Invariant measures for hyperbolic dynamical systems

N. Chernov

September 14, 2006

Contents

1 Markov partitions **3**

2 Gibbs Measures **17**
- 2.1 Gibbs states 17
- 2.2 Gibbs measures 33
- 2.3 Properties of Gibbs measures 47

3 Sinai-Ruelle-Bowen measures **56**

4 Gibbs measures for Anosov and Axiom A flows **67**

5 Volume compression **86**

6 SRB measures for general diffeomorphisms **92**

This survey is devoted to smooth dynamical systems, which are in our context diffeomorphisms and smooth flows on compact manifolds. We call a flow or a diffeomorphism hyperbolic if all of its Lyapunov exponents are different from zero (except the one in the flow direction, which has to be zero). This means that the tangent vectors asymptotically expand or contract exponentially fast in time. For many reasons, it is convenient to assume more than just asymptotic expansion or contraction, namely that the expansion and contraction of tangent vectors happens uniformly in time. Such hyperbolic systems are said to be uniformly hyperbolic.

Historically, uniformly hyperbolic flows and diffeomorphisms were studied as early as in mid-sixties: it was done by D. Anosov [2] and S. Smale [77], who introduced his Axiom A. In the seventies, Anosov and Axiom A diffeomorphisms and flows attracted much attention from different directions: physics, topology, and geometry. This actually started in 1968 when Ya. Sinai constructed Markov partitions [74, 75] that allowed a symbolic representation of the dynamics, which matched the existing lattice models in statistical mechanics. As a result, the theory of Gibbs measures for one-dimensional lattices was carried over to Anosov and Axiom A dynamical systems. This was done in fundamental works by Ya. Sinai [76], D. Ruelle [67] and R. Bowen [11] in 1972-76. The resulting theory was a fascinating alloy of physical, measure-theoretic, geometric, and topological ideas and methods. For a while, it was the most active topic in the theory of dynamical systems. By 1980, the theory of Anosov and Axiom A systems was completed almost to perfection.

In the eighties, researchers turned their attention to nonuniformly hyperbolic diffeomorphisms and flows. Basic results in this direction were obtained by Ya. Pesin already in mid-seventies [56, 57]. However, those systems happened to be not nearly as nice or easy to study as uniformly hyperbolic ones. The progress in this direction is due to works by M. Benedics, L. Carleson, M. Jakobson, A. Katok, F. Ledrappier, Ya. Pesin, D. Ruelle, M. Viana, L.-S. Young and others in the eighties and nineties. By now it seems that we gained some reasonably good understanding of nonuniformly hyperbolic dynamics, even though there are still more open questions than answers in this area.

At the same time, a revival of the interest to Anosov and Axiom A systems occured in statistical mechanics in the nineties. It was discovered that several physical phenomena of nonequilibrium origin can be modelled with the help of Anosov and Axiom A systems. Most notably, those are entropy production and chaotic scattering studies by G. Gallavotti, P. Gaspard, D. Ruelle, and others. This new interest led to new results about old Anosov and Axiom A systems, for example, Gallavotti-Cohen fluctuation theorem for entropy production rates, and escape-rate formula for Axiom A maps and Anosov diffeomorphisms with holes. Also motivated by applications in physics, new results on the decay of correlations for Anosov flows were obtained.

The survey is organized as follows. Section 1 contains the construction of Markov partitions and symbolic dynamics – this is the very basis for the theory of Gibbs measures. Section 2 thoroughly covers the theory of Gibbs measures for Anosov and Axiom A diffeomorphisms. It is divided into three subsections. One (2.1) gives the classical theory of Gibbs measures for one-dimensional lattices in statistical mechanics. Based on this, Subsection 2.2 develops the modern theory of Gibbs measures for Anosov and Axiom A maps. Subsection 2.3 contains many properties of Gibbs measures – statistical, topological, and other (some of them very recent). Section 3 is devoted to one special Gibbs measure – the so called Sinai-Ruelle-Bowen (SRB) measure. The theory of Gibbs measures in Sections 2 and 3 is presented in great detail and with complete proofs. We feel that this is appropriate in view of the recent revival of interest to the subject, and given that very few books and surveys give full coverage of the topic (probably none since 1980). Sections 2 and 3 can be used by graduate students to study the theory of Gibbs measures. Many remarks and examples can be regarded are good exercises.

Section 4 covers Anosov and Axiom A flows and their Gibbs measures, including SRB measures. It also contains recent results on the decay of correlations. Section 5 is devoted to the volume compression in Anosov and Axiom A diffeomorphisms, it is motivated by the entropy production studies in statistical mechanics. Section 6 contains the discussion of nonuniformly hyperbolic diffeomorphisms and recent studies of the leakage of mass near Axiom A basic sets and Anosov diffeomorphisms with holes, including the escape-rate formula and conditionally invariant measures.

1 Markov partitions

Let $T : M \to M$ be an Anosov diffeomorphism of class $C^{1+\alpha}$ of a compact C^∞ smooth Riemannian manifold M. For any point $x \in M$ we denote by E^s_x, E^u_x the stable and unstable linear subspaces in $\mathcal{T}_x M$, and by $W^s(x)$, $W^u(x)$ the local stable and unstable manifolds through the point x. For exact definitions, see 6.4.a in [44] and more in [38]

Definition. A subset $R \in M$ is called a *rectangle* if its diameter is small (compared to the shortest closed geodesic on M) and for any $x, y \in R$

$$[x, y] := W^s(x) \cap W^u(y) \in R$$

For $x \in R$, let

$$W^s(x, R) = W^s(x) \cap R \quad \text{and} \quad W^u(x, R) = W^u(x) \cap R$$

Any rectangle R has a direct product structure in the following sense. Pick $z \in R$ and put $C = W^u(z, R)$ and $D = W^s(z, R)$. Then

$$R = [C, D] := \{[x, y] : x \in C, \ y \in D\}$$

Also, for every $x \in R$ there is a unique representation $x = [x^u, x^s]$, where $x^u \in C$ and $x^s \in D$, so the sets C and D play the role of coordinate axes in R. Observe that $x \in \operatorname{int} R$ if and only if $x^u \in \operatorname{int} C$ and $x^s \in \operatorname{int} D$. In particular, R is open (closed) if both C and D are open (closed).

A rectangle R is said to be *proper*, if it is closed and $R = \overline{\operatorname{int} R}$. In the above notations, R is proper if and only if $C = \overline{\operatorname{int} C}$ and $D = \overline{\operatorname{int} D}$ as subsets of the manifolds $W^u(x)$ and $W^s(x)$, respectively.

The intersection of a finite number of open (closed) rectangles is an open (closed) rectangle. It is not true, however, that the intersection of proper rectangles is a proper one.

A subset $R' \subset R$ of a proper rectangle R is called a u-subrectangle if it is a proper rectangle itself and $W^u(x, R') = W^u(x, R)$ for all $x \in R'$. A u-subrectangle $R' \subset R$ stretches completely across R in the unstable direction(s). Similarly, s-subrectangles are defined.

We say that a finite collection $\mathcal{R} = \{R_1, \ldots, R_m\}$ of proper rectangles makes a partition of M, if $\cup R_i$ covers M and $\operatorname{int} R_i \cap \operatorname{int} R_j = \emptyset$ for $i \neq j$.

Definition. A partition $\mathcal{R}$ of M into proper rectangles is called a *Markov partition* if it satisfies the following *Markov property*:

$$TW^s(x, R_i) \subset W^s(Tx, R_j) \tag{1.1}$$

and

$$T^{-1}W^u(Tx, R_j) \subset W^u(x, R_i) \tag{1.2}$$

for $x \in \operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j)$.

Let $x \in \operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j)$. Then $TW^s(x, R_i)$ may not coincide with $W^s(Tx, TR_i \cap R_j)$, if R_i if not small enough – the set TR_i may somehow wrap around the manifold M and 'cross R_j twice'. We will always assume that the diameter of every rectangle is small enough, so the above 'double crossing' will not be possible. Thus, we can sharpen (1.1) and (1.2):

$$TW^s(x, R_i) = W^s(Tx, TR_i \cap R_j) \tag{1.3}$$

and

$$T^{-1}W^u(Tx, R_j) = W^u(x, R_i \cap T^{-1}R_j) \tag{1.4}$$

for all $x \in \operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j)$. Therefore,

$$R_i \cap T^{-1}R_j = [T^{-1}W^u(Tx, R_j), W^s(x, R_i)] \tag{1.5}$$

and

$$TR_i \cap R_j = [W^u(Tx, R_j), TW^s(x, R_i)] \tag{1.6}$$

Based on this, one can verify that $R_i \cap T^{-1}R_j$ is a proper rectangle. Hence, $R_i \cap T^{-1}R_j$ is an s-subrectangle in R_i, and $TR_i \cap R_j$ is a u-subrectangle in R_j. By using continuity arguments, one can show that the above identities (1.1)-(1.4) hold also for any point $x \in R_i \cap T^{-1}R_j$ provided $\operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j) \neq \emptyset$.

Remark 1.1. The Markov property (1.1), (1.2) is *equivalent* to the following one: $TR_i \cap R_j$ is a u-subrectangle in R_j and $R_i \cap T^{-1}R_j$ is an s-subrectangle in R_i whenever $\operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j) \neq \emptyset$.

We note that if $\operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j) \neq \emptyset$, then for any u-subrectangle $R' \subset R_i$ the set $TR' \cap R_j$ is a u-subrectangle in R_j. Similarly, any s-subrectangle $R'' \subset R_j$ the set $R_i \cap T^{-1}R''$ is an s-subrectangle in R_i.

One can verify directly that for any closed rectangle R

$$\partial R = \partial^s R \cup \partial^u R$$

where

$$\partial^s R = \cup_{x \in R} \partial W^u(x, R)$$

$$\partial^u R = \cup_{x \in R} \partial W^s(x, R)$$

for $W^u(x, R)$ and $W^s(x, R)$ as subsets of the manifolds $W^u(x)$ and $W^s(x)$, respectively.

We put $\partial^u \mathcal{R} = \cup_i \partial^u R_i$ and $\partial^s \mathcal{R} = \cup_i \partial^s R_i$. Also, $\partial \mathcal{R} = \partial^s \mathcal{R} \cup \partial^s \mathcal{R}$.

Observe that for any partition $\mathcal{R}$ of M into proper rectangles the set $\partial \mathcal{R}$ is closed and nowhere dense.

One can now verify directly, that if $\mathcal{R}$ is a Markov partition, then

$$T\left(\partial^s \mathcal{R}\right) \subset \partial^s \mathcal{R} \quad \text{and} \quad T^{-1}\left(\partial^u \mathcal{R}\right) \subset \partial^u \mathcal{R} \tag{1.7}$$

Remark 1.2. Assume that $\operatorname{int} R_i$ is connected for each $R_i \in \mathcal{R}$. Then (1.7) implies the Markov property. The connectedness of $\operatorname{int} R_i$ is actually essential.

Remark 1.3. Let $\mathcal{R}$ be a Markov partition. Assume that $\operatorname{int} R_i$ has a finite number of connected components for each $R_i \in \mathcal{R}$. Then the closures of the connected components of all R_i, $R_i \in \mathcal{R}$, are proper rectangles which make another (finer) Markov partition. It will satisfy the assumptions of the previous remark.

Let $T : M \to M$ be an Axiom A diffeomorphism, see 6.4.a in [44] and [38] for exact definitions and basic properties. We follow here earlier notation by Bowen [11]. Denote by $\Omega(T)$ the set of its nonwandering points. According to Smale's decomposition theorem,

$$\Omega(T) = \Omega_1 \cup \cdots \cup \Omega_s$$

where Ω_i are *basic sets*. Let Ω_r be a basic set for T. One can define rectangles $R \subset \Omega_r$ and Markov partitions of Ω_r in exactly the same way as above. All the results (except Remarks 1.2 and 1.3) hold true for basic sets of Axiom A diffeomorphisms almost without change. The openness and closedness of a rectangle R and sets $W^{u,s}(x, R)$ is understood as of subsets of Ω_r and $W^{u,s}(x) \cap \Omega_r$, respectively. Exact definitions and detailed proofs can be found in [11]. Remarks 1.2 and 1.3 make no sense in the Axiom A case, since basic sets are often totally disconnected (this is best illustrated by Smale's horseshoe).

Furthermore, let Ω be a compact locally maximal topologically transitive hyperbolic set, see 6.4.a in [44]. Then Ω has all the characteristic properties of an Axiom A basic set. Hence, rectangles and Markov partitions are defined for Ω just as well. Since we will only work on basic sets, hyperbolic locally maximal transitive sets will not be different from Axiom A basic sets. So, we also call them Axiom A basic sets.

In our further discussion, we cover both types of diffeomorphisms, Anosov and Axiom A (it goes without saying that the latter includes compact locally maximal transitive hyperbolic sets). Whenever necessary, we separate these cases, however.

Define a transition matrix $A = A(\mathcal{R})$ by

$$A_{ij} = \begin{cases} 1 & \text{if } \ \mathrm{int}\, R_i \cap T^{-1}(\mathrm{int}\, R_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

Proposition 1.1 *Let $i_0, i_1, \ldots, i_n$ be an arbitrary sequence such that $A_{i_r,i_{r+1}} = 1$ for all $0 \le r \le n-1$. Then the set $R' = \cap_{r=0}^{n} T^{-r} R_{i_r}$ is an s-subrectangle in R_{i_0}, and*

$$W^u(x, R') = T^{-n} W^u(T^n x, R_{i_n}), \qquad \forall x \in R' \tag{1.9}$$

Similarly, the set $R'' = \cap_{r=0}^{n} T^{n-r} R_{i_r}$ is a u-subrectangle in R_{i_n}, and

$$W^s(y, R'') = T^n W^s(T^{-n} y, R_{i_0}), \qquad \forall y \in R'' \tag{1.10}$$

Proof. See (1.5), (1.6), Remark 1.1 and use induction on n. □

Remark 1.4. Observe that the sets given in (1.9) and (1.10) are closed and their diameters do not exceed $C\lambda^n \mathrm{diam}\, \mathcal{R}$, where $\mathrm{diam}\, \mathcal{R} = \max_i \mathrm{diam}\, R_i$.

Example 1.1. Let $T : \mathbf{T}^2 \to \mathbf{T}^2$ be the hyperbolic toral automorphism defined by the matrix $A_T = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. It is often colloquially called "Arnold's cat map". A Markov partition for T consisting of three rectangles is shown in Fig. 1. For an exercise, the reader may want to find the transition matrix for this partition.

If the matrix A_T is not symmetric, the stable and unstable lines for T on the torus $\mathbf{T}^2$ may not be orthogonal. Then, the atoms of Markov partitions are, geometrically, parallelograms rather than rectangles. In early works on Markov partitions, [74, 75], the term 'parallelogram' was used instead of 'rectangle'.

