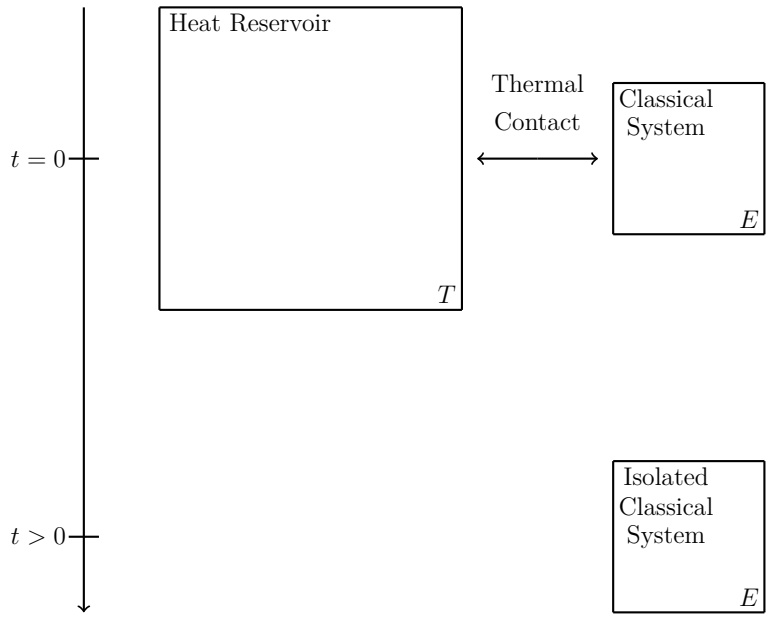

Figure 3.2 The classical system, whose Hamiltonian is E, is allowed to weakly interact with a heat reservoir at temperature T. This thermal interaction, which only allows a change of energy of the system due to heat transfer, is maintained a sufficiently long time such that the system reaches equilibrium with the reservoir. At this time, which is set as $t = 0$, the probability density of the system is given by $\varrho_0^{eq} = e^{-\beta E}/Z$. After $t = 0$ the system is isolated and its energy changes only due to work done by performing a change in the external parameter y. At any instant the probability density evolves accordingly to Liouville's theorem, $\varrho_t(s^t) = \varrho_0^{eq}(s^0)$.

temperature $\beta = 1/k_B T$. Hence, the initial density of states in the phase space is given by

$$\varrho_0(x) = \varrho_0^{eq}(x) = \frac{\exp\{-\beta E(x, y_0)\}}{Z(\beta, y_0)}. \tag{3.22}$$

All thermodynamic information about the macroscopic system is contained in its thermodynamic potentials, for example its thermodynamic entropy identified as the information theory entropy of the equilibrium probability density,

$$S(\langle E \rangle) := k_B H[\varrho_t^{eq}] \tag{3.23}$$

For $t > 0$ the system is isolated [2] and work is done on the system, represented by a change in the external parameter y, in such a way that one can drive the system out of equilibrium. Figure 3.2 represents this process.

[2] In agreement with [5] a system is called isolated when its internal energy changes only due to work done by the external agent. It is usually called a adiabatic process in thermodynamics literature.

Theorem 6 (Kawai-Parrondo-Van den Broeck theorem [5])**.** *The accumulated dissipated work up to time t,*

$$\langle W_{diss}\rangle_t := \langle W\rangle_t - \Delta F,$$

where $\langle W\rangle_t$ *is the average work done on the system and* $\Delta F = F(y_t) - F(y_0)$ *is the difference in the free energy* $F = -k_B T \ln Z$ *is*

$$\langle W_{diss}\rangle_t = k_B T D[\varrho_t||\varrho_t^{eq}].$$

Proof. Defining a generalized free-energy similarly as done in [9],

$$F_G[\varrho_t] = \int E(x, y_t)\varrho_t(x)dx + k_B T \int \varrho_t(x)\ln\varrho_t(x)dx, \tag{3.24}$$

consequently the generalized free energy can be written as the equilibrium free-energy of reference, $F(y_t) = -k_B T \ln Z(y_t)$, plus to an additional term F^{add}:

