Contents

Preface to the English Translation

This translation is a slighltly revised version of the original French version. The main changes occur in Chapters 5 and 6 and consist of clarification of some proofs and a new presentation of the basics in iteration of polynomials.

I have made the choice not to add any new material to preserve the main objective of this book, that is to be an introduction to the subject for the beginner, allowing her or him a quick access to the field.

Finally I wish to thank Greg Anderson who has done a fantastic job translating the text. Actually I have to thank him twice: first for the accuracy and the style of the translation and secondly for having been so patient with my recurrent changes in the text. I also wish to thank Adrien Douady for helpful discussions on the text.

Introduction

Then sure to the atoms is given
No rest whatever throughout the great realms of infinity; rather,
Driven unceasingly on with diversified motion, the more part
After impinging rebound, leaving wide open spaces between them:
Some, scarce parted at all, by collision are set in vibration.
All that rebound from the impact to trifling distances only,
By their own intricate shapes being interlocked and entangled
In more compact aggregations, compose more enduring materials,
Substances of stone, hard metal of iron, and other things like them;
No great number all told, odd passengers through the great space-
void.
All the others spring far apart, rebounding from impact,
Leaving wide space between: these the gaseous atmosphere furnish,
And the bright sunlight withal.

Lucretius, *On the Nature of Things*, Book II,
translated by W. Hannaford Brown.

"Mathematics is the language of the physical sciences, and the symbols of this language are triangles, circles, and other geometric figures." Since the seventeenth century, Galileo's point of view has been the core of the dominant philosophical approach to the relation of mathematics with physics. Mathematical formalism is undeniably a powerful and universal tool for stating the laws of nature. In this vein, geometry was developed to understand optics and the movements of the stars, whereas infinitesimal calculus arose from the formalization of mechanics. Following mathematical principles, these areas have since grown considerably, according to a logic in which physical preoccupations tend to fade away. The pair of mathematics-physics is historically inseparable, with mathematics serving to model physical reality with the intent to rationally understand and clearly expose its laws.

Of course this perspective is overly schematic, and the link between mathematics and physics is more complex. The Greeks of Plato's Academy studied geometry *for its own sake*, purely as an intellectual exercise, and with no concern for applicability at all. As noted by Dieudonné, Appolonius was familiar with conics long before they were used to describe planetary trajectories, and more and more physicists are using mathematical results that are allegedly "abstract" or that have been developed, at least, purely mathematically.

Even with this caveat, this point of view remains "oriented," as it implies an important *asymmetry* in the communication between mathematics and physics. Imagine a planet inhabited by sentient beings with mathematics very different from our own. The perspective outlined above leads to the conclusion that their physics is necessarily very different as well.

Let us now clarify a possible misunderstanding. The existence of this asymmetry does not rely on any value judgement. If it is clear that physics needs mathematics, the converse is no less true. Without physics, mathematics is no more than an empty shell, a production of meaningless tautologies. A telling example is that of potential theory. The work to understand the laws of electromagnetism led over two centuries to the manufacture of the most powerful tools of modern analysis (measure theory, harmonic analysis, distributions, etc.). The distinction originates in the fact that physicists cannot do without a mathematical model, whereas mathematicians have forgotten the physical origin of these concepts. Modern measure theory is a completely autonomous theory whose physical roots play no more than a historical role.

One of the aims of the present work is to develop an approach due to Ruelle [**Rue88**] showing the defects of this generally accepted point of view. The relation of mathematics and physics is more symmetric than one might think. We will show that certain purely mathematical results can be obtained as *corollaries* of physical results (laws), in particular of thermodynamics. Here the mathematical problem is not motivated by physics, and it resolution uses the full force of a formalism developed over a century to describe a natural phenomenon, namely the equilibrium of a gas. It is this example which prompts Ruelle [**Rue88**] to say that our mathematics is *natural*. In other words, we want to provide an example where mathematicians have used a physical model to resolve a mathematical problem, thus inverting the commonly assumed relation between the two sciences.

Thermodynamics now forms a part of *statistical mechanics*, a science created in the previous century by (among others) Clausius, Boltzmann, Carnot, Kelvin, Maxwell, and Gibbs. Its epistemological birth is an acceptance of failure: the laws of classical mechanics, applicable for studying a fluid at the microscopic level, are completely useless at the macroscopic level due to the enormous number of interacting particles. So the only measurable quantities are averages and the movement of the particles must be thought of as random, and they must be examined statistically.

The origin of this science begins like a textbook example of the perspective outlined above and that we precisely want to correct. This (physical) science gave birth to the (mathematical) theory of dynamical systems. The first successes came with Poincaré, then Birkhoff and his famous theorem identifying averages over space with averages over time. Finally Kolmogorov and Sinai took over the idea of entropy (invented by Clausius and Boltzmann) for use as a fundamental invariant in ergodic theory. This remains in the classical pattern: a new physical science invents novel concepts (like entropy) which are "borrowed" (stolen?) by mathematicians who, formalizing them, produce mathematical tools whose physical origins have been completely expunged.