Next, let $\mathcal{R}' = \{R'_i\}$ and $\mathcal{R}'' = \{R''_j\}$ be two partitions of M into proper rectangles. For any i, j the intersection $R'_i \cap R''_j$ is a rectangle, which is not necessarily a proper one. However, the set $R_{ij} = \overline{\mathrm{int}\,(R'_i \cap R''_j)}$ is a proper

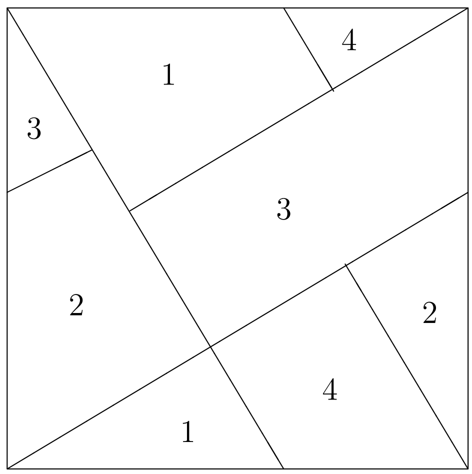

Figure 1: A Markov partition for a hyperbolic toral automorphism.

rectangle (or empty). Then the collection of nonempty rectangles $\{R_{ij}\}$ makes a partition into proper rectangles, the fact one can verify directly. We denote this partition by $\mathcal{R}' \vee \mathcal{R}''$.

Observe that if $\mathcal{R}'$ and $\mathcal{R}''$ are Markov partitions, then so is $\mathcal{R}' \vee \mathcal{R}''$. Hence, if $\mathcal{R}'$ is a Markov partition, then for any $m, n \geq 0$ the partition $\vee_{i=-m}^{n} T^i \mathcal{R}'$ is a Markov one, as well.

Let $m \geq 1$. Assume that $\mathcal{R}'$ is a Markov partition for the diffeomorphism T^m. One can verify directly that

$$\mathcal{R} = \mathcal{R}' \vee T^{-1}\mathcal{R}' \vee \cdots \vee T^{-m+1}\mathcal{R}'$$

is a Markov partition for T.

Theorem 1.2 *Let $T : M \to M$ be an Anosov diffeomorphism or $T : \Omega \to \Omega$ a map on an Axiom A basic set. Then for any $\varepsilon > 0$ there is a Markov partition of M (resp., Ω) whose rectangles have diameters less than ε.*

Two proofs of this theorem exist. The original (but little known) proof by Sinai [75] works only for Anosov diffeomorphisms, but is more explicit and provides a somewhat better control on the construction of Markov partitions. It was first to be extended to hyperbolic dynamical systems of physical interest, such as billiards [14, 15]. The other proof, by Bowen [11], is based on the shadowing property. It is very elegant, works for Anosov and Axiom A diffeomorphisms, and was recently extended to very general hyperbolic systems with singularities [45] and nonuniformly hyperbolic systems [46]. We outline both proofs.

Proof of Theorem 1.2: Anosov case. For simplicity, we assume that $\dim M = 2$, so that the manifolds W^u and W^s are one-dimensional (we call them fibers). For any curve $\gamma \subset M$ denote by $|\gamma|$ its length. For any two points x, y on the same local stable or unstable fiber denote by $W^s(x, y)$ and $W^u(x, y)$, respectively, the segment of that fiber between x and y.

Let $\delta_0 \ll \delta_1 \ll 1$. Let $\mathcal{N}_{\delta_0} = \{x_1, \ldots, x_n\}$ be a δ_0-dense set in M. For each $1 \leq i \leq n$ let $y_{i,1}, y_{i,2}$ be two points on the local stable fiber $W^s(x_i)$ that are the distance δ_1 from x_i, i.e. $|W^s(y_{i_1}, y_{i,2})| = 2\delta_1$ and $|W^s(x_i, y_{i,j})| = \delta_1$ for $j = 1, 2$. Let $x_{i,1}, x_{i,2} \in \mathcal{N}_{\delta_0}$ be two points that are δ_0-close to $y_{i,1}, y_{i,2}$, respectively. Put $z_{i,j} = [x_i, x_{i_j}]$ for $j = 1, 2$, $W^s_{i,0} = W^s(z_{i,1}, z_{i,2})$, and $S_0 = \cup_{i=1}^n W^s_{i,0}$. Similarly, define unstable segments $W^u_{i,0}$ and put $U_0 = \cup_{i=1}^n W^u_{i,0}$.

Due to the compactness of M, there is a constant $\kappa = \kappa(M, T) > 0$ such that $|W^s(y_{i,j}, z_{i,j})| \leq \kappa\delta_0$ and $|W^u(x_{i,j}, z_{i,j})| \leq \kappa\delta_0$ for $j = 1, 2$. We then have the following facts:
(i) $2\delta_1 - \kappa\delta_0 \leq |W^s_{i,0}| \leq 2\delta_1 + \kappa\delta_0$;
(ii) each curve $W^s_{i,0}$ terminates on two curves $W^u_{i_1,0}, W^u_{i_2,0} \subset U_0$, and the distance from the endpoints of $W^s_{i,0}$ to the corresponding endpoints of $W^u_{i_1,0}$ and $W^u_{i_2,0}$ is $\geq \delta_1 - 2\kappa\delta_0$.
These facts, of course, hold true if one interchanges the superscripts u and s. We assume that $\kappa\delta_0 < \delta_1/100$.

According to the previous observations, it is enough to show the existence of Markov partitions for T^m with very large values of m. The set $S_0 \cup U_0$ defines a partition $\mathcal{R}_0$ of M into proper connected rectangles with $\partial^s\mathcal{R}_0 = S_0$ and $\partial^u\mathcal{R}_0 = U_0$, which is not a Markov partition for T^m yet. We will also need that $T^m S_0 \subset S_0$ and $T^{-m} U_0 \subset U_0$ (according to Remark 1.2, this will be enough). We will adjust the curves in S_0 and U_0 to ensure the above inclusions.

Let m be so large that $C\lambda^m < \delta_0/\delta_1$. Then, for any $1 \leq i \leq n$, $|T^m W^s_{i,0}| < 3\delta_0$ and $|T^m W^u_{i_j,0}| > \delta_1^2/\delta_0 > 100\delta_1$ for $j = 1, 2$. The curves $T^m W^u_{i_j,0}$, $j = 1, 2$, are like long parallel railroad tracks on which the short curve $T^m W^s_{i,0}$ can be moved back and forth to allow adjustment. Let $x_{i_3} \in \mathcal{N}_{\delta_0}$ be a point δ_0-close to $T^m W^s_{i,0}$. Put $z'_{i,j} = W^s_{i_3,0} \cap T^m W^u_{i_j,0}$ for $j = 1, 2$ and $W^s_{i,1} = T^{-m} W^s(z'_{i,1}, z'_{i,2})$. Observe that, for each i, two parallel stable fibers $W^s_{i,1}$ and $W^s_{i,0}$ lie just the distance $\leq \text{const} \cdot \delta_0 C\lambda^m < \text{const} \cdot \delta_0^2/\delta_1$ apart, which is $< \delta_0/100$ provided δ_0/δ_1 is small enough.

Similarly, one can define $W^u_{i,1}$. Note, however, that the new curves $W^s_{i,1}$ and $W^u_{i,1}$ may not terminate on each other, so we have to cut back or extend them slightly (by no more than $\delta_0/100$ to restore that property.

Put $S_1 = \cup_{i=1}^n W^s_{i,1}$ and $U_1 = \cup_{i=1}^n W^u_{i,1}$. Observe that $T^m S_1 \subset S_0$ and $T^{-m} U_1 \subset U_0$. The above adjustment is just the first step of a iterational procedure. Having two systems of curves $\{W^s_{i,p}\}$ and $\{W^u_{i,p}\}$, $1 \leq i \leq n$, for some $p \geq 0$, we can adjust them by the above algorithm and get two new systems $\{W^s_{i,p+1}\}$ and $\{W^u_{i,p+1}\}$. Observe that $T^m S_{p+1} \subset S_p$ and $T^{-m} U_{p+1} \subset U_p$, where the meaning of the notations is clear.

It is not so hard to verify that our iterational procedure converges (exponentially fast in p), and we arrive at two limit systems of stable and unstable fibers, $\{W^s_{i,\infty}\}$ and $\{W^u_{i,\infty}\}$, that satisfy the required inclusions, $T^m S_\infty \subset S_\infty$

and $T^{-m}U_\infty \subset U_\infty$. These define a Markov partition for T^m. The theorem is proved.

Remark 1.5. We produced a Markov partition into *connected* rectangles. Each one is a domain in M bounded by two local stable and two local unstable fibers. Since $\partial^s\mathcal{R}$ consists of a finite number of local stable fibers and we have the invariance (1.7), the set $K^s = \cap_n T^n(\partial^s\mathcal{R})$ is finite and T-invariant. So, K^s consists of a finite number of periodic points, whose stable fibers contain the entire set $\partial^s\mathcal{R}$. The same goes to the set $K^u = \cap_n T^{-n}(\partial^u\mathcal{R})$. The finite set $K^s \cup K^u$ is called the *core* of the partition $\mathcal{R}$. For the partition shown in Fig. 1, the core is just one fixed point $(0,0)$. We note also that for Anosov diffeomorphisms on two-dimensional tori there is a direct construction of Markov partitions by Adler and Weiss [5].

Multidimensional Anosov case. The above proof can be carried out in any dimensions, along the lines of [75], but we will not do that here. We only point out a new problem arising in higher dimensions. To illustrate it, assume that $\dim W^u(x) = 2$ and $\dim W^s(x) = 1$ for all $x \in M$. Again, the first step of the proof is to take a partition $\mathcal{R}_0$ of M into nice piecewise smooth connected rectangles, for example, 3-D boxes of more or less the same size. To ensure the Markov property, an iterational procedure of adjustments of the boundaries of the boxes is used. Let $R_0 \in \mathcal{R}_0$, fix an $x \in R_0$ and put $C_0 = W^u(x, R_0)$. If m is large, the surface T^mC_0 will be large in all directions, and its boundary $\partial(T^mC_0)$ will go through many boxes $R \in \mathcal{R}_0$. The first adjustment consists of carving a new boundary curve on the surface $W^u(x)$, close to ∂C_0, whose image under T^m will fall into $\partial^s\mathcal{R}_0$. That new curve will enclose a new region, $C_1 \subset W^u(x)$. Observe that the new boundary curve, ∂C_1, is far more irregular than the old one, ∂C_0, since it has to mimic, on a small scale, the boundaries of the boxes $R \in \mathcal{R}_0$ that $T^m(\partial C_0)$ goes through. Further adjustments will create more and more fractal type irregularities on the boundary curves ∂C_n, $n \geq 1$, on smaller and smaller scales. As a result, the limit region $C_\infty = \lim_{n\to\infty} C_n$ will have boundary that is not smooth at any single point. The careful construction of C_∞, therefore, requires lengthy and delicate work, see [75]. The nonsmoothness of the boundaries of Markov partitions in high dimensions was first noticed by Bowen and proved in [12] by a different technique (counting periodic points).

Proof of Theorem 1.2: Axiom A case. We outline the proof leaving the verification of some claims as exercises. A detailed proof is available in ref.

[11], see also [43], p.592.

Let Ω_r be a an Axiom A basic set. For a small $\beta > 0$ find an $\alpha > 0$ such that any α-pseudo-orbit is β-shadowed by an orbit in Ω_r. Find a $\gamma < \alpha/2$ such that $\text{dist}(Tx, Ty) < \alpha/2$ when $\text{dist}(x, y) < \gamma$. Let $P = \{p_1, \ldots, p_r\}$ be a γ-dense set in Ω_r. Note that

$$\Sigma(P) = \{\underline{\omega} \in \prod_{i=-\infty}^{\infty} \{1, \ldots, r\} : \quad \text{dist}(Tp_{\omega_i}, p_{\omega_{i+1}}) < \alpha \quad \forall\, i \in \mathbb{Z}\}$$

is a topological Markov chain of finite type. For each α-pseudo-orbit $\underline{\omega} \in \Sigma(P)$ there is a unique point $\pi(\underline{\omega}) \in \Omega_r$ that β-shadows it.

Claim. The map $\pi : \Sigma(P) \to \Omega_r$ is surjective and continuous. Hint: prove the continuity by way of contradiction, using the expansiveness of $T : \Omega_r \to \Omega_r$.

As a result, the sets $V_i := \pi(C_i)$, where $C_i = \{\underline{\omega} : \omega_0 = i\}$ is a cylinder whose 0-th coordinate is set to i, are closed for all $1 \le i \le r$. For $\underline{\omega}, \underline{\omega}' \in C_i$ define $\underline{\omega}^* = [\underline{\omega}, \underline{\omega}']$ by

$$\omega_j^* = \begin{cases} \omega_j & \text{for } j \ge 0 \\ \omega_j' & \text{for } j \le 0 \end{cases}$$

Then $[\cdot, \cdot]$ commutes with π, that is $\pi([\underline{\omega}, \underline{\omega}']) \in W^s(\pi(\underline{\omega})) \cap W^u(\pi(\underline{\omega}')) = [\pi(\underline{\omega}), \pi(\underline{\omega}')]$. Hence, V_i are closed rectangles. They cover Ω_r.

Claim. The cover $\{V_i\}$ satisfies the following *semi-Markov* property: if $x = \pi(\underline{\omega})$ for $\underline{\omega} \in \Sigma(P)$, then

$$TW^s(x, V_{\omega_0}) \subset W^s(Tx, V_{\omega_1}) \quad \text{and} \quad T^{-1}W^u(Tx, V_{\omega_1}) \subset W^u(x, V_{\omega_0}) \quad (1.11)$$

To prove this, note that for any $y = \pi(\underline{\omega}') \in W^s(x, V_{\omega_0})$ with $\omega_0' = \omega_0$ we have $y = [x, y] = \pi(\underline{\omega}, \underline{\omega}')$.

The rectangles V_i may overlap and not be proper, so we have to cut them into smaller pieces. For each $x \in \Omega_r$ put $\mathcal{V}(x) = \{V_j : x \in V_j\}$ and $\mathcal{V}^*(x) = \{V_k : V_k \cap V_j \neq \emptyset \text{ for some } V_j \in \mathcal{V}(x)\}$. The collection $\{V_i\}$ makes a finite cover of Ω_r by closed sets, so $\cup \partial V_i$ is nowhere dense in Ω_r. Moreover, the set

$$Z^* = \{x \in \Omega_r : W^s(x) \cap \partial^s V = \emptyset,\ W^u(x) \cap \partial^u V = \emptyset, \quad \forall V \in \mathcal{V}(x) \cup \mathcal{V}^*(x)\}$$

is open and dense in Ω_r.