$$\begin{aligned} F^{add}[\varrho_t] &:= k_B T D[\varrho_t||\varrho_t^{eq}] \\ &= k_B T \left(\int \varrho_t(x)\ln\varrho_t(x)dx - \int \varrho_t(x)\ln\left[\frac{exp\{-\beta E(x,y_t)\}}{Z(y)}\right]dx \right) \\ &= \int E(x,y_t)\varrho_t(x)dx + k_B T\int \varrho_t(x)\ln\varrho_t(x)dx + k_B T\ln Z(y_t) \\ &= F_G[\varrho_t] - F(y_t) \end{aligned}$$

$$\therefore F_G[\varrho_t] = F(y_t) + F^{add}[\varrho_t] \tag{3.25}$$

It is clear from its definition that the generalized free-energy of an equilibrium probability density is equal to the equilibrium free-energy, and F^{add} vanishes in this case.

The generalized free energy difference in the process is

$$\begin{aligned} \Delta F_G &= F_G[\varrho_t] - F_G[\varrho_0] && (3.26) \\ &= \int E(x,y_t)\varrho_t(x)dx + k_B T\int \varrho_t(x)\ln\varrho_t(x)dx \\ &- \int E(x,y_0)\varrho_0(x)dx - k_B T\int \varrho_0(x)\ln\varrho_0(x)dx. && (3.27) \end{aligned}$$

Therefore, the equation (3.26) may be written as

$$H[\varrho_t] - H[\varrho_0] = \beta \left[\langle W \rangle_t - \Delta F_G \right]. \tag{3.28}$$

At initial time the system and the heat reservoir are in thermodynamic equilibrium, therefore, according to equation (3.23) the entropy $H[\varrho_0]$ is identified as the thermodynamic entropy of the system. The changes in the external parameter may drive the system out of equilibrium, and the differential entropy $H[\varrho_t]$ is not necessarily the thermodynamic entropy of the system, because the equilibrium probability density at final time is given by $\varrho_t^{eq}(x) = exp\{-\beta E(x, y_t)\}/Z(y_t)$, which may be different from $\varrho_t(x)$.

The differential entropy of a probability density, due to the Liouville's theorem, is invariant under time evolution, i.e., $H[\varrho_t] = H[\varrho_0]$ (see page 5 in [16]). Therefore the equation (3.28) implies the dissipation theorem [5]:

$$\begin{aligned} \beta \langle W_{diss} \rangle_t = \beta \left[\langle W \rangle_t - \Delta F \right] &= \beta F^{add}[\varrho_t] \\ &= D[\varrho_t || \varrho_t^{eq}]. \end{aligned}$$

□

Chapter 4

Results

The previous chapters discussed the realm of unpredictability and classical physics. It has been made clear in chapter 3 the connection between unpredictability, represented by non-predictive information, and thermodynamic inefficiency, represented by generalized dissipated work. This raises the question whether model unpredictability and energetic inefficiency are connected in the framework of classical physics. The goal of this chapter is to discuss such issue.

It is possible to introduce a discrete time stochastic process from a classical dynamical system $(\Gamma, \Sigma, p_t, \phi)$. Figure 2.5 demonstrate how a discrete time model, called discrete time classical dynamical system, emerges from a continuous one. Letting $\mathcal{A}$ be a partition of Γ, then it is possible to define a sequence of state random variables

$$\begin{aligned} &X_t : \Gamma \to \mathcal{A}, \\ &\quad x_t \in \alpha_t \Leftrightarrow X_t(x_t) = \alpha_t. \end{aligned} \tag{4.1}$$

Random variables are usually defined as maps taking values in $\mathbb{R}$, as done in appendix A. It is always possible to label the partition $\mathcal{A}$ according to an index set $I_{|\mathcal{A}|} := \{1, \ldots, |\mathcal{A}|\}$; the element of the partition which the state x_t belongs would be recognized as α_{i_t}, for some $i_t \in I_{|\mathcal{A}|}$, and the state random variable should be defined by the rule $x_t \in \alpha_{i_t} \mapsto i_t$. It is indeed a very common notation in dynamical systems literature [17]. However, letting implicit the element of the partition which x_t belongs, and denoting it as α_t, do not make any difference to the aims of this study and implies a much simpler notation: $p(X_t = \alpha_t) := p_t(\alpha_t)$.