The mathematical story we will use to illustrate Ruelle's ideas originates with a remark of Poincaré in his *Mémoire sur les groupes Kleinéens* [**Poi83**]. Let G be a Fuchsian group, which is to say a group of fractional linear transformations fixing the unit disk D. Assume that G is given by a set of generators which identify in pairs the edges of a polygon in D (in the hyperbolic geometry) in such a way that the images of this polygon under the group form a tiling of D. This group acts symmetrically on the complement of D, and the unit circle ∂D is invariant under the action of G on the Riemann sphere. Imagine now that we slightly perturb the given set of generators, preserving the relations among them. This obtains a new group of fractional linear transformations with the property that the associated polygon and its images still form a tiling of a Jordan domain Δ. The curve $\partial\Delta$ is invariant under the action of this perturbed Kleinian group (and this limit set of the group is analogous to the Julia sets in iteration.) Poincaré remarks that if $\partial\Delta$ is not a circle, then it must be a highly irregular curve: a remarkable intuition which presages Mostow's rigidity theorem and much other significant research!

This intuition was rigorously confirmed several years later by Fricke and Klein. These authors demonstrate that the curve in question, if it is not a circle, is locally non-rectifiable (in particular the 1-dimensional Hausdorff measure is not σ-finite). We note in passing (see [**Pom75**]) that historically this was the first example of such a curve, with Van Koch's curve appearing 10 years later. A better understanding of this phenomenon would have to wait for Mostow [**Mos68**]. Finally Bowen, using Mostow's work and the thermodynamical formalism, shows that the curve, if not a circle, has Hausdorff dimension strictly greater than 1. In fact he exhibits a formula for this dimension, which we now proceed to explain.

First of all, Bowen generalizes the problem to the general framework of *conformal repellers*. Seen from this angle, the action of the group on $\partial\Delta$ can be modeled by the action of a shift operator on the set $A^{\mathbb{N}}$ of infinite sequences of letters from a finite alphabet A. (An example is multiplication by 10 mod 1 on the interval $[0, 1]$ "coded" by decimal expansion.) This set can then be interpreted as the configuration space of a one dimensional lattice gas. The Hausdorff measure appears as a Gibbs state of the system associated to an interaction ϕ depending on the geometry and on a temperature T. This temperature is such that the associated equilibrium state μ minimizes the free energy

$$U - TS.$$

Here U is the potential energy associated with ϕ, S is the entropy of μ, the temperature T being such that the free energy is 0. The dimension is then $(kT)^{-1}$ where k is Boltzmann's constant.

This result is in all ways remarkable: deep thermodynamical methods work perfectly on this problem, even though it can be given no physical interpretation. The mathematical problem that has been modeled here by physics does not itself have any "natural" origin. This situation is exactly symmetric to the one with which we are accustomed, where physicists, to the great surprise of mathematicians, use abstract mathematical results to model experimental situations. Here it is mathematicians who superimpose the thermodynamical model over their problem. What has happened is not just the "borrowing" of ideas or intuition, but the direct and unabashed application of a physical model.

This survey is an attempt to explain this theory both mathematically and physically.

As far as physics is concerned, we will try to explain the thermodynamical ideas necessary to understand the proof clearly enough for a non-physicist to follow. We will include historical remarks whenever possible in this introduction to indicate the evolution of the ideas. For this reason we begin with a chapter exploring the physical origins of the notion of ergodicity before rigorously laying down the basics of ergodic theory and the proof of Birkhoff's theorem.

In the following chapter we develop the various formulations of entropy in (almost) the historical order of appearance. We delay discussion of information theory until the most pedagogically sound framework is in place for an approach to this sophisticated idea. The chapter ends with a "toy model" of thermodynamics. With this one can easily define thermodynamic functions and understand the fundamental ties that unite them. This lies at the heart of what follows.

The next chapters rigorously develop the theory of thermodynamics of lattices, the Perron-Frobenius-Ruelle theorem (this is the key to the vault for all of this work), and its "thermodynamical" consequences. We lay particular stress on the ideas of Gibbs state and equilibrium state, two essentially physical ideas which yield mathematical contributions.

Taken together, what we have mentioned is commonly called "the thermodynamical formalism," and the rest of this work is dedicated to applications of it. This descriptive term seems quite convenient, in so far as the formalism is present in these applications. We begin with a general study of conformal repellers, a theory which encompasses that of Bowen mentioned earlier in connection with Poincaré's problem. The thermodynamical formalism allows us to relate the Hausdorff dimension of the "limit set" to the spectral radius of a Ruelle operator. This fact yields some information about this dimension, and also allows us to examine its variation with parameters on which the conformal repeller depends. For the case of Poincaré's problem, this amounts to consideration of the space of all possible deformations of a Fuchsian group; this is the Teichmüller space of the group, and it comes equipped with a natural finite-dimensional complex structure. The question to be asked, then, concerns the regularity of the functions on Teichmüller space which arise as thermodynamic functions. This has all been examined, but the number of parameters in this case and the special techniques of Teichmüller theory prevent us from pursuing these matters in this survey.

We therefore abandon this problem of Poincaré, instead focusing our attention on the case of iteration of quadratic polynomials ($z \mapsto z^2 + c$). For values of c lying inside the main cardioid of the Mandelbrot set, the associated Julia set is a Jordan curve on which the polynomial acts as a conformal repeller, so all the thermodynamical theory is available. The complexity here is manageable due to the fact that the associated "Teichmüller space" (i.e. the set of values of c) has dimension 1.

As a concrete illustration of the power of the thermodynamical formalism, we will use it to rigorously establish the three following results.