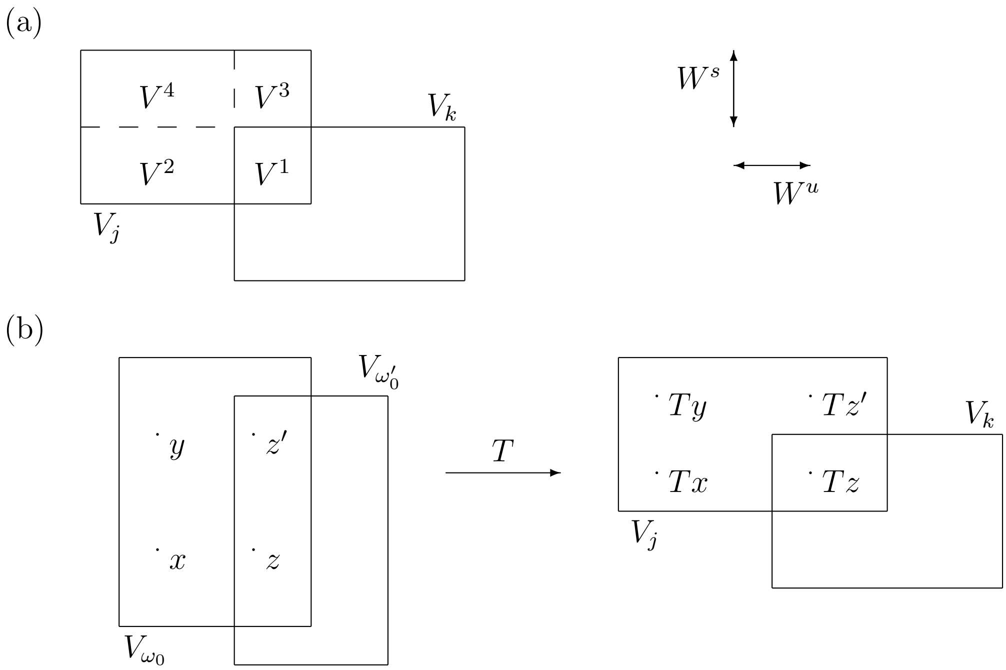

Figure 2: (a) The cutting; (b) A contradiction.

For $V_j \cap V_k \neq \emptyset$ we put (Fig. 2a)

$$\begin{aligned} V_{j,k}^1 &= \{x \in V_j : W^u(x, V_j) \cap V_k \neq \emptyset,\ W^s(x, V_j) \cap V_k \neq \emptyset\} = V_j \cap V_k \\ V_{j,k}^2 &= \{x \in V_j : W^u(x, V_j) \cap V_k \neq \emptyset,\ W^s(x, V_j) \cap V_k = \emptyset\} \\ V_{j,k}^3 &= \{x \in V_j : W^u(x, V_j) \cap V_k = \emptyset,\ W^s(x, V_j) \cap V_k \neq \emptyset\} \\ V_{j,k}^4 &= \{x \in V_j : W^u(x, V_j) \cap V_k = \emptyset,\ W^s(x, V_j) \cap V_k = \emptyset\} \end{aligned}$$

Claim. Each $V_{j,k}^t$ is a rectangle, and for each $x \in V_j \cap Z^*$ and $V_k \in \mathcal{V}^*(x)$ we have $x \in \operatorname{int} V_{j,k}^t$ for some t. The proof is a direct inspection.

For $x \in Z^*$, we put

$$R(x) = \cap\{\operatorname{int} V_{j,k}^t : x \in V_{j,k}^t\}$$

and

$$\mathcal{R} = \{\overline{R(x)} : x \in Z^*\} := \{R_1, \ldots, R_m\}$$

which will be the desired Markov partition.

Claim. Each R_i, $1 \le i \le m$, is a proper rectangle, because it is a closure of an open set, $R(x)$.

Claim. The union $\cup R_i$ covers Ω_r, since each point x of the dense open set Z^* is covered by $\overline{R(x)} \in \mathcal{R}$.

Claim. The sets $\operatorname{int} R_1, \ldots, \operatorname{int} R_m$ are disjoint. To prove this, show that if $y \in R(x) \cap Z^*$, then $R(y) = R(x)$. First check that $\mathcal{V}(y) = \mathcal{V}(x)$, then note that y must belong in the same sets $V_{k,j}^t$ as x does.

It remains to verify the Markov property (1.1), after that (1.2) will follow by considering T^{-1}.

Let $x, y \in Z^* \cap T^{-1}Z^*$ and $y \in W^s(x, R(x))$. We want to show that $R(Tx) = R(Ty)$.

Claim. $\mathcal{V}(Tx) = \mathcal{V}(Ty)$. Indeed, assume that $Tx \in V_j$, find an $\underline{\omega} \in \pi^{-1}(x)$ such that $\omega_1 = j$, and use the semi-Markov property (1.11) to show that $Ty \in V_j$.

Claim. Tx and Ty belong in the same rectangles $V_{j,k}^t$. If not, then $\exists V_k \in \mathcal{V}^*(Tx)$ such that $W^u(Ty, V_j) \cap V_k = \emptyset$ while $\exists z : Tz \in W^u(Tx, V_j) \cap V_k$, see Fig. 2b. Find an $\underline{\omega} \in \pi^{-1}(x)$ such that $\omega_1 = j$ and an $\underline{\omega}' \in \pi^{-1}(z)$ such that $\omega_1' = k$. Use the semi-Markov property (1.11) to show that $z \in W^u(x, V_{\omega_0})$. Hence, $V_{\omega_0'} \in \mathcal{V}^*(x)$, and the fact $R(x) = R(y)$ implies that $z' := [z, y] \in V_{\omega_0'}$. Use (1.11) again to show that $Tz' \in V_k$, a contradiction.

The last two claims imply that for $x \in Z^* \cap T^{-1}Z^*$

$$T\Big(W^s(x, R(x)) \cap Z^* \cap T^{-1}Z^*\Big) \subset W^s(Tx, R(Tx))$$

Now (1.1) follows by continuity, since the set $W^s(x, R(x)) \cap Z^* \cap T^{-1}Z^*$ is open and dense as a subset of $W^s(x, R(x))$ and the set $Z^* \cap T^{-1}Z^*$ is open and dense in every rectangle $R_i \in \mathcal{R}$. □

Symbolic dynamics. Let $\mathcal{R} = \{R_1, \ldots, R_m\}$ be a Markov partition and A its transition matrix (1.8). The latter defines a topological Markov chain (TMC)

$$\Sigma_A = \{\underline{\omega} \in \prod_{i=-\infty}^{\infty} \{1, \ldots, m\} : A_{\omega_i \omega_{i+1}} = 1 \text{ for all } -\infty < i < \infty\}$$

with a left shift homeomorphism $\sigma : \Sigma_A \to \Sigma_A$ defined by $(\sigma(\underline{\omega}))_i = \omega_{i+1}$.

Theorem 1.3 *For any sequence $\underline{\omega} \in \Sigma_A$ the set $\cap_{n=-\infty}^{\infty} T^{-n} R_{\omega_n}$ consists of exactly one point that we denote by $\pi(\underline{\omega})$. The map $\pi : \Sigma_A \to M$ is continuous, surjective, $\pi \circ \sigma = T \circ \pi$ and π^{-1} is one-to-one on the T-invariant residual set $M \setminus M^{\#}$, where*

$$M^{\#} := \cup_{n=-\infty}^{\infty} T^n(\partial \mathcal{R})$$

is a countable union of closed nowhere dense sets.

The theorem is stated for Anosov systems. Its statement for Axiom A basic sets is obtained by replacing M with Ω_r.

Proof. The existence and uniqueness of the point $\pi(\underline{\omega})$ follows from Proposition 1.1 and Remark 1.4. The rest can be verified directly.

Let $-\infty < n' \le n'' < \infty$ and $\omega_{[n',n'']} = \omega_{n'}\omega_{n'+1}\cdots\omega_{n''}$ be an admissible word, which means that $A_{\omega_i \omega_{i+1}} = 1$ for all $n' \le i < n''$. The set

$$C(\omega_{[n',n'']}) = \{\underline{\omega}' \in \Sigma_A : \omega'_i = \omega_i \text{ for all } n' \le i \le n''\}$$

is called a cylinder. Observe that $\pi(C(\omega_{[n',n'']}))$ is a proper rectangle. In particular, for any $n \ge 1$ the set $\pi(C(\omega_{[0,n]}))$ is an s-subrectangle in R_{ω_0} and $\pi(C(\omega_{[-n,0]}))$ is a u-subrectangle in R_{ω_0}. By allowing either $n' = -\infty$ or $n'' = \infty$ (but not both) in the above definition, we get a one-sided infinite sequence

$\omega_{(-\infty,n'']}$ or $\omega_{[n',+\infty)}$, respectively. Then for any $n \in \mathbb{Z}$ the set $\pi(C(\omega_{(-\infty,n]}))$ is a compact subset of a global unstable manifold, and $\pi(C(\omega_{[n,+\infty)}))$ is a compact subset of a global stable manifold.

Remark 1.6. Let $x \in M \setminus M^{\#}$ and $\underline{\omega} = \pi^{-1}x$. For every $n \in \mathbb{Z}$ denote by $\omega_{(-\infty,n]} = \{\omega_i\}_{-\infty}^{n}$ and $\omega_{[n,+\infty)} = \{\omega_i\}_{n}^{+\infty}$ one-sided subsequences of $\underline{\omega}$. Then $\cup_n \pi(C(\omega_{(-\infty,n]}))$ is the global unstable manifold through the point x, and $\cup_n \pi(C(\omega_{[n,+\infty)}))$ is the global stable manifold through the point x. This is not necessarily true if $x \in M^{\#}$.

The space Σ_A is equipped with the product topology, in which σ is a homeomorphism. Following some authors, e.g. [28], we will say that the shift $\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive if there is a dense *positive semiorbit* (rather than a dense orbit, as in some standard definitions of topological transitivity). A TMC (Σ_A, σ) is then topologically transitive iff for any $1 \le i, j \le m$ there is an admissible word that starts with i and ends with j. In terms of the matrix A, the transitivity means that for all i, j there is $k = k_{ij} \ge 1$ such that the (i, j)-th entry of A^k is positive, in which case A is called an *irreducible* matrix. The TMC (Σ_A, σ) is topologically mixing iff there is $k \ge 1$ such that for all i, j one can find an admissible word that start with i and end with j and has length k, i.e. the length of the connecting word is independent of i, j. In this case $A^k > 0$, i.e. A^k contains no zeroes, and A is said to be *aperiodic* or *primitive*.

Let Ω_r be an Axiom A basic set and $\mathcal{R}$ its Markov partition. The topological properties of $T : \Omega_r \to \Omega_r$ are related to those of the corresponding TMC (Σ_A, σ) with the help of Smale's decomposition theorem. One can verify directly that σ is always topologically transitive, and furthermore, it is topologically mixing if and only if so is the map $T : \Omega_r \to \Omega_r$.

Consider now an Anosov diffeomorphism $T : M \to M$ and its Markov partition $\mathcal{R}$. Just as above, one can verified directly that the corresponding TMC (Σ_A, σ) is topologically transitive if and only if so is $T : M \to M$, in which case the set of nonwandering points $\Omega(T)$ for the diffeomorphism T is the entire manifold M. It is not known if nontransitive Anosov diffeomorphisms (such that $\Omega(T) \neq M$) exist. In any case, our first proof of Theorem 1.2 provides a Markov partition of M without assuming that $\Omega(T) = M$.

Let $T : M \to M$ be a transitive Anosov diffeomorphism. Then $\Omega(T) = M$, and it follows from Smale's decomposition theorem that $T : M \to M$, and hence $\sigma : \Sigma_A \to \Sigma_A$, is topologically mixing, because M cannot be possibly decomposed into a finite union of disjoint closed sets.

Finally, we note that for a transitive Anosov diffeomorphism $T : M \to M$ every global stable manifold W^s is dense in M (this follows from Remark 1.6).

2 Gibbs Measures

We now turn to invariant measures for Anosov and Axiom A diffeomorphisms. Let $\mathcal{R}$ be a Markov partition for an Anosov diffeomorphism $T : M \to M$ or an Axiom A basic set Ω_r. Let (Σ_A, σ) be the corresponding symbolic system, i.e. a topological Markov chain (TMC). Let ρ be any σ-invariant measure on Σ_A. Then its projection on M (Ω_r) will be a T-invariant measure. This easily follows from Theorem 1.3. It is therefore natural to study σ-invariant measures on TMC's first.

It so happened that such studies originally started in statistical physics, for quite different reasons. Gibbs states were introduced in statistical physics by Dobrushin [23, 24, 25] in the end of sixties and were later rediscovered by Lanford and Ruelle [47]. Because of its origin in physics, the theory of σ-invariant measures on TMC's carries a strong physics flavor and uses terminology borrowed from mechanics, such as energy, potential, pressure, etc.

We present here two independent (but equivalent) versions of the theory of Gibbs measures. One is the original, born in statistical physics, it defines Gibbs states through interactions, Gibbs ensembles and thermodynamic limit. We do this in Subsection 2.1. The other is a later version adopted to the needs of dynamical system theory, it defines Gibbs measures through potentials and topological pressure. This is done in Subsection 2.2, where we also prove the equivalence of both versions for topologically mixing TMC's. The two versions complement each other and both are necessary to prove the whole spectrum of fascinating properties of Gibbs measures, which are collected in our Subsection 2.3.

2.1 Gibbs states

We introduce basic terminology and notation used throughout this section, for the convenience of future references.

Notation. We fix a finite set $S = \{1, \ldots, m\}$. Let $\Omega = S^{\mathbb{Z}}$ denote the space of doubly infinite sequences of elements of S equipped with the product topology. For any $\Lambda \subset \mathbb{Z}$ put $\Omega_\Lambda = S^\Lambda$. We denote by ω_Λ elements of Ω_Λ (one can think of ω_Λ as a function $\Lambda \to S$, or as a sequence $\{\omega_i\}$ labelled by $i \in \Lambda$). For any $X \subset \Lambda$ we denote by $\omega_\Lambda|_X \in \Omega_X$ the restriction of ω_Λ to X. For any $\underline{\omega} \in \Omega$ denote by $\underline{\omega}|_X$ the restriction of $\underline{\omega}$ to X. For $X \subset \Lambda$, the map $\pi_{X\Lambda} : \omega_\Lambda \mapsto \omega_\Lambda|_X$ is a projection $\Omega_\Lambda \to \Omega_X$. For brevity, we write π_Λ for $\pi_{\Lambda Z}$, i.e. $\pi_\Lambda : \underline{\omega} \mapsto \underline{\omega}|_\Lambda$. For two disjoint subsets $X, Y \subset \mathbb{Z}$ we denote by $\omega_X \vee \omega_Y$ an element of $\Omega_{X \cup Y}$ whose restrictions on X and Y coincide with ω_X and ω_Y, respectively. We also put $\text{dist}(X, Y) = \min\{|i - j| : i \in X, \, j \in Y\}$. Let $\Lambda^c = \mathbb{Z} \setminus \Lambda$. For $\Lambda \subset \mathbb{Z}$ and $k \in \mathbb{Z}$ we put $\Lambda - k = \{i - k : i \in \Lambda\}$ the translate of Λ and denote by $\sigma^k(\omega_\Lambda) \in \Omega_{\Lambda - k}$ the corresponding translate of $\omega_\Lambda \in \Omega_\Lambda$. If $\Lambda = [n', n''] := \{n', n' + 1, \ldots, n''\}$ for some $n' \le n''$, we call Λ an *interval* (in $\mathbb{Z}$) of length $|\Lambda| := n'' - n' + 1$. Observe that $\pi_\Lambda^{-1}(\omega_\Lambda) = C(\omega_\Lambda)$ is a cylinder, as it was defined in the previous section. We denote by $\mathcal{F}$ the set of all finite nonempty subsets of $\mathbb{Z}$, and by $\mathcal{I} \subset \mathcal{F}$ the set of all intervals in $\mathbb{Z}$ (including intervals of length one, i.e. single point sets $\{i\} \subset \mathbb{Z}$). We call elements $s \in S$ *states*, and elements $\omega_\Lambda \in \Omega_\Lambda$ *configurations* (of states) on Λ. In physics interpretation, this means that each site $i \in \Lambda$ is in a certain state, $\omega_i \in S$.