In classical physics, model unpredictability is usually quantified by Kolmogorov-Sinai entropy. The aim of this chapter is to establish a connection between Kolmogorov-Sinai

entropy and dissipated work, which is often the energetic inefficiency quantifier. It is done in two steps: lemma 4 announces a lower bound on chaos and then it is used to determine a lower bound on energetic inefficiency.

Lemma 4. *Let* $(\Gamma, \Sigma, p_t, \phi)$ *be a discrete time classical dynamical system, and* $\mathcal{A}$ *an arbitrary finite partition of the finite volume phase space* Γ*. Then,*

$$h(\phi) \geq \overline{H[\varrho_t]} + \overline{c_t(\mathcal{A})} + \overline{d_t(\mathcal{A})},$$

where,

$$c_t(\mathcal{A}) := 1 - \sum_{\alpha_0,\ldots,\alpha_t \in \mathcal{A}} p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})\tilde{v}(\alpha_{t-1},\ldots,\alpha_0|\alpha_t), \tag{4.2}$$

and

$$d_t(\mathcal{A}) := -\sum_{\alpha_t \in \mathcal{A}} p(\alpha_t) \ln v(\alpha_t). \tag{4.3}$$

Proof. We start by defining a coarse-grained density of states in the phase space conditional to the path history

$$\varrho_t^{cg}(x|\alpha_0,\ldots,\alpha_{t-1}) := \sum_{\alpha_t} \frac{p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})}{v(\alpha_t)} \mathbb{I}_{\alpha_t}(x), \tag{4.4}$$

where the indicator function $\mathbb{I}_{\alpha_t} : \Gamma \to \{0,1\}$ is defined as

$$\mathbb{I}_{\alpha_t}(x) := \begin{cases} 1, & x \in \alpha_t \\ 0, & x \notin \alpha_t \end{cases}, \tag{4.5}$$

and the phase space volume is

$$v[\alpha_t] := \int_{\alpha_t} dx. \tag{4.6}$$

The coarse-grained Shannon entropy, i.e., the entropy of ϱ_t^{cg}, then becomes

$$H[\varrho_t^{cg}] = -\sum_{\alpha_t} p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1}) \ln \frac{p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})}{v(\alpha_t)}. \tag{4.7}$$

The Shannon entropy of $\varrho_t^{cg}(x|\alpha_0,\ldots,\alpha_{t-1})$ averaged over all possible paths $(\alpha_0,\ldots,\alpha_{t-1})$ can be compared to the Shannon entropy of ϱ_t,

$$\begin{aligned}
\left\langle H[\varrho_t^{cg}]\right\rangle - H\left[\varrho_t\right] &= -\sum_{\alpha_0,\ldots\,\alpha_t} p(\alpha_0,\ldots,\alpha_t)\ln\frac{p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})}{v(\alpha_t)} \\
&+ \int dx\varrho_t(x)\ln\varrho_t(x) \\
&= \sum_{\alpha_0,\ldots\,\alpha_t}\int_{\alpha_t\cap\phi(\alpha_{t-1})\cap\cdots\cap\phi^t(\alpha_0)} dx\varrho_t(x)\left[\ln\varrho_t(x) - \ln\frac{p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})}{v(\alpha_t)}\right] \\
&\geq \sum_{\alpha_0,\ldots\,\alpha_t}\int_{\alpha_t\cap\phi(\alpha_{t-1})\cap\cdots\cap\phi^t(\alpha_0)} dx\left[\varrho_t(x) - \frac{p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})}{v(\alpha_t)}\right] \\
&= 1 - \sum_{\alpha_0,\ldots\,\alpha_t} p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})\frac{v[\alpha_t\cap\phi(\alpha_{t-1})\cap\cdots\cap\phi^t(\alpha_0)]}{v[\alpha_t]},
\end{aligned}$$

where the inequality follows from the relation $a(\ln a - \ln b)\geq a-b$, $a,b\in\mathbb{R}_+^*$ (see page 6 in [16]).