1. The Hausdorff dimension $d(c)$ of the Julia set for $z \mapsto z^2 + c$ depends real-analytically on c in the main cardioid.
2. In the main cardioid, $d(c) > 1$ except when $c = 0$.
3. To second order near 0, we have

$$d(c) = 1 + \frac{|c|^2}{4 \log 2} + o(|c|^2).$$

We mention that the first and third of these results are due to Ruelle [**Rue78**], while the second should be credited to Bowen [**Bow78**]. We will give a proof of the second order behavior of $d(c)$ different from Ruelle's original proof, which is by his own admission heuristic. Ours is rigorous, and is perhaps one of the only novel features of this monograph, another being, we hope, to fill a gap in the literature and to do a service to the analysts who would like to study problems on the Hausdorff dimension of limit sets of groups or semi-groups, and perhaps even to physicists interested in the analytic aspects of the subject.

In the final chapter we address the problem of phase transitions, a topic all the more fascinating because of the paradox that comes with it. This is an imminently concrete and tangible problem, but the theory which is supposed to explain it is difficult and still mostly incomplete. After a preliminary review of well known physical results, show first of all that the thermodynamical formalism generalizes to the case of polynomials whose Julia set contains no critical point, a theory due to Denker and Urbanski [**DU91**]. We then show that in the case of existence of a rationally indifferent fixed point, the "temperature" corresponding to the Hausdorff dimension undergoes a phase transition. These results are then used to show that the function $d|_{\mathbb{R}}$ is continuous to the left of $\frac{1}{4}$ (see [**BZ96**]).

We conclude this introduction with the confession that the work which follows is not completely self-contained. Along the way we will require the Koebe distortion theorem, a proof of which can be found in any good book on complex variables (for example [**Pom75**]), and the work of Mañé, Sad, and Sullivan [**MSS83**] which has been followed by the recent theorem of Slodkowski, which deserves another book of this same size and for which we refer to [**Sło91**] and [**AM92**]. We will also make use of the Schauder-Tychonov fixed point theorem, a proof of which can be found in [**Rud73**], as well as the Kato-Rellich theorem on perturbation of the spectrum, as in [**RS75**]. Finally, we have deliberately (and this is surely a mistake) avoided any quantum aspects of the subject.

A word concerning the style: this is a work of synthesis. My aim in writing it was to understand and make understood certain mechanisms and trains of thought. To do this, I had to read many works, all written by people more knowledgeable than myself. I have not hesitated on occasion to take certain developments from them almost word for word, and this is in the interest of clarity: since it is so clearly explained already, why change it? A work of synthesis is in many ways an act of plunder. Among all the works I have consulted, I would mention here the ones that I have "looted" from the most.

- Douady and Hubbard [**DH85**]: *Étude dynamique des polynômes complexes*
- Jancovici [**Jan73**]: *Thermodynamique et physique statistique*
- Khinchin [**Khi49**]: *Mathematical Foundations of Statistical Physics*
- Landau and Lifschitz [**LL59**]: *Statistical Mechanics*
- Milnor [**Mil**]: *Dynamics in one complex variable: Introductory Lectures*
- Parry and Pollicot [**PP90**]: *Zeta functions and the periodic structure of hyperbolic dynamics*
- Ruelle [**Rue78**]: *Thermodynamic formalism*
- Walters [**Wal81**]: *Introduction to ergodic theory*

CHAPTER 3

Entropy in Ergodic Theory

3.1. Physical Framework: Thermodynamics of Lattices

In this chapter we rigorously work out the mathematical ideas which will accompany us through the rest of our work. These ideas originated in thermodynamics. As we have done previously, we will begin with a description of the original physical infrastructure. The presentation here is adapted from Ruelle [**Rue78**].

This setup differs from the ones we have seen before in that it is no longer time which serves as the group of transformations, but space. We consider the space $L = \mathbb{Z}^d$ as a model for our physical "body," which can be thought of as a crystalline structure rather than a fluid. Each element of L can be in one of the states from the finite set A. The simplest example is the case where $A = \{0, 1\}$, where 0 or 1 is associated to a point of the lattice depending on whether it is "occupied" or not. The set $\Omega = A^L$ is the set of configurations of the system. If $S \subset L$, we write $\Omega_S = A^S$.

An *interaction* is a function Φ defined on

$$\bigcup_{\Lambda \subset L \text{ finite}} \Omega_\Lambda$$

such that the contribution of $x \in L$ to the interaction, defined by

$$|\Phi|_x = \sum_{X \ni x} \frac{1}{|X|} \sup_{\xi \in \Omega_X} |\Phi(\xi)|, \qquad |X| = \operatorname{Card} X$$

is finite for every x.

We will always make the hypothesis that Φ is translation invariant. In this case, $|\Phi|_x$ is in fact independent of x, so it can serve as a norm on the set of interactions. This hypothesis gives a privileged position to the action of L as group on itself, and it is by virtue of this action that ergodic theory comes into play.

An interaction is called *finite* if there exists a finite set $\Delta \subset L$ such that for every $\xi \in \Omega$,

$$\Phi(\xi|_X) = 0$$

unless $X - x$ is contained in Δ for all $x \in X$. (The notation $\xi|_X$ indicates the restriction of ξ to X.) For example, if $d = 1$ and $\Delta = \{-1, 0, 1\}$, then $\Phi(\xi|_X)$ must be zero unless X consists of two consecutive integers. The subspace of finite interactions is dense in the Banach space of all interactions.