Denote by $\mathcal{M}$ the space of Borel probability measures on Ω. For any $\Lambda \subset \mathbb{Z}$ let $\mathcal{M}_\Lambda$ denote the space of Borel probability measures on Ω_Λ. For any $X \subset \Lambda \subset \mathbb{Z}$ the projection $\pi_{X\Lambda} : \Omega_\Lambda \to \Omega_X$ defines a map $\mathcal{M}_\Lambda \to \mathcal{M}_X$ that we denote by the same symbol, $\pi_{X\Lambda}$. Again, we write π_Λ instead of $\pi_{\Lambda Z}$, for brevity. Obviously, any measure $\mu \in \mathcal{M}$ is uniquely determined by the set of its 'finite-dimensional' projections $\{\pi_\Lambda \mu : \Lambda \in \mathcal{F}\}$. Furthermore, any collection of measures $\{\mu_\Lambda\}$, $\mu_\Lambda \in \mathcal{M}_\Lambda$, for all finite $\Lambda \subset \mathbb{Z}$, determines a unique measure $\mu \in \mathcal{M}$ (in the sense $\pi_\Lambda \mu = \mu_\Lambda$) if and only if the 'finite dimensional distributions' agree, i.e. iff $\pi_{X\Lambda} \mu_\Lambda = \mu_X$ whenever $X \subset \Lambda$.

Now, let A be an $m \times m$ transition matrix that defines the space $\Sigma_A \subset \Omega$. For any $\Lambda \subset \mathbb{Z}$ let

$$\Sigma_\Lambda := \pi_\Lambda(\Sigma_A)$$

This is the set of configurations on Λ that can be extended to admissible sequences in Σ_A.

We note that, from the standpoint of modern statistical mechanics, we only cover Gibbs measures on a one-dimensional lattice $\mathbb{Z}$. The theory of

Gibbs measures on higher-dimensional lattices $\mathbb{Z}^d$, $d \geq 2$, is substantially different and more complicated, see, e.g. a discussion in [68].

Now we introduce the very basic definitions.

Interactions. An *interaction* is a real valued function Φ on $\cup_{\Lambda\in\mathcal{I}}\Sigma_\Lambda$, i.e. on the set of all admissible configurations on intervals. We have two standing assumptions on interactions:
(Φ1) Φ is shift invariant, i.e. for all $\Lambda \in \mathcal{I}$ and $k \in \mathbb{Z}$ we have $\Phi(\sigma^k\omega_\Lambda) = \Phi(\omega_\Lambda)$.
(Φ2) Φ decays exponentially, i.e.

$$||\Phi|| := \sup_n \theta^{-n} \sup\{|\Phi(\omega_\Lambda)| : \Lambda \in \mathcal{I}, |\Lambda| = n\} < +\infty \tag{2.1}$$

for some $0 < \theta < 1$.

We now fix an interaction Φ satisfying (Φ1) and (Φ2) with some θ.

Energy. Let $\Lambda \in \mathcal{F}$. The *energy* of a configuration $\omega_\Lambda \in \Sigma_\Lambda$ is

$$U(\omega_\Lambda) = \sum_{X\subset\Lambda, X\in\mathcal{I}} \Phi(\omega_\Lambda|_X)$$

Next, let $\Lambda \in \mathcal{F}$ and $M \subset \Lambda^c$ such that $\mathrm{dist}(\Lambda, M) = 1$. Let $\omega_\Lambda \in \Sigma_\Lambda$ and $\omega_M \in \Sigma_M$ be two configurations such that $\omega_\Lambda \vee \omega_M \in \Sigma_{\Lambda\cup M}$. Then the *energy of interaction between* ω_Λ *and* ω_M is defined by

$$W(\omega_\Lambda, \omega_M) = \sum_{X\in\mathcal{I}}{}^{*} \Phi(\omega_\Lambda \vee \omega_M|_X) \tag{2.2}$$

where Σ^* is taken over intervals $X \subset \Lambda \cup M$ such that $X \cap \Lambda \neq \emptyset$ and $X \cap M \neq \emptyset$. If $\mathrm{dist}(\Lambda, M) \geq 2$, then the sum (2.2) is empty. In this case we define another quantity similar to (2.2) as follows. Fix an 'external' configuration $\xi \in \Sigma_{(\Lambda\cup M)^c}$ such that $\omega_\Lambda \vee \omega_M \vee \xi \in \Sigma_A$. Now define the *full energy of interaction between* ω_Λ *and* ω_M, given the external configuration ξ, by

$$E(\omega_\Lambda, \omega_M, \xi) = \sum_{X\in\mathcal{I}}{}^{**} \Phi(\omega_\Lambda \vee \omega_M \vee \xi|_X)$$

where Σ^{**} is taken over all intervals $X \subset \mathbb{Z}$ such that $X \cap \Lambda \neq \emptyset$ and $X \cap M \neq \emptyset$.

Lemma 2.1 *Let $\Lambda \in \mathcal{I}$ be any interval of length $|\Lambda| = n$. Then*

$$|U(\omega_\Lambda)| \le n||\Phi||(1-\theta)^{-1} \tag{2.3}$$

for any $\omega_\Lambda \in \Sigma_\Lambda$. Next, let $M \cap \Lambda = \emptyset$ and dist$(\Lambda, M) = 1$. Then

$$\begin{aligned} |W(\omega_\Lambda, \omega_M)| &\le 2||\Phi||(\theta + 2\theta^2 + 3\theta^3 + \cdots) \\ &= 2||\Phi||\theta(1-\theta)^{-2} \end{aligned} \tag{2.4}$$

for any $\omega_M \in \Omega_M$ such that $\omega_\Lambda \vee \omega_M \in \Sigma_{\Lambda\cup M}$. Also, let dist$(\Lambda, M) = d \ge 2$. Then

$$|E(\omega_\Lambda, \omega_M, \xi)| \le 2||\Phi||\theta^d(1-\theta)^{-2} \tag{2.5}$$

for any $\xi \in \Sigma_{(\Lambda\cup M)^c}$ such that $\omega_\Lambda \vee \omega_M \vee \xi \in \Sigma_A$. Observe that (2.3) grows at most linearly in n, while (2.4) is uniformly bounded, and (2.5) is uniformly small as $d \to \infty$.

The proof is left as an exercise.

Gibbs ensembles. Let $\Lambda \in \mathcal{F}$ be a finite set. The *Gibbs ensemble* (without boundary condition) is a probability measure on the finite set Σ_Λ defined by

$$\mu_\Lambda(\omega_\Lambda) = Z_\Lambda^{-1} \exp\left[-U(\omega_\Lambda)\right] \tag{2.6}$$

where Z_Λ is the normalizing factor called the *partition function*:

$$Z_\Lambda = \sum_{\omega_\Lambda \in \Sigma_\Lambda} \exp\left[-U(\omega_\Lambda)\right]$$

Observe that $\mu_\Lambda(\omega_\Lambda) > 0$ for all $\omega_\Lambda \in \Sigma_\Lambda$.

Now, let $\Lambda \in \mathcal{F}$, and $\xi \in \Sigma_{\Lambda^c}$. The *Gibbs ensemble with boundary condition ξ* is a probability measure on the set Σ_Λ defined by

$$\mu_{\Lambda,\xi}(\omega_\Lambda) = Z_{\Lambda,\xi}^{-1} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda, \xi)\right] \tag{2.7}$$

where we set $W(\omega_\Lambda, \xi) = +\infty$ in the case $\omega_\Lambda \vee \xi \notin \Sigma_A$. Here again $Z_{\Lambda,\xi}$ is the normalizing factor:

$$Z_{\Lambda,\xi} = \sum_{\omega_\Lambda \in \Sigma_\Lambda} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda, \xi)\right]$$

For any $\xi \in \Sigma_{\Lambda^c}$, let

$$\Sigma_{\Lambda,\xi} := \{\omega_\Lambda \in \Sigma_\Lambda : \omega_\Lambda \vee \xi \in \Sigma_A\}$$

Observe that $\mu_{\Lambda,\xi}(\omega_\Lambda) > 0$ if and only if $\omega_\Lambda \in \Sigma_{\Lambda,\xi}$, i.e. the support of the measure $\mu_{\Lambda,\xi}$ is $\Sigma_{\Lambda,\xi} \subset \Sigma_\Lambda$.

Remark 2.1. We can extend both measures μ_Λ and $\mu_{\Lambda,\xi}$ from Σ_Λ to Ω_Λ by setting $\mu_\Lambda(\omega) = \mu_{\Lambda,\xi}(\omega) = 0$ for all $\omega \in \Omega_\Lambda \setminus \Sigma_\Lambda$.

Thermodynamic limits. Let $\{\Lambda_i\}_{i=1}^\infty$ be a sequence of finite subsets of $\mathbb{Z}$. We write $\Lambda_i \to \mathbb{Z}$, as $i \to \infty$, if for any finite $\Lambda \in \mathcal{F}$ we have $\Lambda \subset \Lambda_i$ for all $i \geq i_\Lambda$.

Let $\Lambda_i \to \mathbb{Z}$. Consider the Gibbs ensembles μ_{Λ_i}. For every $\Lambda \in \mathcal{F}$ the measure $\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i} \in \mathcal{M}_\Lambda$ is defined whenever $\Lambda \subset \Lambda_i$, i.e. for all sufficiently large i. We call a measure $\rho \in \mathcal{M}$ a *thermodynamic limit* of the Gibbs ensembles $\{\mu_{\Lambda_i}\}$ if for each $\Lambda \in \mathcal{F}$ the sequence of measures $\{\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i}\}$ weakly converges to the measure $\pi_\Lambda\rho$.

Next, let $\xi \in \Sigma_A$ be a fixed configuration, and again $\Lambda_i \to \mathbb{Z}$. For each i, put $\xi_i = \xi|_{\Lambda_i^c}$ and consider the Gibbs ensemble μ_{Λ_i,ξ_i} with boundary condition ξ_i. We call a measure $\rho_\xi \in \mathcal{M}$ a *thermodynamic limit* of Gibbs ensembles μ_{Λ_i,ξ_i}, if for each $\Lambda \in \mathcal{F}$ the sequence of measures $\{\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i,\xi_i}\}$ weakly converges to the measure $\pi_\Lambda\rho_\xi$.

Remark 2.2. One can easily verify that any thermodynamic limit measure ρ is concentrated on Σ_A, i.e. $\rho(\Sigma_A) = 1$. The same is true for any ρ_ξ. Indeed, note that $(\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i})(\Sigma_\Lambda) = 1$, hence $(\pi_\Lambda\rho)(\Sigma_\Lambda) = \rho(\pi_\Lambda^{-1}(\Sigma_\Lambda)) = 1$, and $\Sigma_A = \cap\{\pi_\Lambda^{-1}\Sigma_\Lambda : \Lambda \in \mathcal{F}\}$.

Theorem 2.2 *There is at least one thermodynamic limit measure ρ. For each $\xi \in \Sigma_A$, at least one thermodynamic limit measure ρ_ξ exists.*

Proof. Let $\Lambda_i \to \mathbb{Z}$. For each $\Lambda \in \mathcal{F}$ the sequence of measures $\{\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i}\}$ on the finite set Σ_Λ contains a subsequence that converges to a probability measure on Σ_Λ. Since there are countably many $\Lambda \in \mathcal{F}$, we can apply a standard diagonal argument and find a subsequence $\{i_k\}$ such that for every finite Λ the sequence of measures $\{\pi_{\Lambda\Lambda_{i_k}}\mu_{\Lambda_{i_k}}\}$ converges to a probability measure, ρ_Λ, on Σ_Λ. The measures $\{\rho_\Lambda\}$ obviously agree, i.e. $\pi_{\Lambda M}\rho_M = \rho_\Lambda$ for all finite $\Lambda \subset M$. Hence, the collection $\{\rho_\Lambda\}$ defines a unique probability measure $\rho \in \mathcal{M}$ such that $\pi_\Lambda \rho = \rho_\Lambda$ for all $\Lambda \in \mathcal{F}$. The measure ρ is a thermodynamic limit of the Gibbs ensembles $\mu_{\Lambda_{i_k}}$. The existence of ρ_ξ is proved by exact same argument. □

Remark 2.3. Later we will see that for any mixing TMC there is in fact exactly one thermodynamic limit measure, $\rho = \rho_\xi$, independent of the boundary condition ξ. It only depends on the interaction Φ. We will arrive at this conclusion much later, however.

The picture is quite different for nonmixing TMC's, as the following examples show.

Example 2.1. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The set Σ_A consists of just two sequences of alternating 1's and 2's. Call them $\underline{\omega}^{(1)}$ and $\underline{\omega}^{(2)}$ so that $\omega_0^{(i)} = i$. The shift σ is transitive but not mixing. For simplicity, let $\Phi \equiv 0$. Then $\rho(\underline{\omega}^{(1)}) = \rho(\underline{\omega}^{(2)}) = 1/2$. But the measure ρ_ξ depends on $\xi \in \Sigma_A$. Check that $\rho_\xi(\xi) = 1$ for each $\xi \in \Sigma_A$. Note that ρ_ξ is not σ-invariant for either $\xi = \underline{\omega}^{(1)}$ or $\xi = \underline{\omega}^{(2)}$.

Example 2.2. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The set Σ_A is now countable, it contains all sequences starting with all ones and ending with all twos. It also contains two constant sequences, call them $\underline{\omega}^{(1)}$ and $\underline{\omega}^{(2)}$, consisting of all ones and all twos, respectively. The shift σ is topologically transitive in the regular sense (a dense orbit exists) but not in the sense of our version of transitivity: there is no dense semiorbit, the matrix A is not irreducible. For simplicity, let $\Phi \equiv 0$. Now ρ is not unique, and depending on the sequence $\{\Lambda_i\}$ one can get any measure ρ such that $\rho(\underline{\omega}^{(1)}) = p$ and $\rho(\underline{\omega}^{(2)}) = q$ with $p + q = 1$. The same is true for ρ_ξ, if only $\xi \neq \underline{\omega}^{(1)}$ and $\xi \neq \underline{\omega}^{(2)}$.

Gibbs states. A probability measure ν on Ω is called a *Gibbs state* if for every $\Lambda \in \mathcal{F}$ there exists a probability measure ν_{Λ^c} on Σ_{Λ^c} such that for all $\omega_\Lambda \in \Omega_\Lambda$

$$(\pi_\Lambda \nu)(\omega_\Lambda) = \int_{\Sigma_{\Lambda^c}} \mu_{\Lambda,\omega'}(\omega_\Lambda)\, d\nu_{\Lambda^c}(\omega') \tag{2.8}$$

where $\mu_{\Lambda,\omega'}$ is the Gibbs ensemble with boundary condition, as defined by (2.7). Here $\omega' \in \Sigma_{\Lambda^c}$ is the variable of integration.