Taking an average in time:

$$\begin{aligned}
\overline{H[\varrho_t^{cg}]} = -\lim_{k\to\infty}\frac{1}{k}\sum_{t=1}^{k}\sum_{\alpha_0,\ldots\,\alpha_{t-1}} p(\alpha_0,\ldots,\alpha_{t-1})\sum_{\alpha_t} p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})\times \\
\times\ln p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1}) \geq H[\varrho_0] + \overline{c_t(\mathcal{A})} + \overline{d_t(\mathcal{A})}
\end{aligned} \tag{4.8}$$

where

$$c_t(\mathcal{A}) := 1 - \sum_{\alpha_0,\ldots,\alpha_t} p(\alpha_t|\alpha_0,\ldots,\alpha_{t-1})\frac{v[\alpha_t\cap\phi(\alpha_{t-1})\cap\cdots\cap\phi^t(\alpha_0)]}{v[\alpha_t]}$$

and

$$d_t(\mathcal{A}) := -\sum_{\alpha_t} p_t[\alpha_t]\ln v[\alpha_t]. \tag{4.9}$$

The over-bars means time average, i.e., $\overline{f_t} = \lim_{k\to\infty}\frac{1}{k}\sum_{t=1}^{k} f_t$. The quantity c_t can be written considering a reversed-time volume measure. Denoting the time-reversed evolution function as $\psi := \phi^{-1}$, and as ψ preserves volume measure, it follows that

$$\begin{aligned}
v[\alpha_t\cap\ldots\phi^t(\alpha_0)] &= v[\psi^t(\alpha_t)\cap\ldots\alpha_0] \\
&= \tilde{v}(\alpha_t,\ldots,\alpha_0).
\end{aligned}$$

Therefore the quantity $c_t(\mathcal{A})$ and $d_t(\mathcal{A})$ can be written as equations (4.2) and (4.3), respectively.

Taking the partition that maximizes the left-hand side of inequality (4.8) we obtain the Kolmogorov-Sinai entropy of ϕ [8], then it follows

$$h(\phi) \geq \overline{H[\varrho_t]} + \overline{c_t(\mathcal{A})} + \overline{d_t(\mathcal{A})}. \tag{4.10}$$

□

Theorem 7. *The time average dissipation is lower bounded in a Hamiltonian process* $(\Gamma, \Sigma, p_t, \phi)$*:*

$$\beta\overline{\langle W_{diss}\rangle_k} \geq \beta[\overline{\langle E\rangle_{\varrho_k} - F_k}] + \overline{c_t(\mathcal{A})} + \overline{d_t(\mathcal{A})} - h(\phi).$$

where $\mathcal{A}$ *is a arbitrary finite partition of* Γ *and* h *is the entropy of* ϕ*.*

Proof. From lemma 6 it follows that

$$\begin{aligned}
\langle W_{diss}\rangle_t &= -k_B T H[\varrho_t] - k_B T \int dx \varrho_t \ln \varrho_t^{eq} \\
&= -k_B T H[\varrho_t] - k_B T \int dx \varrho_t \ln \left[\frac{e^{-\beta E}}{Z}\right] \\
&= -k_B T H[\varrho_t] + \int dx \varrho_t E + k_B T \ln Z \\
&= -k_B T H[\varrho_t] + \langle E\rangle_{\varrho_t} - F(y_t).
\end{aligned} \tag{4.11}$$

The desired result follows from taking time average of equation (4.11) and using lemma 4. □