Having fixed an interaction, we then define an *energy function* as follows. For each finite subset $\Lambda \subset L$, we set

$$U_\Lambda^\Phi(\xi) = \sum_{X \subset \Lambda} \Phi(\xi|_X).$$

Imitating the discussion from Section 2.3, we define the partition function for the finite set Λ by

$$Z_\Lambda^\Phi = \sum_{\xi \in \Omega_\Lambda} \exp(-U_\Lambda^\Phi(\xi)),$$

and the Gibbs ensemble as the probability measure defined on Ω_Λ by

$$P_\Lambda(\{\xi\}) = \frac{1}{Z_\Lambda} \exp(-U_\Lambda^\Phi(\xi)).$$

The art of statistical mechanics consists in a "passage to the limit" as "Λ tends to L" to obtain a thermodynamic limit called a *Gibbs state*. In general, the term state stands for a measure on Ω.

We will not develop this aspect of things from the physical standpoint, but we will return to the mathematical analog in the following section. From that perspective, we will see that this state minimizes a certain quantity, called pressure, which is analogous to the free energy defined in Section 2.3. This quantity will be defined in a novel (for us) manner in which the group action (in fact the shift in this case) plays a starring role.

The remainder of this section is taken up with a "physical" motivation for the definitions. We saw in Chapter 2 that, at a fixed temperature, the Gibbs ensemble maximizes the quantity $\log Z$, which is simply $-\beta/N$ times the free energy. We wish to give some sense to the thermodynamic limit of the quantities $\log Z_\Lambda$. In Chapter 4 we will see that the appropriately defined Gibbs state is an equilibrium state in the sense that it maximizes this quantity.

We begin by defining

$$A_\Phi(\xi) = -\sum_{X \ni 0} \frac{1}{|X|} \Phi(\xi|_X),$$

which is well-defined by the definition of interaction and satisfies $\|A\|_\infty \leq |\Phi|$. We now define the *modified partition function* by

$$Z_\Lambda^*(A_\Phi) = \sum_{\xi \in \Omega_\Lambda} \exp\left(\sum_{x \in \Lambda} A_\Phi(\tau^x \xi^*)\right).$$

In this formula, for each $\xi \in \Omega_\Lambda$, ξ^* is a point in the full configuration space Ω for which $\xi^*|_\Lambda = \xi$, and $\tau^x(\xi)(y) = \xi(y+x)$. We also set

$$\Lambda(n) = \{0, \ldots, n\}^d.$$

In the next section we will show (for a special case) that as n tends to $+\infty$,

$$\frac{1}{|\Lambda(n)|} \log Z_{\Lambda(n)}^*(A_\Phi)$$

converges to a quantity called the *pressure of* A_Φ, denoted by $P(A_\Phi)$. By way of motivation we will show here that

$$\frac{1}{|\Lambda(n)|} \log Z_{\Lambda(n)}^\Phi$$

has the same limit. To see this, consider first a finite interaction Φ, and write

$$\begin{aligned} U_\Lambda^\Phi(\xi^*|_\Lambda) + \sum_{x\in\Lambda} A_\Phi(\tau^x\xi^*) &= \sum_{x\in\Lambda}\left(\sum_{x\in X\subset\Lambda}\frac{\Phi(\xi^*|_X)}{|X|} - \sum_{0\in X}\frac{\Phi(\tau^x\xi^*|_X)}{|X|}\right) \\ &= \sum_{x\in\Lambda}\left(\sum_{x\in X,\, X\cap^c\Lambda\neq\emptyset}\frac{\Phi(\xi^*|_X)}{|X|}\right), \end{aligned}$$

where we have used the invariance of Φ under translations. This quantity is bounded in absolute value by $N(\Lambda)|\Phi|$, where $N(\Lambda)$ is the number of elements y of Λ such that $y+\Delta$ is not contained in Λ. The desired conclusion follows easily by remarking that $N(\Lambda(n))/|\Lambda(n)|$ tends to 0 as n tends to ∞. The general case (Φ not necessarily finite) can be handled with an equicontinuity argument, but we do not develop this here.

One might ask why the limit quantity $P(A_\phi)$ is called pressure. The answer is already given in the calculation of Section 2.3. We saw there that the free energy is given by $-N\log Z/\beta$. In the computation above, the temperature does not come into play, and we have set $\beta = 1$. The difference between the situation here and that of Section 2.3 is that here the system has infinite volume, and $P(A_\Phi)$ thus has the dimensions of energy per unit volume, also known as pressure.

We turn now to a characteristically mathematical treatment of this topic in the extremely special case of $d = 1$. Moreover, $\mathbb{Z}$ will be replaced by $\mathbb{N}$; this will not have any profound effect on the ideas, although certain things will be easier in this case.

3.2. Entropy of an Invariant Measure

In what follows, we will systematically place our work in the context of symbolic dynamics. This will simplify proofs, and it will be sufficient for the applications we have already seen. For a more general account, the reader may profitably consult [**Wal81**].

3.2.1. Entropy of the Shift. Let $A = \{1, \dots, m\}$ be a finite alphabet, and consider the dynamical system $X = A^{\mathbb{N}}$ with T the shift map on X. Everything we do here remains valid in the case of a sub-shift of finite type, to be defined in Section 3.4 The measure space $(X, \mathcal{B})$ comes equipped with a natural filtration $(\mathcal{B}_n)$ where $\mathcal{B}_n$ is the σ-algebra generated by the set $\mathcal{A}_n$ of cylinders $x_1 \cdots x_n$.