Remark 2.4. Note that $\nu(\Sigma_A) = 1$ for any Gibbs state ν. One can argue as in Remark 2.2.

Lemma 2.3 (i) *Let* $f_i : \Sigma_{\Lambda_i} \to \mathbb{R}$ *be some functions. Assume that the sequence of functions* $f_i \circ \pi_{\Lambda_i}$ *on* Σ_A *uniformly converges to a continuous function* $f : \Sigma_A \to \mathbb{R}$ *as* $i \to \infty$ *(and* $\Lambda_i \to \mathbb{Z}$*). Then*

$$\lim_{i\to\infty} \sum_{\omega\in\Sigma_{\Lambda_i}} f_i(\omega)\mu_{\Lambda_i}(\omega) = \int_{\Sigma_A} f(\omega)\, d\rho(\omega) \tag{2.9}$$

(ii) *Furthermore, (2.9) remains true if we substitute* $\mu_{\Lambda_i} \mapsto \mu_{\Lambda_i,\xi_i}$ *and* $\rho \mapsto \rho_\xi$.

An important note: in the case (ii) the functions f_i need only be defined on Σ_{Λ,ξ_i}, which is the support of the measure μ_{Λ_i,ξ_i}. The uniform convergence is then understood in the sense $\sup_{\Sigma_{\Lambda_i,\xi_i}} |f_i - f| \to 0$.

The proof is left as an exercise.

Theorem 2.4 *Any thermodynamic limit measure* ρ *of Gibbs ensembles* μ_{Λ_i} *is a Gibbs state. For every configuration* $\xi \in \Sigma_A$*, any thermodynamic limit measure* ρ_ξ *of Gibbs ensembles* μ_{Λ_i,ξ_i} *with boundary condition* $\xi_i = \xi|_{\Lambda_i^c}$ *is a Gibbs state.*

Proof. To prove the first statement, we let $\Lambda \subset \Lambda_i$, $\omega_\Lambda \in \Sigma_\Lambda$ and write

$$\begin{aligned}
(\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i})(\omega_\Lambda) &= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \mu_{\Lambda_i}(\omega_\Lambda \vee \omega') \\
&= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \left[Z_{\Lambda_i}^{-1} e^{-U(\omega')}\right] \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega')\right] \\
&= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \left[(\pi_{\Lambda_i\setminus\Lambda,\Lambda_i}\mu_{\Lambda_i})(\omega')\right] \mu_{\Lambda,\omega'}(\omega_\Lambda)
\end{aligned} \tag{2.10}$$

where for each $\omega' \in \Sigma_{\Lambda_i\setminus\Lambda}$

$$\mu_{\Lambda,\omega'}(\omega_\Lambda) := Z_{\Lambda,\omega'}^{-1} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega')\right]$$

(assuming that $\mu_{\Lambda,\omega'}(\omega_\Lambda) = 0$ if $\omega_\Lambda \vee \omega' \notin \Sigma_{\Lambda_i}$) and $Z_{\Lambda,\omega'}$ is the normalizing factor:

$$Z_{\Lambda,\omega'} := \sum_{\omega_\Lambda\in\Sigma_\Lambda} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega')\right]$$

Observe that the functions on Σ_{Λ^c} defined by $\omega' \mapsto \mu_{\Lambda,\pi_{\Lambda_i\setminus\Lambda}(\omega')}(\omega_\Lambda)$ uniformly converge, as $i \to \infty$, to the continuous function $\omega' \mapsto \mu_{\Lambda,\omega'}(\omega_\Lambda)$. Then we apply the result of Lemma 2.3 (i), adapted to the set $\mathbb{Z} \setminus \Lambda$ instead of $\mathbb{Z}$, and get

$$\lim_{i\to\infty} (\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i})(\omega_\Lambda) = \int_{\Sigma_{\Lambda^c}} \mu_{\Lambda,\omega'}(\omega_\Lambda)\, d(\pi_{\Lambda^c}\rho)(\omega')$$

This is equivalent to (2.8) by the definition of ρ. Observe that we can set $\nu_{\Lambda^c} = \pi_{\Lambda^c}\rho$ in (2.8). This proves the first statement of Theorem 2.4.

To prove the second one, we use our handy notation $\xi_i = \xi|_{\Lambda_i^c}$ and write

$$\begin{aligned}
(\pi_{\Lambda\Lambda_i}\mu_{\Lambda_i,\xi_i})(\omega_\Lambda) &= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \mu_{\Lambda_i,\xi_i}(\omega_\Lambda \vee \omega') \\
&= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \left[Z^{-1}_{\Lambda_i,\xi_i} e^{-U(\omega')-W(\omega',\xi_i)}\right] \times \\
&\qquad \times \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega') - E(\omega_\Lambda,\xi_i,\omega')\right] \\
&= \sum_{\omega'\in\Sigma_{\Lambda_i\setminus\Lambda}} \left[(\pi_{\Lambda_i\setminus\Lambda,\Lambda_i}\mu_{\Lambda_i,\xi_i})(\omega')\right] \mu_{\Lambda,\omega',\xi_i}(\omega_\Lambda) \qquad (2.11)
\end{aligned}$$

where for each $\omega' \in \Sigma_{\Lambda_i\setminus\Lambda}$

$$\mu_{\Lambda,\omega',\xi_i}(\omega_\Lambda) := Z^{-1}_{\Lambda,\omega',\xi_i} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega') - E(\omega_\Lambda,\xi_i,\omega')\right]$$

(assuming that $\mu_{\Lambda,\omega',\xi_i}(\omega_\Lambda) = 0$ if $\omega_\Lambda \vee \omega' \vee \xi_i \notin \Sigma_A$) and $Z_{\Lambda,\omega',\xi_i}$ is the normalizing factor:

$$Z_{\Lambda,\omega',\xi_i} := \sum_{\omega_\Lambda\in\Sigma_\Lambda} \exp\left[-U(\omega_\Lambda) - W(\omega_\Lambda,\omega') - E(\omega_\Lambda,\xi_i,\omega')\right]$$

Then we use Lemma 2.3 (ii). Note that the new extra term $E(\omega_\Lambda,\xi_i,\omega')$ is uniformly small as $i\to\infty$, cf. Lemma 2.1. □

We will see that the Gibbs state is unique for any topologically mixing TMC. For nonmixing TMC's, there may be more than one Gibbs state, as Examples 2.1 and 2.2 show. In this case one can show that the set of Gibbs states is a compact convex subset of $\mathcal{M}$, see [68] for further discussion.

Despite the assumed shift invariance of the interaction Φ, Gibbs states may not be σ-invariant, as Example 2.1 shows. For mixing TMC's, the (unique) Gibbs state is, indeed, σ-invariant, but this fact will be established much later.

Lemma 2.5 *If ν is a Gibbs state, then its image under σ, call it $\nu^* = \sigma^*\nu$, is a Gibbs state. Hence, the set of all Gibbs states is σ-invariant.*

Proof. Since $(\pi_\Lambda \nu^*)(\omega_\Lambda) = (\pi_{\Lambda+1}\nu)(\sigma^{-1}\omega_\Lambda)$, it is enough to note that $\mu_{\Lambda,\omega'}(\omega_\Lambda) = \mu_{\Lambda+1,\sigma^{-1}\omega'}(\sigma^{-1}\omega_\Lambda)$ and put $\nu^*_{\Lambda^c} = \nu_{\Lambda^c+1}$. The details are left to the reader. □

The equation (2.8) is remarkable in that it gives an exact formula for the measures of cylinders, $\nu(C(\omega_\Lambda))$, for any Gibbs state ν. This has many important consequences.

Theorem 2.6 *A probability measure ν on Ω is a Gibbs state if and only if for any finite $\Lambda \subset \mathbb{Z}$ and two configurations $\omega_\Lambda \in \Omega_\Lambda$ and $\omega' \in \Omega_{\Lambda^c}$, the conditional ν-probability that $\underline{\omega}|_\Lambda = \omega_\Lambda$, given that $\underline{\omega}|_{\Lambda^c} = \omega'$, is $\mu_{\Lambda,\omega'}(\omega_\Lambda)$.*

Proof. Let $\Lambda_n = [-n, n]$. For large n, such that $\Lambda \subset \Lambda_n$, put $\omega'_n := \omega'|_{\Lambda_n \setminus \Lambda}$. It is enough to prove that

$$\lim_{n\to\infty} \frac{\nu(\omega_\Lambda \vee \omega'_n)}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \nu(\omega'_\Lambda \vee \omega'_n)} \to \mu_{\Lambda,\omega'}(\omega_\Lambda)$$

Put $\Sigma_{\Lambda^c_n,\omega'_n} = \{\xi \in \Sigma_{\Lambda^c_n} : \omega'_n \vee \xi \in \Sigma_{\Lambda^c}\}$. According to (2.8), it is enough to prove that

$$\lim_{n\to\infty} \sup_{\xi \in \Sigma_{\Lambda^c_n,\omega'_n}} \left| \frac{\mu_{\Lambda_n,\xi}(\omega_\Lambda \vee \omega'_n)}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \mu_{\Lambda_n,\xi}(\omega'_\Lambda \vee \omega'_n)} - \mu_{\Lambda,\omega'}(\omega_\Lambda) \right| \to 0 \qquad (2.12)$$

In virtue of (2.7),

$$\begin{aligned} \frac{\mu_{\Lambda_n,\xi}(\omega_\Lambda \vee \omega'_n)}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \mu_{\Lambda_n,\xi}(\omega'_\Lambda \vee \omega'_n)} &= \frac{\exp[-U(\omega_\Lambda \vee \omega'_n) - W(\omega_\Lambda \vee \omega'_n, \xi)]}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \exp[-U(\omega'_\Lambda \vee \omega'_n) - W(\omega'_\Lambda \vee \omega'_n, \xi])} \\ &= \frac{\exp[-U(\omega_\Lambda) - W(\omega_\Lambda, \omega'_n) - E(\omega_\Lambda, \xi, \omega'_n)]}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \exp[-U(\omega'_\Lambda) - W(\omega'_\Lambda, \omega'_n) - E(\omega'_\Lambda, \xi, \omega'_n)]} \\ &\to \frac{\exp[-U(\omega_\Lambda) - W(\omega_\Lambda, \omega')]}{\sum_{\omega'_\Lambda \in \Omega_\Lambda} \exp[-U(\omega'_\Lambda) - W(\omega'_\Lambda, \omega')]} \end{aligned}$$

as $n \to \infty$, uniformly in $\xi \in \Sigma_{\Lambda^c_n,\omega'_n}$ (again, employ Lemma 2.1). This proves (2.12) and Theorem 2.6. □

In physics, Gibbs states are measures given by their conditional distributions on finite configurations, as in Theorem 2.6, or by a set of equations, such as (2.8), called DLR equations [68].

The following examples cover a few rare cases where the thermodynamic limit measures can be computed directly.

Example 2.3. Assume that $a_{ij} \equiv 1$, so that $(\Sigma_A, \sigma) = (\Omega, \sigma)$ is a full shift. Let $\Phi(\{i\}) = \psi_i$ for $1 \le i \le m$, and $\Phi(\omega_\Lambda) = 0$ for all intervals Λ of length ≥ 2. Then the only thermodynamic limit measure ρ is a Bernoulli measure on Ω, i.e. a direct product of identical measures, p, on S, such that $p(i) = e^{-\psi_i} / \sum_j e^{-\psi_j}$.

Example 2.4. Assume again that $\Sigma_A = \Omega$, a full shift. Let $\Phi(\omega_\Lambda) = 0$ unless $\Lambda = \{n, n+1\}$ is an interval of length two. Put $\Phi(\omega_n, \omega_{n+1}) = \psi_{\omega_n, \omega_{n+1}}$, so that the $m \times m$ matrix (ψ_{ij}) completely determines the interaction Φ. Now the only thermodynamic limit measure ρ is a Markov measure on Ω. To see that, consider the $m \times m$ matrix with entries $b_{ij} = e^{-\psi_{ij}}$. Its entries are positive, so the Perron-Frobenius theorem for positive matrices [33, 73] implies that its largest eigenvalue λ is real positive and simple, and the corresponding left and right eigenvectors, call them $U = (u_i)$ and $V = (v_i)$, respectively, have positive components. Assume that U and V are normalized so that $\sum_i u_i v_i = 1$. By taking a thermodynamic limit one can show that for any word $\omega_0 \cdots \omega_n \in \Omega_{[0,n]}$

$$(\pi_{[0,n]}\rho)(\omega_0 \cdots \omega_n) = \lambda^{-n} u_{\omega_0} b_{\omega_0\omega_1} b_{\omega_1\omega_2} \cdots b_{\omega_{n-1}\omega_n} v_{\omega_n}$$

This is obviously a Markov measure with transition probabilities $\pi_{ij} = \lambda^{-1} b_{ij} v_j / v_i$ and stationary vector $p_i = u_i v_i$.

Example 2.5. We can adapt the previous example to any proper subshift, $\Sigma_A \neq \Omega$, which is mixing. In this case the characteristic matrix $B = (b_{ij})$ has entries $b_{ij} = a_{ij} e^{-\psi_{ij}}$. The Perron-Frobenius theorem applies since $B^k > 0$ for some $k > 0$. In particular, for $\Phi \equiv 0$ we obtain a Markov measure, also called Parry measure, on Σ_A, cf. 3.3.f in [44]. For this measure λ is the largest eigenvalue of the transition matrix A, since $B = A$.

We now discuss more detailed properties of Gibbs states. The discussion is restricted to topologically mixing TMC's, which we call just *mixing*, for brevity. Recall that (Σ_A, σ) is mixing iff $A^k > 0$ for some $k \ge 1$.

The following lemma is a modification of a standard one, see Lemma 1.18 in [11].

Lemma 2.7 *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers such that* $|a_{m+n} -$

$a_m - a_n| \leq R$ *for all* $m, n \geq 1$ *and some constant* $R > 0$. *Then* $P := \lim_{n\to\infty} a_n/n$ *exists. Furthermore,* $|a_n - Pn| \leq 2R$ *for all* n.

Proof. Fix an $m \geq 1$. For $n \geq 1$, write $n = km + l$ with $0 \leq l \leq m-1$. Then it follows by induction on k that $|a_n - ka_m - a_l| \leq kR$. Hence,

$$\left|\frac{a_n}{n} - \frac{ka_m}{km+l} - \frac{a_l}{km+l}\right| \leq \frac{kR}{km+l}$$

Letting $n \to \infty$ gives

$$\frac{a_m}{m} - \frac{R}{m} \leq \liminf_n \frac{a_n}{n} \leq \limsup_n \frac{a_n}{n} \leq \frac{a_m}{m} + \frac{R}{m}$$

Hence, $P := \lim a_n/n$ exists. Next, assume that $a_m > Pm + 2R$ for some m. Then $a_{2^n m} > 2^n mP + (2^n+1)R$ which follows by induction on n. Hence $\limsup a_n/n \geq P + R/m$, a contradiction. A similar contradiction results from the assumption $a_m < Pm - 2R$. □

Let F and G be any variable quantities. We adopt notation

$$F \asymp G \quad \Longleftrightarrow \quad C_1 \leq F/G \leq C_2$$

for some constants $C_1, C_2 > 0$, that only depend on the interaction Φ and the transition matrix A.