The hidden information due to coarse-graining is defined as the difference between entropy after coarse-graining and entropy before coarse-graining, see for example equation (1) in [18]. The information hidden in $\mathcal{A}$ is lower bounded [16, 18]

$$H[p_t(\alpha)] - H[\varrho_t] \geq d_t(\mathcal{A}). \tag{4.12}$$

Equation (4.12) can be demonstrated defining the coarse-grained density $\sigma_t^{cq}(x) := \frac{p_t(\alpha)}{v(\alpha)}\mathbb{I}_\alpha(x)$. It follows that,

$$
\begin{aligned}
H[\sigma_t^{cq}] - H[\varrho_t] &= -\sum_\alpha p_t(\alpha) \ln \frac{p_t(\alpha)}{v(\alpha)} + \int \varrho_t(x) \ln \varrho_t(x) dx \\
&= \sum_\alpha \int_\alpha \varrho_t(x) \left[\ln \varrho_t - \ln \frac{p_t(\alpha)}{v(\alpha)} \right] dx \\
&\geq \sum_\alpha \int_\alpha \left[\ln \varrho_t - \ln \frac{p_t(\alpha)}{v(\alpha)} \right] dx \\
&= 0. \qquad (4.13)
\end{aligned}
$$

As $H[\sigma_t^{cq}] = H[p_t(\alpha)] - d_t(\mathcal{A})$ the result is demonstrated. Following the same procedure lemma 4 states that the quantity $I_h := \overline{c_t(\mathcal{A})} + \overline{d_t(\mathcal{A})}$ is interpreted as the average information hidden in the coarse-graining.

Chapter 5

Conclusions and outlook

In this dissertation, we have studied some relations between classical information theory and statistical mechanics of classical Hamiltonian systems. Our result is a lower bound for the dissipated energy in a classical Hamiltonian process, connecting it to the Kolmogorov-Sinai entropy associated with its dynamics.

Results concerning classical physics may work as guides to quantum mechanics in some very specific situations. For example, the fluctuation dissipation theorem in classical physics [**?**] was extended to the quantum case in [**?**], defining entropy production by replacing probability distributions to density operators and relative entropy to quantum relative entropy (see equation (4) in [**?**]). The connection between energetic inefficiency and model inefficiency first presented in [9] was extended to a quantum scenario in [**?**]. The concept of Kolmogorov-Sinai entropy was extended to the case of stochastic quantum system in [**?**]. It raises the question whether entropy production and the "quantum" Kolmogorov-Sinai entropy would be connected in some specific quantum scenario, for instance, the case of stochastic quantum scenario. Our result could work as a guide to such task.

Appendix A

Probability spaces

In a random experiment the possible outcomes constitute the sample space Γ. For instance, in the experiment of tossing a coin the sample space may be written as $\Gamma = \{0, 1\}$. The sample space may be defined by convenience, the only restriction is that any outcome must be represented by an element of Γ [19]. An important concept of random experiments is probability, and there are two ways to understand it [4].

The first is called objective point of view, where the probability is a property of an event. The probability of an event $A \in \Gamma$ must be empirically calculated by

$$p(A) := \lim_{N\to\infty} \frac{n(A)}{N}, \tag{A.1}$$

if it exists, where $n(A)$ is the number of occurrence of the event A in N measurements.

The second is called subjective point of view, and $p(A)$ is understood as an amount of ignorance of the observer concerning the event A. In this approach probability is a property of the observer. This point of view needs a consistent mathematical formalism since there is not a rule of calculation as in A.1. This mathematical formalism of probability theory is due to A. N. Kolmogorov, and it is defined as follows [19].

Definition 21. *A family Σ of subsets of Γ is called σ-algebra if it has the properties*

$$\begin{aligned}
&(\sigma 1) \quad \Gamma \in \Sigma, \\
&(\sigma 2) \quad A \in \Sigma \Rightarrow A^c \in \Sigma, \\
&(\sigma 3) \quad \forall n \in \mathbb{N}, A_n \in \Sigma \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \Sigma,
\end{aligned}$$

where the complement of A is $A^c := \{\gamma | \gamma \in \Gamma, \gamma \notin A\}$.