Now let μ be a T-invariant probability measure on X. For each n, we can define the entropy of the random experiment which consists of choosing a cylinder in $\mathcal{A}_n$ with probability according to μ. This entropy is

$$H_\mu(\mathcal{A}_n) = -\sum_{A\in\mathcal{A}_n}\mu(A)\log(\mu(A)).$$

PROPOSITION 3.1. *For every $n, p \in \mathbb{N}$, we have $H_\mu(\mathcal{A}_{n+p}) \leq H_\mu(\mathcal{A}_n)+H_\mu(\mathcal{A}_p)$.*

PROOF. Set $u_n = H_\mu(\mathcal{A}_n)$. First off, we observe that since μ is in the first place a measure, and in the second place invariant, we have for every cylinder A and every $p \geq 0$,

$$\mu(A) = \sum_{B\in\mathcal{A}_p}\mu(AB) = \sum_{B\in\mathcal{A}_p}\mu(BA).$$

From this we can write

$$u_{n+p} = \sum_{A\in\mathcal{A}_n,\, B\in\mathcal{A}_p} -\mu(AB)\log(\mu(AB)).$$

We also have the equalities

$$\sum_B \frac{\mu(AB)}{\mu(A)} = 1, \text{ and } \sum_B \frac{\mu(AB)}{\mu(A)}\frac{\mu(B)}{\mu(AB)} = \frac{1}{\mu(A)}$$

which, combined with the convexity of the function $-\log$, imply the inequality

$$\log(\mu(A)) \leq -\sum_{B\in\mathcal{A}_p} \frac{\mu(AB)}{\mu(A)} \log\frac{\mu(B)}{\mu(AB)}.$$

This in turn yields

$$\mu(A)\log(\mu(A)) \leq -\sum_B \mu(AB)\log(\mu(B)) + \sum_B \mu(AB)\log(\mu(AB))$$

and the claim follows by summing over $A\in\mathcal{A}_n$ and using invariance by T. □

Proposition 3.2. *Let (u_n) be a sequence of nonnegative reals such that $u_{n+p} \leq u_n + u_p$ for all $n, p \in \mathbb{N}$. The the sequence (u_n/n) is convergent.*

Proof. We will in fact show that the sequence converges to $\ell = \inf\{u_n/n,\ n \geq 1\}$. Choose $\epsilon > 0$. From the definition of infimum, there is an integer n_0 such that $\ell \leq u_{n_0}/n_0 \leq \ell + \epsilon$. For $n \geq n_0$, use the Euclidean algorithm to write $n = pn_0 + r$, so that the condition on the sequence u_n implies $u_n \leq pu_{n_0} + u_r$. So if n (and therefore p) is large enough, we have

$$\frac{u_n}{n} \leq \frac{pu_{n_0} + u_r}{pn_0} \leq \frac{u_{n_0}}{n_0} + \frac{\max(u_0, u_1, \dots, u_{n_0-1})}{pn_0} \leq \ell + 2\epsilon.$$

□

These two results allow us to define the *entropy of the shift with respect to the measure* μ by

$$h_\mu(T) = \lim_{n\to\infty} \frac{H_\mu(\mathcal{A}_n)}{n}.$$

Proposition 3.3. *The function $\mu \mapsto h_\mu(T)$ is linear on the compact convex set $M(X,T)$ of T-invariant probability measures on X.*

Remark 3.4. The preceding discussion already implies that $\mu \mapsto h_\mu(T)$ is upper semi-continuous.

Proof. Consider $\mu', \mu'' \in M(X,T)$, $\alpha \in [0,1]$, and $\mu = \alpha\mu' + (1-\alpha)\mu''$. If $A \subset X$ is a Borel set, then convexity of the function $x\log x$ implies

$$\begin{aligned}
0 &\leq -\mu(A)\log\mu(A) + \alpha\mu'(A)\log\mu'(A) + (1-\alpha)\mu''(A)\log\mu''(A)\\
&= -\mu'(A)\left(\log\mu(A) - \log(\alpha\mu'(A))\right)\\
&\quad -(1-\alpha)\mu''(A)\left(\log\mu(A) - \log\left((1-\alpha)\mu''(A)\right)\right)\\
&\quad -\mu'(A)\alpha\log\alpha - \mu''(A)(1-\alpha)\log(1-\alpha)\\
&\leq -\mu'(A)\alpha\log\alpha - \mu''(A)(1-\alpha)\log(1-\alpha).
\end{aligned}$$

Consequently, for every $n \geq 0$ we have

$$0 \leq H_\mu(\mathcal{A}_n) - \alpha H_{\mu'}(\mathcal{A}_n) - (1-\alpha)H_{\mu''}(\mathcal{A}_n) \leq \frac{2}{e},$$

so the result follows upon division by n and taking the limit. □

3.2.2. Application: The Shannon-McMillan Theorem. As in the previous chapter, we would like to "motivate" this concept of entropy by showing that it appears "naturally" in certain contexts. The best example is the Shannon-McMillan Theorem. Before proving it, we will establish a formula which is useful for the calculation of entropy.

Consider $x \in X$, so $x = x_1 x_2 \cdots x_n \cdots$. For $n \in \mathbb{N}$, define

$$J_n(x) = \begin{cases} \mu(x_2 \cdots x_n)/\mu(x_1 \cdots x_n) & \text{if the denominator is nonzero,} \\ +\infty & \text{otherwise.} \end{cases}$$

Note that the function J_n is bounded below by 1 because of the invariance of μ.