Theorem 2.8 (Partition functions) *Let* (Σ_A, σ) *be mixing. For* $n \geq 1$, *put* $\Lambda_n = [0, n-1]$, *an interval of length* n. *There is a finite limit*

$$P = \lim_{n\to\infty} \frac{1}{n} \ln Z_{\Lambda_n} \tag{2.13}$$

Moreover,

$$Z_{\Lambda_n} \asymp e^{Pn} \tag{2.14}$$

and

$$Z_{\Lambda_n,\xi} \asymp e^{Pn} \tag{2.15}$$

for any $\xi \in \Sigma_{\Lambda_n^c}$.

Definition. We call the number $P = P_\Phi$ the (*topological*) *pressure* of the interaction Φ for the TMC (Σ_A, σ).

Proof. To prove (2.13) and (2.14) it is enough to show that

$$R := \sup_{m,n} |\ln Z_{\Lambda_{m+n}} - \ln Z_{\Lambda_m} - \ln Z_{\Lambda_n}| < \infty$$

and apply the previous lemma to the sequence $a_n = \ln Z_{\Lambda_n}$. So, we need to show that

$$Z_{\Lambda_{m+n}} \asymp Z_{\Lambda_m} Z_{\Lambda_n} \tag{2.16}$$

For fixed n, m, put $\Lambda_n = [0, n-1]$, $\Lambda'_k = [n, n+k-1]$, and $\Lambda'_m = [n+k, n+k+m-1]$, so that $\Lambda_n \cup \Lambda'_k \cup \Lambda'_m = \Lambda_{n+k+m} := [0, n+m+k-1]$. Since $A^k > 0$ is an entirely positive matrix, then for any $\omega_1 \in \Sigma_{\Lambda_n}$ and $\omega_2 \in \Sigma_{\Lambda'_m}$ there is an $\omega' \in \Sigma_{\Lambda'_k}$ such that $\omega_1 \vee \omega' \vee \omega_2 \in \Sigma_{\Lambda_{n+k+m}}$. The number of configurations ω' satisfying the above condition is obviously $\leq |S|^k$ =const. Therefore

$$\begin{aligned} Z_{\Lambda_{n+k+m}} &= \sum_{\omega \in \Sigma_{\Lambda_{n+k+m}}} \exp[-U(\omega)] \\ &= \sum_{\omega_1 \in \Sigma_{\Lambda_n}} \sum_{\omega_2 \in \Sigma_{\Lambda'_m}} \exp[-U(\omega_1)] \exp[-U(\omega_2)] \times \\ &\quad \times \sum_{\omega' \in \Sigma_{\Lambda'_k}}{}^{*} \exp[-U(\omega') - W(\omega', \omega_1 \vee \omega_2)] \end{aligned}$$

where the inner sum $\sum^*$ is taken over ω' such that $\omega_1 \vee \omega' \vee \omega_2 \in \Sigma_{\Lambda_{n+k+m}}$. The last exponential is uniformly bounded away from 0 and ∞, cf. Lemma 2.1. Hence $Z_{\Lambda_{n+k+m}} \asymp Z_{\Lambda_m} Z_{\Lambda_n}$. Redoing the above calculation with $m = 0$ shows that $Z_{\Lambda_{n+k}} \asymp Z_{\Lambda_n}$. Hence $Z_{\Lambda_{m+n+k}} \asymp Z_{\Lambda_{m+n}}$. This proves (2.16), hence (2.13) and (2.14).

To prove (2.15), it is enough to show that

$$Z_{\Lambda_n, \xi} \asymp Z_{\Lambda_{n-2k}} \tag{2.17}$$

because $Z_{\Lambda_{n-2k}} \asymp Z_{\Lambda_n}$. We have

$$\begin{aligned} Z_{\Lambda_n,\xi} &= \sum_{\omega \in \Sigma_{\Lambda_n,\xi}} \exp[-U(\omega) - W(\omega, \xi)] \\ &= \sum_{\omega_1 \in \Sigma_{[k,n-k-1]}} \exp[-U(\omega_1)] \times \\ &\quad \times \sum_{\omega' \in \Sigma_{[0,k-1]}}{}^{*} \sum_{\omega'' \in \Sigma_{[n-k,n-1]}}{}^{*} \exp[-U(\omega') - U(\omega'') - W(\omega_1, \omega' \vee \omega'') - W(\omega, \xi)] \end{aligned}$$

where the inner sums are taken over ω', ω'' such that $\omega' \vee \omega_1 \vee \omega'' \in \Sigma_{\Lambda_n}$. The last exponential is uniformly bounded away from 0 and ∞ (we employ again Lemma 2.1). There is at least one ω' and at least one ω'' in the last two sums, and there are no more than $|S|^{2k}$ =const of pairs of ω', ω''. This proves (2.17) and (2.15). □

The exact formula (2.8) is very helpful, but it has a shortcoming: the integrand may be zero, which happens exactly when $\omega_\Lambda \vee \omega' \notin \Sigma_A$. For mixing TMC's, there is the following improvement of (2.8):

Lemma 2.9 *Let (Σ_A, σ) be mixing and ν be a Gibbs state. Let $\Lambda \in \mathcal{I}$ be an interval of length $|\Lambda| = n$, and let Λ' and Λ'' two intervals of length k adjacent to Λ on both sides, so that $M = \Lambda' \cup \Lambda \cup \Lambda''$ is a longer interval of length $n + 2k$. Then for any $\omega_\Lambda \in \Sigma_\Lambda$*

$$(\pi_\Lambda \nu)(\omega_\Lambda) = \int_{\Sigma_{M^c}} (\pi_{\Lambda M} \mu_{M,\xi})(\omega_\Lambda) \, d\nu_{M^c}(\xi) \tag{2.18}$$

Here

$$(\pi_{\Lambda M} \mu_{M,\xi})(\omega_\Lambda) = \sum_{\omega'}{}^{*} \sum_{\omega''}{}^{*} \mu_{M,\xi}(\omega_\Lambda \vee \omega' \vee \omega'') \tag{2.19}$$

where the sums are taken over $\omega' \in \Sigma_{\Lambda'}$ and $\omega'' \in \Sigma_{\Lambda''}$ such that $\omega_\Lambda \vee \omega' \vee \omega'' \vee \xi \in \Sigma_A$. Furthermore,

$$(\pi_{\Lambda M} \mu_{M,\xi})(\omega_\Lambda) \asymp \exp[-Pn - U(\omega_\Lambda)] \tag{2.20}$$

for any $\xi \in \Sigma_{M^c}$. In particular, the integrand in (2.18) is never zero.

Proof. The equations (2.18) and (2.19) follow from (2.8) immediately. To prove (2.20), we observe three facts: (i) for each pair ω', ω'' in (2.19)

$$\exp[-U(\omega_\Lambda \vee \omega' \vee \omega'') - W(\omega_\Lambda \vee \omega' \vee \omega'', \xi)] \asymp \exp[-U(\omega_\Lambda)]$$

by the Lemma 2.1, (ii) $Z_{M,\xi} \asymp e^{Pn+2Pk} \asymp e^{Pn}$ due to (2.15). In addition, the mixing of (Σ_A, σ) implies that (iii) for every $\xi \in \Sigma_{M^c}$ there is at least one pair of ω', ω'' in (2.19), and the total number of such pairs does not exceed $|S|^{2k}$ =const. Now (2.20) follows from the facts (i)-(iii) easily. □

Theorem 2.10 *Let* (Σ_A, σ) *be mixing and* ν *be a Gibbs state. There are constants* $c_1, c_2 > 0$ *such that for any interval* $\Lambda \in \mathcal{I}$ *of length* $|\Lambda| = n$ *and* $\omega_\Lambda \in \Sigma_\Lambda$

$$c_1 \le \frac{\nu(C(\omega_\Lambda))}{\exp[-Pn - U(\omega_\Lambda)]} \le c_2 \tag{2.21}$$

where $C(\omega_\Lambda) = \pi_\Lambda^{-1}(\omega_\Lambda)$ *is a cylinder for the configuration* ω_Λ.

Proof. This follows from Lemma 2.9, because the integration in (2.18) is just averaging, which preserves the upper and lower uniform bounds obtained in (2.20). □

Remark 2.5. The ν-measure of every cylinder $C(\omega_\Lambda) \subset \Sigma_A$ is positive due to (2.21). Hence $\nu(U) > 0$ for every open set $U \subset \Sigma_A$, i.e. ν has full support on Σ_A.

Theorem 2.11 (Uniqueness of Gibbs states) *Let* (Σ_A, σ) *be mixing. Then the Gibbs state* $\nu = \nu_\Phi$ *is unique.*

Proof. Let ν, ν' be two distinct Gibbs states. Then $\nu - \nu'$ is a signed measure on Σ_A with norm $||\nu - \nu'|| > 0$. Note that

$$\begin{aligned} ||\nu - \nu'|| &= \lim_{n\to\infty} \sum_{\omega_{\Lambda_n} \in \Sigma_{\Lambda_n}} |(\nu - \nu')(C(\omega_{\Lambda_n}))| \\ &= \lim_n \sum_{\omega_{\Lambda_n} \in \Sigma_{\Lambda_n}} \left| \int_{\Sigma_{M_n^c}} (\pi_{\Lambda_n M_n} \mu_{M_n,\xi})(\omega_\Lambda)\, d(\nu_{M_n^c} - \nu'_{M_n^c})(\xi) \right| \end{aligned} \tag{2.22}$$

where $\Lambda_n = [-n, n]$, $M = [-n-k, n+k]$, and we used (2.18).

We now need two lemmas on signed measures.

Lemma 2.12 . *Let* μ *be a signed measure,* $||\mu|| < \infty$*, on a measurable space* X*. Let* $f_1, \dots, f_N$ *be nonnegative real-valued measurable functions on* X *such that* $f_1 + \cdots + f_n = K$ *=const. Then* $|\mu(f_1)| + \cdots + |\mu(f_N)| \le K||\mu||$.

Proof. By the Hahn decomposition theorem [66], there is a partition $X = X_+ \cup X_-$, $X_+ \cap X_- = \emptyset$ such that $\mu = \mu_+ - \mu_-$, where μ_+ and μ_- are (nonnegative) measures concentrated on X_+ and X_-, respectively, and $||\mu|| =$

$||\mu_+|| + ||\mu_-||$. Let $f_i = f_{i+} + f_{i-}$, where $f_{i+} \equiv 0$ on X_- and $f_{i-} \equiv 0$ on X_+. Then $\mu(f_i) = \mu_+(f_{i+}) - \mu_-(f_{i-})$, and the lemma follows from

$$\begin{aligned}\sum_i |\mu(f_i)| &\leq \sum_i \mu_+(f_{i+}) + \sum_i \mu_-(f_{i-}) \\ &= \mu_+\left(\sum_i f_{i+}\right) + \mu_-\left(\sum_i f_{i-}\right) \\ &= K||\mu_+|| + K||\mu_-||\end{aligned}$$

Lemma 2.13 . *Under the conditions of the previous lemma, let $\mu(X) = 0$ and let $f_i(x) \geq \varepsilon f_i(x')$ for all $x, x' \in X$, $1 \leq i \leq N$, and some constant $\varepsilon > 0$. Then $|\mu(f_1)| + \cdots + |\mu(f_N)| \leq (1-\varepsilon)K||\mu||$.*

Proof. Fix an $x' \in X$ and define $g_i(x) = f_i(x) - \varepsilon f_i(x')$ for $1 \leq i \leq N$. Note that $g_i \geq 0$ and $\sum_i g_i = (1-\varepsilon)K$. Since $\mu(X) = 0$, we have $\mu(g_i) = \mu(f_i)$. Then we just apply the previous lemma to the functions g_i. □

We now get back to (2.22). Observe that for each $\omega_{\Lambda_n} \in \Sigma_{\Lambda_n}$

$$(\pi_{\Lambda_n M_n} \mu_{M_n,\xi})(\omega_\Lambda) \geq \varepsilon \cdot (\pi_{\Lambda_n M_n} \mu_{M_n,\xi'})(\omega_\Lambda)$$

for all $\xi, \xi' \in \Sigma_{M_n^c}$ and some constant $\varepsilon > 0$, due to (2.20). Then the last lemma implies

$$||\nu - \nu'|| \leq (1-\varepsilon)||\nu_{M_n^c} - \nu'_{M_n^c}|| \leq (1-\varepsilon)||\nu - \nu'||$$

hence $||\nu - \nu'|| = 0$, so that $\nu = \nu'$. Theorem 2.11 is proved. □

Corollary 2.14 *Let (Σ_A, σ) be mixing. Then the unique Gibbs state ν is σ-invariant.*

Proof: see Lemma 2.5.

Theorem 2.15 *Let (Σ_A, σ) be mixing. Then the unique Gibbs state ν is ergodic.*

Proof. The proof will be based solely on the relation (2.21) and the σ-invariance of ν (the last corollary). This generality of the argument will be essential later.

We will show that for any measurable sets $C_1, C_2 \subset \Sigma_A$

$$\liminf_{n\to\infty} \nu(C_1 \cap \sigma^{-n} C_2) \asymp \nu(C_1)\nu(C_2) \tag{2.23}$$

From this, the ergodicity follows immediately.

As usual, it is enough to prove the so called *cluster property* (2.23) for cylinders, so we assume that $C_1 = C(\omega_{\Lambda_1})$ and $C_2 = C(\omega_{\Lambda_2})$ for some intervals Λ_1, Λ_2 of lengths $|\Lambda_1| = n_1$, $|\Lambda_2| = n_2$ and admissible configurations $\omega_{\Lambda_1}, \omega_{\Lambda_2}$. For large n, we have $l_n :=\text{dist}(\Lambda_1, \Lambda_2 + n) > 2k$. Then we partition the interval between Λ_1 and $\Lambda_2 + n$ into three subintervals: the left one (adjacent to Λ_1) of length k, call it Λ', the right one (adjacent to $\Lambda_2 + n$) of length k, call it Λ'', and the middle one in between of length $l_n - 2k$, call it Λ. Then

$$\begin{aligned}\nu(C(\omega_{\Lambda_1}) \cap \sigma^{-n} C(\omega_{\Lambda_2})) &= \sum_{\omega_\Lambda \in \Sigma_\Lambda} \sum_{\omega'} \sum_{\omega''} \nu(C(\omega_{\Lambda_1} \vee \omega' \vee \omega_\Lambda \vee \omega'' \vee \sigma^{-n}\omega_{\Lambda_2})) \\ &\asymp \sum_{\omega_\Lambda \in \Sigma_\Lambda} \sum_{\omega'} \sum_{\omega''} \exp[-P(n_1 + n_2 + l_n) \\ &\qquad -U(\omega_{\Lambda_1} \vee \omega' \vee \omega_\Lambda \vee \omega'' \vee \sigma^{-n}\omega_{\Lambda_2})] \end{aligned} \tag{2.24}$$

where the inner sums are taken over $\omega' \in \Sigma_{\Lambda'}$ and $\omega'' \in \Sigma_{\Lambda''}$ such that the configuration $\omega_{\Lambda_1} \vee \omega' \vee \omega_\Lambda \vee \omega'' \vee \sigma^{-n}\omega_{\Lambda_2}$ is admissible. There is at least one such pair of ω', ω'', and the number of such pairs is $\leq |S|^{2k}$ =const. Now, observe that

$$|U(\omega_{\Lambda_1} \vee \omega' \vee \omega_\Lambda \vee \omega'' \vee \sigma^{-n}\omega_{\Lambda_2}) - U(\omega_{\Lambda_1}) - U(\omega_\Lambda) - U(\omega_{\Lambda_2})| \leq \text{const}$$

by Lemma 2.1. Also, $e^{-Pl_n} \asymp e^{-P(l_n - 2k)}$, and

$$\sum_{\omega_\Lambda \in \Sigma_\Lambda} \exp[-P(l_n - 2k) - U(\omega_\Lambda)] \asymp \sum_{\omega_\Lambda \in \Sigma_\Lambda} \nu(C(\omega_\Lambda)) = 1$$

due to (2.21). Combining these facts and (2.24) proves (2.23) and Theorem 2.15. □

Remark 2.6. The relation (2.23) actually implies more than ergodicity for ν, it is equivalent to the strong mixing of ν, cf. Proposition 20.3.6 in [43]. We do not need this for the moment, and will later obtain an even stronger property – exponential mixing.