Definition 22. *A probability on a measurable space* (Γ, Σ) *is normalized, nonnegative and* σ*-additive map defined on* Σ*, i.e,* $p : \Sigma \to \mathbb{R}_+$ *such that*

$$(p1) \quad p(\Gamma) = 1,$$

$$(p2) \quad \forall n \in \mathbb{N}, A_n \in \Sigma, \forall i \neq j, A_i \cup A_j = \emptyset \Rightarrow p\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} p\left(A_n\right).$$

The triple (Γ, Σ, p) *is called probability space.*

Definition 23. *Let* (Γ, Σ, p) *be a probability space. The joint probability is defined as*

$$\begin{aligned} p' :& \Sigma \times \Sigma \to \mathbb{R}_+ \\ & p'(A, B) := p(A, B) = p(A \cap B), \end{aligned}$$

and the conditional probability is defined as

$$\begin{aligned} p'' :& \Sigma \times \Sigma \to \mathbb{R}_+ \\ & p''(A, B) := p(A|B) = \begin{cases} \frac{p(A\cap B)}{p(B)}, & p(B) > 0 \\ p(A), & p(B) = 0 \end{cases}. \end{aligned}$$

Definition 24. *A random variable* $X : \Gamma \to \mathbb{R}$ *is a map measurable to* Σ*, i.e.,* $[X = x] \in \Sigma$*,* $\forall x \in \mathbb{R}$*, where* $[X = x] := \{\gamma \in \Gamma | X(\gamma) = x\}$*. The probability of the random variable* X *is the map*

$$\begin{aligned} p_X :& \mathbb{R} \to \mathbb{R}_+ \\ & p_X(x) := p(X = x) = p(x) = p[X = x]. \end{aligned}$$

A very useful relation regarding probability spaces and convex functions is the Jensen theorem. It is used in chapter 2 to prove some relations in information theory.

Theorem 8. *Let* (Γ, Σ, p) *be a probability space and* X *a random variable of it. If* $f : \mathbb{R} \to \mathbb{R}$ *is a convex function, i.e.,* $f[(1-t)x + tx] \leq (1-t)f(x) + tf(y)$*,* $\forall t \in [0, 1]$*,* $\forall x \in \mathbb{R}$*, then*

$$\int_{\Gamma} dp f(X) \geq f\left(\int_{\Gamma} dp X\right). \tag{A.2}$$

The proof of inequality (A.2) can be found in [19].

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley, 2012.

[3] M. M. Wilde, *Quantum information theory*. Cambridge University Press, 2013.

[4] E. T. Jaynes, "Information theory and statistial mechanics," *Physical Review*, vol. 106, 1957.

[5] R. Kawai, J. Parrondo, and C. V. den Broeck, "Dissipation: the phase-space perspective," *Physical Review Letters*, vol. 98, 2007.

[6] V. I. Arnold and A. Avez, *Ergodic problems of classical mechanics*. Benjamin, 1968.

[7] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic theory*. Springer-Verlag, 1982.

[8] R. Frigg, "In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour?-bridging the gap between dynamical systems theory and communication theory," *The British Journal for the Philosophy of Science*, vol. 19, 2004.

[9] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, "Themodynamics of prediction," *Physical Review Letters*, vol. 109, 2012.

[10] A. I. Khinchin, *Mathematical foundations of information theory*. Dover Publications, 1957.

[11] E. T. Jaynes, *Probability theory: the logic of science*. Cambridge University Press, 2003.

[12] G. E. Crooks, "Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems," *Journal of statistical physics*, vol. 90, 1998.

[13] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Physical Review Letters*, vol. 78, 1997.

[14] N. A. Lemos, *Mecânica analítica*. Editora Livraria da Física, 2007.

[15] R. C. Tolman, *The principles of statistical mechanics*. Clarendon Press, 1938.

[16] F. Benatti, *Deterministic chaos in infinite quantum systems*. Springer, 1993.

[17] Y. Sinai, "Kolmogorov-Sinai entropy," *Scholarpedia*, vol. 4(3), 2009.

[18] A. Alonso-Serrano and M. Visser, "Coarse graining Shannon and von Neumann entropies," *Entropy*, vol. 27, 2017.

[19] B. R. James, *Probabilidade: um curso em nível intermediário*. IMPA, 2015.