PROPOSITION 3.5. *The sequence (J_n) is an $L^1(\mu)$ martingale with respect to the filtration $(\mathcal{B}_n)$.*

PROOF. Consider $x_1 \cdots x_n \in \mathcal{A}_n$, and assume $\mu(x_1 \cdots x_n) > 0$. On this cylinder, we have

$$E(J_{n+1}|_{\mathcal{B}_n}) = \frac{1}{\mu(x_1 \cdots x_n)} \sum_{x_{n+1}} \mu(x_1 \cdots x_n x_{n+1}) \frac{\mu(x_2 \cdots x_n x_{n+1})}{\mu(x_1 \cdots x_n x_{n+1})} = J_n.$$

Furthermore, $E(J_n) = \sum_{x_1 \cdots x_n \in \mathcal{A}_n} \mu(x_2 \cdots x_n) \leq m$. □

COROLLARY 3.6. *The function $J_\mu(x) = \lim_{n\to\infty} J_n(x)$ exists almost everywhere. It is called the* Jacobian *of T with respect to the measure μ.*

PROPOSITION 3.7. *The function*

$$M(x) = \sup\{\log J_n(x) \text{ for } n \geq 0\}$$

is in $L^1(X, \mu)$.

PROOF. For $\lambda > 0$ we define $E_\lambda = \{M > \lambda\}$. This set can also be described as

$$E_\lambda = \left\{ x \in X \text{ such that } \inf_n \frac{\mu(x_1 \cdots x_n)}{\mu(x_2 \cdots x_n)} < e^{-\lambda} \right\}.$$

Let (C_n) be the maximal cylinders for which $\mu(C_n) < e^{-\lambda}\mu(T(C_n))$. By maximality, the C_n are pairwise disjoint, and we have also $E_\lambda = \bigcup C_n$. Therefore

$$\mu(E_\lambda) = \sum \mu(C_n) < e^{-\lambda} \sum \mu(T(C_n)) \leq m e^{-\lambda},$$

and we conclude that $\int M d\mu = \int_0^\infty \mu(E_\lambda) d\lambda \leq m$. □

We are now in a position to establish the following formula, known as *Rohlin's formula.*

THEOREM 3.8. $h_\mu(T) = \int_X \log J_\mu d\mu$

PROOF. From the three previous results and the dominated convergence theorem we obtain

$$\int_X \log J_\mu d\mu = \lim_{n\to\infty} \int_X \log J_n d\mu.$$

But observe that

$$\int_X \log J_n d\mu = \sum_{A \in \mathcal{A}_n} \mu(A) \log \mu(T(A)) - \sum_{A \in \mathcal{A}_n} \mu(A) \log \mu(A) = u_n - u_{n-1}.$$

We know that this sequence converges to a limit L, so its Césaro sums converge to the same limit. But the n-th order Césaro sum is $u_n/n - u_0/n$, and this converges to $h_\mu(T)$. □

EXAMPLE 3.9. We now use this formula to calculate the entropy of a Markov shift. First of all we obtain the Jacobian of the shift for this case. Recall that a Markov shift is determined by a stochastic transition matrix Π and a (left) eigenvector p for the eigenvalue 1. If $x = x_1 \cdots x_n \cdots$, we have

$$J_n(x) = \frac{\mu(x_2 \cdots x_n)}{\mu(x_1 \cdots x_n)} = \frac{p_{x_2}}{p_{x_1} p_{x_1 x_2}} = J_\mu(x).$$

Thus the Jacobian only depends on the first two variables, and consequently

$$\begin{aligned} h_\mu(T) &= \int_X J_\mu d\mu = \sum_{i,j \in A} p_i p_{ij} (\log p_j - \log p_i - \log p_{ij}) \\ &= \sum_j p_j \log p_j - \sum_i p_i \log p_i - \sum_{i,j} p_i p_{ij} \log p_{ij} = -\sum_{i,j} p_i p_{ij} \log p_{ij}. \end{aligned}$$

We note that in the case of a Bernoulli shift, this formula reduces to

$$H_\mu(\mathcal{A}_1) = -\sum p_i \log p_i.$$

THEOREM 3.10 (Shannon-McMillan). *If (X, T, μ) is ergodic then*

$$-\frac{1}{n} \log(\mu(x_1 \cdots x_n)) \overset{\mu-\text{a.e.}}{\longrightarrow} h_\mu(T).$$

PROOF. We have

$$-\frac{1}{n} \log(\mu(x_1 \cdots x_n)) = \frac{1}{n} \sum_{k=1}^{n} \log(J_\mu(T^k(x))) + \frac{1}{n} \sum_{k=1}^{n} (\log J_k - \log J_\mu)(T^{n-k}(x)).$$

The first term is handled by the ergodic theorem, and its limit is $\int \log(J_\mu) d\mu = h_\mu(T)$. To deal with the second term, for $N > 0$ we set

$$F_N = \sup\{|\log J_k - \log J_\mu| \text{ for } k \geq N\}.$$

We now have

$$\frac{1}{n} \sum_{k=1}^{n} |\log J_k - \log J_\mu| \circ T^{n-k} \leq \frac{1}{n} \sum_{k=1}^{N-1} F_0 \circ T^{n-k} + \frac{1}{n} \sum_{k=N}^{n} F_N \circ T^{n-k}.$$

By the dominated convergence theorem, $\int F_N d\mu$ tends to 0 as N approaches ∞, and let us fix N so that this integral is less than ϵ. Now we let n tend to ∞, so that the first term tends to 0 a.e. according to Birkhoff's ergodic theorem. This same theorem shows that the second term tends to $\int F_N d\mu \leq \epsilon$. □