Corollary 2.16 *If Φ_1 and Φ_2 are two interactions satisfying ($\Phi1$) and ($\Phi2$), then their Gibbs states ν_{Φ_1} and ν_{Φ_2} either coincide or are mutually singular. The latter means that there is a $B \subset \Sigma_A$ such that $\nu_{\Phi_1}(B) = 1$ and $\nu_{\Phi_2}(B) = 0$ (of course, the set B can be neither open nor closed).*

2.2 Gibbs measures

The machinery of Gibbs states developed in statistical physics was adapted to hyperbolic dynamical systems with Markov partitions by Sinai [76], Bowen [11] and Ruelle [68]. In their works, an alternative construction of Gibbs measures was developed.

Hölder continuous functions. Let (Σ_A, σ) be a TMC and $\varphi \in \mathcal{C}(\Sigma_A)$ a real-valued continuous function. For each $n \geq 1$ put

$$\text{var}_n\varphi = \sup\{|\varphi(\underline{\omega}) - \varphi(\underline{\omega}')| : \ \omega_i = \omega_i' \ \ \forall |i| \leq n\}$$

Since φ is uniformly continuous, $\text{var}_n\varphi \to 0$ as $n \to \infty$.

We introduce the class of Hölder continuous functions on Σ_A. For any $0 < \alpha < 1$ define a metric d_α on Σ_A such that $d_\alpha(\underline{\omega}, \underline{\omega}') = \alpha^n$, where $n = 1 + \max\{n \geq 0 : \ \omega_i = \omega_i' \ \ \forall |i| < n\}$ (in the case $\omega_0 \neq \omega_0'$ we set $n = 0$). A function φ on Σ_A will be Hölder continuous with respect to any (and then all) d_α if $\text{var}_n\varphi \leq b\,\theta^n$ for some $b > 0$ and $\theta \in (0, 1)$ and all $n \geq 0$. For $0 < \theta < 1$, let

$$\mathcal{H}_\theta(\Sigma_A) = \{\varphi : \ \sup_n \theta^{-n}\text{var}_n\varphi < +\infty\}$$

Obviously, $\mathcal{H}_{\theta_1}(\Sigma_A) \subset \mathcal{H}_{\theta_2}(\Sigma_A)$ if $\theta_1 < \theta_2$. This gives a "filtration" of the space of all Hölder continuous functions

$$\mathcal{H}(\Sigma_A) = \cup_{0<\theta<1}\mathcal{H}_\theta(\Sigma_A)$$

For $n \geq 0$, denote by $S_n\varphi$ the function

$$S_n\varphi(\underline{\omega}) = \sum_{i=0}^{n-1} \varphi(\sigma^i\underline{\omega})$$

called a *partial* (or *ergodic*) *sum.*

Next, let $n \geq 0$ and

$$\mathcal{FR}_n(\Sigma_A) = \{\varphi : \ \varphi(\underline{\omega}) = \varphi(\underline{\omega}') \ \text{ if } \ \omega_i = \omega_i' \ \ \forall |i| \leq n\}$$

Then $\mathcal{FR}(\Sigma_A) = \cup_n \mathcal{FR}_n(\Sigma_A)$ is the class of all "finite range" functions.

Definition of Gibbs measures. Let (Σ_A, σ) be mixing and $\varphi \in \mathcal{H}(\Sigma_A)$. A σ-invariant Borel probability measure ρ on Σ_A is called a *Gibbs measure* for the *potential function* φ if there are constants $c_1, c_2 > 0$ and P such that for any interval $\Lambda = [0, n-1]$ of length n and $\omega_\Lambda \in \Sigma_\Lambda$

$$c_1 \leq \frac{\rho(C(\omega_\Lambda))}{\exp[-Pn + S_n\varphi(\underline{\omega})]} \leq c_2 \tag{2.25}$$

with any $\underline{\omega} \in C(\omega_\Lambda)$.

Remark 2.7. Let $\Lambda = [0, n-1]$ and $\underline{\omega}, \underline{\omega}' \in C(\omega_\Lambda)$. Suppose that $\text{var}_n\varphi \leq b\theta^n$ for all $n \geq 1$. Then

$$|S_n\varphi(\underline{\omega}) - S_n\varphi(\underline{\omega}')| \leq 2b/(1-\theta) = \text{const}$$

This explains why the choice of $\underline{\omega} \in C(\omega_\Lambda)$ in (2.25) does not matter.

Remark 2.8 The requirement of σ-invariance cannot be dropped: there is a measure ρ on some Σ_A that satisfies (2.25) and is *not* σ-invariant. For example, let $\Sigma_A = \Omega$, a full shift, and $\varphi \equiv 0$. Let ρ_0 be the Bernoulli measure with uniform distribution, i.e. $\rho_0(C(i)) = 1/|S|$ for all $1 \leq i \leq |S|$. Then take $\rho = g\rho_0$, where g is any positive density on Ω bounded away from 0 and ∞ and normalized by $\rho_0(g) = 1$.

Definition. Let (Σ_A, σ) be mixing and $\varphi \in \mathcal{H}(\Sigma_A)$. Let $n \geq 1$, $\Lambda = [0, n-1]$, and

$$Z_n(\varphi) = \sum_{\omega_\Lambda \in \Sigma_\Lambda} \exp[\sup_{C(\omega_\Lambda)} S_n\varphi] \tag{2.26}$$

Then the number

$$P_\varphi = \limsup_{n\to\infty} \frac{1}{n} \ln Z_n(\varphi) \tag{2.27}$$

is called the (*topological*) *pressure* of the function φ on Σ_A.

Lemma 2.17 *Let (Σ_A, σ) be mixing, $\varphi \in \mathcal{H}(\Sigma_A)$ and ρ a Gibbs measure for the potential φ. Then the value P in (2.25) coincides with P_φ. Moreover, if we redefine (2.26), more generally, by picking an arbitrary point $\underline{\omega}$ in every cylinder $C(\omega_\Lambda)$ and setting*

$$Z_n(\varphi) = \sum_{\omega_\Lambda \in \Sigma_\Lambda,\, \underline{\omega} \in C(\omega_\Lambda)} \exp[S_n\varphi(\underline{\omega})]$$

then

$$Z_n(\varphi) \asymp e^{P_\varphi n} \tag{2.28}$$

In particular, $P_\varphi = \lim_n \frac{1}{n} \ln Z_n(\varphi)$ *in (2.27).*

Proof. Note that

$$\sum_{\omega_\Lambda \in \Sigma_\Lambda} \rho(C(\omega_\Lambda)) = 1$$

hence (2.25) implies

$$\sum_{\omega_\Lambda \in \Sigma_\Lambda} \exp[-Pn + S_n\varphi(\underline{\omega})] \asymp 1$$

which proves (2.28) with $P = P_\varphi$. Lemma 2.17 follows. □

Theorem 2.18 *Let* (Σ_A, σ) *be mixing,* $\varphi \in \mathcal{H}(\Sigma_A)$ *and* ρ *a Gibbs measure for the potential* φ*. Then the measure-theoretic entropy of* ρ *is given by*

$$h_\rho(\sigma) = P_\varphi - \int_{\Sigma_A} \varphi(\underline{\omega})\, d\rho$$

Proof. Let ξ be the partition of Σ_A into cylinders $C(i) = \{\underline{\omega} \in \Sigma_A : \omega_0 = i\}$, and put

$$\xi_n = \xi \vee \sigma^{-1}\xi \vee \cdots \vee \sigma^{-n+1}\xi$$

Since ξ is a generating partition, then $h_\rho(\sigma) = \lim_{n\to\infty} n^{-1} H_\rho(\xi_n)$.

Note that each atom $B \in \xi_n$ is a cylinder, $B = C(\omega_[0, n-1])$ based on the interval $[0, n-1]$. Therefore, $\rho(B) \asymp \exp[-Pn + S_n\varphi(\underline{\omega}_B)]$ where $\underline{\omega}_B \in B$ is an arbitrary point, and we have

$$\begin{aligned} H_\rho(\xi_n) &= -\sum_{B\in\xi_n} \rho(B) \ln \rho(B) \\ &= -\sum_{B\in\xi_n} \rho(B)\, [-Pn + S_n\varphi(\underline{\omega}_B)] + O(1) \\ &= Pn - \int_{\Sigma_A} S_n\varphi(\underline{\omega})\, d\rho + O(1) \end{aligned} \tag{2.29}$$

where $O(1)$ stands for a uniformly bounded quantity. At the last step, we used Remark 2.7.

Since ρ is σ-invariant, we have

$$\int_{\Sigma_A} S_n\varphi(\underline{\omega})\, d\rho = n \int_{\Sigma_A} \varphi(\underline{\omega})\, d\rho$$

Substituting this in (2.29), dividing by n and letting $n \to \infty$ proves the theorem. □

There is a remarkable similarity between the relations (2.21) and (2.25). In addition, both measures, ν in (2.21) and ρ in (2.25), are σ-invariant. We will use this analogy to prove the following:

Theorem 2.19 *Any Gibbs measure ρ for a potential $\varphi \in \mathcal{H}(\Sigma_A)$ is ergodic.*

Proof. Just like (2.21) and the σ-invariance of ν implied the ergodicity of ν in the proof of Theorem 2.15, we now obtain the ergodicity of ρ. The proof of Theorem 2.15 carries over with only minor technical modifications, which we leave out. □

Theorem 2.20 (Existence and uniqueness of Gibbs measures) *Let (Σ_A, σ) be mixing. For any $\varphi \in \mathcal{H}(\Sigma_A)$ there is a unique Gibbs measure $\rho = \rho_\varphi$.*

Proof. We prove the uniqueness first. Let ρ and ρ' be two Gibbs measures for the same potential φ. Then they have the same value $P = P_\varphi$, due to Lemma 2.17. Hence, (2.25), which holds for both ρ and ρ', implies

$$\frac{c_1}{c_2} \le \frac{\rho(C(\omega_\Lambda))}{\rho'(C(\omega_\Lambda))} \le \frac{c_2}{c_1}$$

for any cylinder $C(\omega_\Lambda) \subset \Sigma_A$. Thus, the measures ρ and ρ' are equivalent (absolutely continuous with respect to each other). Since both are ergodic, they must coincide, cf. 3.6.a in [44].

The existence of ρ will follow from the next theorem. □

Theorem 2.21 (Gibbs measures = Gibbs states) *The class of Gibbs states $\{\nu_\Phi\}$ on Σ_A coincides with that of Gibbs measures $\{\rho_\varphi\}$ in the following sense. For any interaction Φ satisfying (Φ1) and (Φ2) there is a potential $\varphi \in \mathcal{H}(\Sigma_A)$ such that ρ_φ exists, $\rho_\varphi = \nu_\Phi$ and $P_\varphi = P_\Phi$. Conversely, for any potential $\varphi \in \mathcal{H}(\Sigma_A)$ there is an interaction Φ satisfying (Φ1) and (Φ2) such that ν_Φ is the Gibbs measure ρ_φ and $P_\Phi = P_\varphi$.*

Proof. Let Φ be an interaction. Put

$$\varphi(\underline{\omega}) = -\sum_{n=0}^{\infty} \Phi(\underline{\omega}|_{[0,n]}) \tag{2.30}$$

Observe that for $\Lambda = [0, n-1]$ we have $S_n\varphi(\underline{\omega}) = -U(\underline{\omega}|_\Lambda) - W(\underline{\omega}|_\Lambda, \underline{\omega}|_{[n,+\infty)})$. Since the W term is uniformly bounded, cf. Lemma 2.1, we have $\exp[S_n\varphi(\underline{\omega})] \asymp \exp[-U(\underline{\omega}|_\Lambda)]$. Now (2.25) follows from (2.21), hence ρ_φ exists and coincides with ν_Φ.

Conversely, let $\varphi \in \mathcal{H}(\Sigma_A)$ be a potential function. We can write $\varphi = \sum_{n=0}^{\infty} \varphi_n$ where $\varphi_n \in \mathcal{FR}_n(\Sigma_A)$ and $|\varphi_n| \leq \mathrm{var}_{n-1}\varphi$ for all $n \geq 1$. The functions φ_n can be defined inductively in the following way. Given $\varphi_0, \ldots, \varphi_{n-1}$ and a cylinder $C = C(\omega_{[-n,n]}) \subset \Sigma_A$, we pick an arbitrary $\underline{\omega}' \in C$ and put $\varphi_n(\underline{\omega}) \equiv \varphi(\underline{\omega}') - \sum_{i=0}^{n-1} \varphi_i(\underline{\omega}')$ for all $\underline{\omega} \in C$.

Next, we set $\Phi(\omega_{[-n,n]}) = -\varphi_n(\underline{\omega})$ for any $\underline{\omega} \in C(\omega_\Lambda)$. This defines Φ on all intervals $[-n, n]$ and then on their translates $[-n+k, n+k]$ by the property (Φ1). On all the intervals of even length we set Φ to zero. One can easily verify that if $\varphi \in \mathcal{H}_{\theta'}(\Sigma_A)$, then Φ satisfies (Φ2) with $\theta = \theta'^{1/2}$. Then the interaction Φ has a Gibbs state, ν_Φ.