3.3. Topological Entropy and Pressure

The results in this section are essentially due to Walters [**Wal81**]. Let ϕ be a continuous function on X. For $n \geq 0$, define

$$S_n \phi = \sum_{k=0}^{n-1} \phi \circ T^k.$$

For a cylinder C and a continuous function $\psi \in C(X)$, we also define

$$\psi_C = \sup\{\psi(x) \text{ for } x \in C\}.$$

THEOREM 3.11. *The following limit exists and defines the pressure of ϕ:*

$$P(\phi) = \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{C\in\mathcal{A}_n} e^{(S_n\phi)_C} \right).$$

PROOF. We set $V_n = \sum_{C\in\mathcal{A}_n} e^{(S_n\phi)_C}$, and note that

$$V_{n+p} = \sum_{A,B} e^{(S_{n+p}\phi)_{AB}} \leq \sum_{A,B} e^{(S_n\phi)_A} e^{(S_p\phi)_B} \leq V_n V_p.$$

Now apply Proposition 3.2 with $u_n = \log V_n$. □

The *topological entropy* is defined to be the pressure of $\phi \equiv 0$. In the case of the shift on $A^{\mathbb{N}}$, the topological entropy is $\log m$. For the case of a sub-shift, see Section 3.4. In the general case, we have the following result, called the *variational principle*.

THEOREM 3.12. *We have $P(\phi) = \sup\{h_\mu(T) + \int_X \phi \, d\mu$ for $\mu \in M(X,T)\}$.*

REMARK 3.13. Proposition 3.3 and the accompanying remark show that this supremum is attained by at least one ergodic measure in $M(X,T)$.

PROOF. We show first that $P(\phi)$ is at least as large as the indicated supremum using the following simple lemma whose proof is left to the reader.

LEMMA 3.14. *Let $p_1, \dots, p_n$ be nonnegative reals with $\sum p_i = 1$, and take arbitrary real numbers $a_1, \dots, a_n$. Then*

$$\sum_{i=1}^{n} p_i(a_i - \log p_i) \leq \log \left(\sum_{i=1}^{n} e^{a_i} \right)$$

with equality if and only if $p_i = e^{a_i}/(\sum e^{a_i})$.

Applying the lemma to $p_A = \mu(A)$ and $a_A = (S_n\phi)_A$, we obtain

$$H_\mu(A) + \frac{1}{n} \sum_{\mathcal{A}_n} p_A (S_n\phi)_A \leq \frac{1}{n} \log V_n.$$

Taking ω_k to be $\sup\{|\phi(x) - \phi(x')|$ such that $x_j = x'_j$ for $1 \leq j \leq k\}$, we then have

$$\sum p_A (S_n\phi)_A \geq n \int_X \phi \, d\mu - \sum_{k=0}^{n} \omega_{n-k},$$

and the bound follows from the fact that ω_n approaches 0 as n tends to ∞.

To prove the converse inequality, we use the following trick that was taught to us by Thierry Bousch. The two side of the equality in the claim are Lipshitz with constant 1 on $C(X)$. To prove equality it thus suffices to show that it holds for a dense subset of $C(X)$. This will be precisely done in the next chapter, in Proposition 4.11 and Theorem 4.12. □

REMARK 3.15. Theorem 3.12 is used in Proposition 4.11, so one might suspect there is a logical vicious circle. This is not the case, since only Lemma 3.14 is used there.

3.4. Sub-shifts

There is a vast literature on this subject; we simply cite [**DGS76**] as a notable reference.

In practice, the words of $A^{\mathbb{N}}$ are not all admissible. Consider an example that we will return to in the sequel, that of a free group G on p symbols, $\gamma_1, \ldots, \gamma_p$. This group is the set of words of the form $x_1 \cdots x_k$, where each x_j is one of the γ_j or another symbol denoted γ_j^{-1}. The rule is that if γ_i is followed by γ_i^{-1} (or vice versa) then the word is identified with the one obtained by erasing these two symbols. Of course it is possible that the entire word might be erased, in which case one obtains the empty word, which is the identity element of G.

Having described this group, we define the (abstract) *limit set* of G as the set of elements (words) of

$$\{\gamma_1, \ldots, \gamma_p, \gamma_1^{-1}, \ldots, \gamma_p^{-1}\}^{\mathbb{N}}$$

in which the sequences $\gamma_i \gamma_i^{-1}$ and $\gamma_i^{-1} \gamma_i$ never appear for any i. We remark that this limit set is a closed subset of $A^{\mathbb{N}}$ (where $A = \{\gamma_1, \ldots, \gamma_p, \gamma_1^{-1}, \ldots, \gamma_p^{-1}\}$) which retains the fundamental property of invariance under the shift.

Let A be a finite alphabet and F a finite set of finite words of A. The subset X of $A^{\mathbb{N}}$ of words in which the elements of F never appear is a closed subset of $A^{\mathbb{N}}$ which is shift-invariant. Such a pair (X, T) is called a *sub-shift of finite type*.

By enlarging the alphabet with the addition of a new letter for each word of F (and replacing T by an iterate) we can always assume that the banned words consist of exactly two letters each.

There is an alternative method of defining sub-shifts of finite type, which we now describe. Again, let $A = \{1, \ldots, m\}$ be a finite alphabet, and let $M = (m_{ij})$ be an $m \times m$ matrix with entries either 0 or 1. Furthermore, assume that each row and each column contains at least one 1. Entries of M^n are denoted m_{ij}^n.

A word $x = x_1 \cdots x_n \cdots \in X = A^{\mathbb{N}}$ is said to be M-*admissible* if it satisfies $m_{x_i x_{i+1}} = 1$ for every $i \geq 1$. We denote the collection of M-admissible words by $X(M)$. Evidently $(X(M), T)$ is a sub-shift of finite type.

All of the definitions in this chapter can be extended to the case of sub-shifts of finite type, where it is to be understood that all summations are restricted to M-admissible cylinders. Moreover, we recover the case of the full shift by taking M to be the matrix with every element equal to 1. Nevertheless, certain topological properties of the shift on $A^{\mathbb{N}}$ can be lost for certain matrices M. Let us now examine the relations between algebraic properties of M and topological properties of $(X(M), T)$.

Let X be a compact space and T a continuous map from X to itself. We say that T is *topologically transitive* if for each pair of open sets U, V, there exists an integer $n \geq 0$ such that $T^n(U) \cap V \neq \emptyset$. The reader may easily check that T is topologically transitive if and only if there exists $x \in X$ such that the collection of iterates $T^n(x)$, $n \geq 0$, is dense in X, and that in the case the set of x with this property is also dense in X.

We say that T is *topologically mixing* if for each pair of open sets U, V, there exists an integer $n \geq 0$ such that $k \geq n$ implies $T^k(U) \cap V \neq \emptyset$. Of course this latter property is stronger than the former, and it is satisfied by the shift on $A^{\mathbb{N}}$ for the simple reason that for any open U there is $n \geq 0$ such that $T^n(U) = A^{\mathbb{N}}$.

We arrive at the following theorem, whose proof is left to the reader.

THEOREM 3.16. *The sub-shift $(X(M), T)$ is topologically transitive if and only if for every $i, j \in A$ there exists $n \geq 1$ for which $m_{ij}^n > 0$ (i.e. M is irreducible). It is topologically mixing if and only if there exists $n \geq 1$ such that for every $i, j \in A$ we have $m_{ij}^k > 0$ (i.e. M is aperiodic).*

There exist irreducible matrices which are not aperiodic. The simplest example is given by the cyclic matrix

$$M(e_j) = e_{j+1} \text{ for } j \leq n-1 \text{ and } M(e_n) = e_1.$$

Here we have $m_{ii}^k > 0$ if and only if k is a multiple of n. Actually, this example allows us to construct all irreducible non-aperiodic matrices. Modulo the order of the generators, every such matrix can be obtained by starting with the matrix above, replacing each 1 with a d-dimensional aperiodic matrix, and each 0 with a d-dimensional zero matrix. We refer the reader to [**DGS76**] for more information about these matrices.

In what follows the matrix M is always assumed to be aperiodic, unless explicitly stated to the contrary.

PROPOSITION 3.17. *For $y \in X(M)$, the set $\bigcup_{n \geq 0} T^{-n}(y)$ is dense in $X(M)$.*

The easy proof is left to the reader. This last proposition will allow us to extend the Perron-Frobenius-Ruelle Theorem to the case of aperiodic sub-shifts.

We bring this algebraic discussion to a close with a return to the example of the free group on p symbols. Here the alphabet A has $n = 2p$ elements and the matrix M is $1_n - I_n$ where 1_n is the matrix with all entries equal to 1, and I_n is the identity matrix of size n. Then as long as $p > 1$ we have

$$M^2 = (n-2)1_n + I_n > 0,$$

and the matrix is therefore aperiodic. We remark that if $p = 1$ then M is not even irreducible and the limit set of G reduces to two points.

In contrast to the case of the usual shift, the calculation of topological entropy is not so obvious, but we will do this in the case where M is irreducible. We have the following theorem, which is (for M aperiodic) a special case of the main theorem of the next chapter.

THEOREM 3.18 (Perron-Frobenius). *If M is a positive irreducible matrix, there exists $\lambda > 0$ such that for any eigenvalue η of M we have $|\eta| \leq \lambda$, and also such that λ is a simple eigenvalue for M and M^t associated to positive eigenvectors.*

Let us provisionally accept this theorem and use it to perform a calculation.

THEOREM 3.19. *The topological entropy of the sub-shift associated to the irreducible matrix M is $\log(\lambda(M))$ where $\lambda(M)$ is the largest eigenvalue of M.*

PROOF. The topological entropy is given by

$$h(T) = \lim_{n \to \infty} \frac{\log(N_n)}{n}$$

where N_n is the number of admissible cylinders of order n. Let $N_n(j)$ be the number of these cylinders which end with j, so we have the formulas

$$\sum_{j=1}^{m} N_n(j) = N_n, \quad N_{n+1}(j) = \sum_{i=1}^{m} m_{ij} N_n(i).$$

Using V_n to denote the row vector $(N_n(j))$, these yield $V_{n+1} = V_n M$, implying that $V_n = V_0 M^n$ with $V_0 = (1, \dots, 1)$.

Now take a positive (left) eigenvector V for M with eigenvalue $\lambda = \lambda(M)$. Setting $a = \min(V(i))$ and $b = \max(V(i))$, we then have for each $j \in A$ that

$$\frac{V(j)}{a}\lambda^n = \sum_{i=1}^{m} \frac{V(i)}{a} m_{ij}^n \geq \sum_{i=1}^{m} V_0(i) m_{ij}^n \geq \frac{V(j)}{b}\lambda^n.$$

Since the term in the middle is $N_n(j)$, we see by summing the inequalities over j that $h(T) = \log(\lambda)$. □

In the special case where each row and each column of M have the same number k of nonzero entries, $h(T) = \log k = h_\mu(T)$ where μ is the Markov measure defined by $p = (1/m, \dots, 1/m)$ and $\Pi = (1/k)M$. Thus for our example of the free group on p letters, the topological entropy is equal to $\log(2p-1)$.