Next, observe that

$$S_n\varphi(\underline{\omega}) = -{\sum_{\Lambda}}^{*} \Phi(\underline{\omega}|_\Lambda)$$

where $\sum^*$ extends over intervals Λ such that $|\Lambda \cap [0, n-1]| > |\Lambda|/2$. It then follows from Lemma 2.1 that

$$|S_n\varphi(\underline{\omega}) + U(\underline{\omega}|_{[0,n-1]})| \leq \text{const}$$

Now (2.25) with $\rho = \nu_\Phi$ follows from (2.21), hence ν_Φ is the Gibbs measure for the potential φ. Note that this proves the existence of a Gibbs measure for any $\varphi \in \mathcal{H}(\Sigma_A)$. $\Box$

Remark 2.9. Given an interaction, Φ, there are many ways to define a potential φ with the same Gibbs measure. For example, we can set $\varphi(\underline{\omega}) = \sum_{n=1}^{\infty} \Phi(\underline{\omega}|_{\Lambda_n})$ where $\{\Lambda_n\}$ is any sequence of intervals of length $|\Lambda_n| = n$ that contain 0. Alternatively, one can put $\varphi(\underline{\omega}) = \sum_{\Lambda \in \mathcal{I}, 0 \in \Lambda} |\Lambda|^{-1} \Phi(\underline{\omega}|_\Lambda)$. Similarly, there are many ways to define an interaction Φ given a potential φ, with the same Gibbs state.

Remark 2.10. The definition (2.30) has an advantage that the potential φ depends only on nonnegative coordinates, i.e. $\varphi(\underline{\omega}) = \varphi(\underline{\omega}')$ if $\omega_i = \omega_i'$ for all $i \geq 0$. Therefore, for any potential function $\varphi' \in \mathcal{H}(\Sigma_A)$ there exists another function, $\varphi \in \mathcal{H}(\Sigma_A)$, that depends only on nonnegative coordinates and such that $\rho_\varphi = \rho_{\varphi'}$.

Theorems 2.19 and 2.20 have an immediate corollary.

Corollary 2.22 *If φ and ψ are two potential functions, then their Gibbs measures ρ_φ and ρ_ψ either coincide or are mutually singular.*

It is interesting to know when $\rho_\varphi = \rho_\psi$.

Definition. We call two functions $\varphi, \psi \in \mathcal{H}(\Sigma_A)$ *cohomologous* (with respect to the shift σ), denoted by $\varphi \sim \psi$ if

$$\varphi = \psi + K - u + u \circ \sigma \tag{2.31}$$

for some constant K and some function $u : \Sigma_A \to \mathbb{R}$. This is obviously an equivalence relation on $\mathcal{H}(\Sigma_A)$.

Theorem 2.23 *Let (Σ_A, σ) be mixing, and $\varphi, \psi \in \mathcal{H}(\Sigma_A)$. The following conditions are equivalent:*
(i) $\rho_\varphi = \rho_\psi$.
(ii) *There is a constant K such that $S_n\varphi(\underline{\omega}) - S_n\psi(\underline{\omega}) = nK$ whenever $\sigma^n\underline{\omega} = \underline{\omega}$.*
(iii) *$\varphi \sim \psi$ in the sense (2.31) with some Hölder continuous function u; see also remarks after the proof.*
(iv) *There are constant K and L such that $|S_n\varphi(\underline{\omega}) - S_n\psi(\underline{\omega}) - nK| \leq L$ for all $\underline{\omega}$ and $n \geq 1$.*
If these conditions hold, then $K = P_\varphi - P_\psi$.

Proof. The implication (iii)⇒(iv) is obvious. To prove (iv)⇒(i), replace $S_n\varphi$ by $S_n\psi$ in (2.25) and observe that c_1, c_2 change by at most a factor of e^L, P decreases by K, and ρ will remain unchanged.

We now prove (i)⇒(ii). Let $\sigma^n\underline{\omega} = \underline{\omega}$ and $j \geq 1$. The bounds (2.25) imply

$$\frac{c_1}{c_2} \leq \frac{\exp[-P_\varphi j + S_j\varphi(\underline{\omega})]}{\exp[-P_\psi j + S_j\psi(\underline{\omega})]} \leq \frac{c_2}{c_1} \tag{2.32}$$

Now put $j = nr$, $r \geq 1$, and notice that $S_j\varphi(\underline{\omega}) = rS_n\varphi(\underline{\omega})$. Substituting these in (2.32) and letting $r \to \infty$ gives (ii) with $K = P_\varphi - P_\psi$.

Lastly, we prove (ii)⇒(iii). Since the shift $\sigma : \Sigma_A \to \Sigma_A$ is mixing, there is an $\underline{\omega} \in \Sigma_A$ whose positive semiorbit $\mathcal{O}_+(\underline{\omega}) := \{\sigma^n \underline{\omega}\}_{n=1}^{\infty}$ is dense in Σ_A. Denote $v = \varphi - \psi - K \in \mathcal{H}(\Sigma_A)$. We define $u : \mathcal{O}_+(\underline{\omega}) \to \mathbb{R}$ by

$$u(\sigma^n \underline{\omega}) = \sum_{j=0}^{n-1} v(\sigma^j \underline{\omega})$$

Note that $\underline{\omega}$ cannot be a periodic point, so the function u is well defined on $\mathcal{O}_+(\underline{\omega})$. There is a standard 'closing' argument showing that u extends to Σ_A by continuity and becomes a Hölder continuous function. It is enough to check that whenever $\sigma^p \underline{\omega}$ and $\sigma^q \underline{\omega}$ agree in places $-r$ to r (i.e. belong in one cylinder $C(\omega_{[-r,r]}) \subset \Sigma_A$), then

$$|u(\sigma^p \underline{\omega}) - u(\sigma^q \underline{\omega})| = \left| \sum_{j=p}^{q-1} v(\sigma^j \underline{\omega}) \right| \leq b\theta^r \tag{2.33}$$

for some constants $b > 0$ and $\theta \in (0,1)$ (we assume, without loss of generality, $p < q$). Consider a periodic point $\underline{\omega}^* \in \Sigma_A$ defined by

$$\omega_i^* = \omega_t \quad \text{for} \ \ i \equiv t \ (\operatorname{mod} q - p)$$

where $p \leq t < q$. Since $\underline{\omega}^*$ is periodic with period $q - p$, we have $\sum_{j=p}^{q-1} v(\sigma^j \underline{\omega}^*) = 0$ by (ii). For each $j = p, \ldots, q-1$, the sequences $\sigma^j \underline{\omega}$ and $\sigma^j \underline{\omega}^*$ agree in places $p - r - j$ through $q + r - j$, hence

$$|v(\sigma^j \underline{\omega}) - v(\sigma^j \underline{\omega}^*)| \leq b_1 \theta_1^{\min\{r+j-p, r+q-j\}}$$

with some $b_1 > 0$ and $\theta_1 \in (0,1)$, by the Hölder continuity of v. Adding up for $j = p, \ldots, q-1$ gives (2.33). Hence, indeed, the extension of u to Σ_A is Hölder continuous. Lastly, note that

$$u(\sigma \underline{\omega}') - u(\underline{\omega}') = v(\underline{\omega}')$$

for all $\underline{\omega}' \in \mathcal{O}_+(\underline{\omega})$, and this equation extends to Σ_A by continuity. □

It is also interesting to investigate the equation (2.31) for u, assuming that $\varphi, \psi \in \mathcal{H}(\Sigma_A)$ are given. Then (2.31) is called a *cohomological equation.*

Let (2.31) hold with an *arbitrary* function u (maybe not even measurable!). Then (2.31) has a Hölder continuous solution u. Indeed, it is enough to verify the condition (ii) of the above theorem.

Next, assume that u, u' are two solutions of the equation (2.31). By inspection one can show that the difference $u - u'$ is constant on each orbit $\{\sigma^n \underline{\omega}\}_{n=-\infty}^{+\infty}$. This observation has a few interesting implications:
(a) A Hölder continuous solution u of the equation (2.31) is unique, up to an additive constant.
(b) If a solution u is continuous, then it is Hölder continuous.
(c) Let u be measurable, and ν any σ-invariant ergodic measure on Σ_A. Then u is Hölder continuous mod 0, i.e. up to a set of ν-measure zero. (This follows from the Birkhoff ergodic theorem.)

Remark 2.11. As it follows now from Remark 2.10, for any potential $\psi \in \mathcal{H}(\Sigma_A)$ there is a cohomologous potential $\varphi \sim \psi$ that depends only on non-negative coordinates. A direct proof of this fact is given by Bowen [11].

Ruelle-Perron-Frobenius operator. We will prove more advanced properties of Gibbs measures, and that requires a different apparatus. Consider the symbolic space

$$\Sigma_A^+ = \{\underline{x} \in \prod_{i=0}^{\infty} \{1, \dots, m\} : A_{x_i x_{i+1}} = 1 \text{ for all } i \geq 0\}$$

of one-sided admissible sequences, with the shift $\sigma : \Sigma_A^+ \to \Sigma_A^+$ defined by $(\sigma(\underline{x}))_i = x_{i+1}$. Now σ is a finite-to-one continuous map of Σ_A^+ onto itself. Note that $\sigma : \Sigma_A^+ \to \Sigma_A^+$ is mixing iff $A^k > 0$ for some $k \geq 1$. Denote by $\mathcal{C}(\Sigma_A^+)$ the space of continuous functions on Σ_A^+. Any function $f \in \mathcal{C}(\Sigma_A^+)$ can be naturally extended to Σ_A by $f(\underline{\omega}) := f(\underline{\omega}|_{[0,+\infty)})$, so that $\mathcal{C}(\Sigma_A^+)$ is identified with a subspace of $\mathcal{C}(\Sigma_A)$ consisting of functions depending on nonnegative coordinates. Then we put $\mathcal{H}(\Sigma_A^+) = \mathcal{C}(\Sigma_A^+) \cap \mathcal{H}(\Sigma_A)$, the space of Hölder continuous functions on Σ_A^+. Similarly, $\mathcal{FR}_n(\Sigma_A^+) = \mathcal{C}(\Sigma_A^+) \cap \mathcal{FR}_n(\Sigma_A)$. For any $\Lambda \subset [0, +\infty)$ we denote by $C^+(\omega_\Lambda) \subset \Sigma_A^+$ the cylinder for a given configuration ω_Λ.

Remark 2.12. Let (Σ_A^+, σ) be mixing and $C^+(\omega_{[0,n]}) \subset \Sigma_A^+$ a cylinder. Then $\sigma^{n+k} C^+(\omega_{[0,n]}) = \Sigma_A^+$.

Let $\varphi \in \mathcal{C}(\Sigma_A^+)$. We define the *Ruelle-Perron-Frobenius operator* $\mathcal{L} = \mathcal{L}_\varphi$ on the space $\mathcal{C}(\Sigma_A^+)$ by

$$(\mathcal{L}_\varphi f)(\underline{x}) = \sum_{\underline{y} \in \sigma^{-1} \underline{x}} e^{\varphi(\underline{y})} f(\underline{y}) \tag{2.34}$$

It has an adjoint operator, $\mathcal{L}^*$, on the space $\mathcal{C}^*(\Sigma_A^+)$ of linear functionals on $\mathcal{C}(\Sigma_A^+)$ defined by

$$(\mathcal{L}^* g)(f) = g(\mathcal{L} f)$$

for all $g \in \mathcal{C}^*(\Sigma_A^+)$ and $f \in \mathcal{C}(\Sigma_A^+)$. Observe that $\mathcal{L}$ takes positive functions to positive functions, hence $\mathcal{L}^*$ takes measures to measures. However, neither $\mathcal{L}$ nor $\mathcal{L}^*$ preserves norm in the corresponding space.

Remark 2.13. By induction on n, one can show that

$$(\mathcal{L}_\varphi^n f)(\underline{x}) = \sum_{\underline{y} \in \sigma^{-n}\underline{x}} e^{S_n \varphi(\underline{y})} f(\underline{y}) \tag{2.35}$$

In particular, if χ_C is the characteristic function of a cylinder $C = C^+(\omega_{[0,n]}) \subset \Sigma_A^+$, then one can use (2.35) to obtain that

$$\mathcal{L}_\varphi^j \chi_C \geq e^{-j||\varphi||} > 0 \tag{2.36}$$

for all $j \geq n + k$.

Remark 2.14. It is a simple exercise to check that for any $f, g \in \mathcal{C}(\Sigma_A^+)$ we have $\mathcal{L}(f \cdot (g \circ \sigma)) = (\mathcal{L} f) \cdot g$. Hence, by induction, $\mathcal{L}^n(f \cdot (g \circ \sigma^n)) = (\mathcal{L}^n f) \cdot g$.

Remark 2.15. The operator (2.34) is a generalization of the classical Perron-Frobenius operator for interval maps. The term $e^{\varphi(\underline{y})}$ plays the role of the reciprocal of the derivative, and f plays the role of a density.

Theorem 2.24 (Ruelle-Perron-Frobenius) *Let (Σ_A^+, σ) be mixing, $\varphi \in \mathcal{H}(\Sigma_A^+)$, and $\mathcal{L} = \mathcal{L}_\varphi$ as defined above. There is a $\lambda > 0$, a function $h \in \mathcal{C}(\Sigma_A^+)$, $h > 0$, and a probability measure ν on Σ_A^+ such that*

$$\mathcal{L} h = \lambda h \quad \text{and} \quad \mathcal{L}^* \nu = \lambda \nu \tag{2.37}$$

and $\nu(h) = 1$. Furthermore,

$$\lim_{n \to \infty} ||\lambda^{-n} \mathcal{L}^n g - \nu(g) \cdot h|| = 0 \tag{2.38}$$

for all $g \in \mathcal{C}(\Sigma_A^+)$.

Note similarities between this theorem and Examples 2.7 and 2.8. The value $\lambda > 0$ is the largest (leading) eigenvalue of $\mathcal{L}$. The function $h > 0$ is the corresponding (right) eigenvector, and the positive functional ν (=measure) is the corresponding adjoint (left) eigenvector. The eigenvectors h and ν are normalized so that $\nu(h) = 1$. We will see later that the invariant (Gibbs) measure is $\rho = h\nu$, which means that ρ is a measure with density h with respect to the (reference) measure ν. In addition, we will show that $P_\varphi = \ln \lambda$.

Proof. Consider a continuous (nonlinear) map $\mu \mapsto ||\mathcal{L}^*\mu||^{-1}\mathcal{L}^*\mu$ on the space $\mathcal{M}^+$ of Borel probability measures on Σ^+. Since $\mathcal{M}^+$ is a compact convex subset of the linear space $\mathcal{C}^*(\Sigma_A^+)$, this continuous map has a fixed point according to the Schauder-Tychonoff theorem, see [27] p. 456. Hence, there exist a measure $\nu \in \mathcal{M}^+$ such that $\mathcal{L}^*\nu = \lambda\nu$ for some $\lambda > 0$. We will see later that the fixed point ν is actually unique.

Remark 2.16. For any cylinder $C = C^+(\omega_{[0,n]}) \subset \Sigma_A^+$

$$\nu(C) \geq \lambda^{-n-k} e^{-(n+k)||\varphi||} > 0 \tag{2.39}$$

Indeed, $\nu(C) = \nu(\chi_C) = \lambda^{-j}\mathcal{L}^{*j}\nu(\chi_C) = \lambda^{-j}\nu(\mathcal{L}^j\chi_C)$, then we can use (2.36).

Remark 2.17. Let $f \in \mathcal{C}(\Sigma_A^+)$, $f \geq 0$ and $\nu(f) = 1$. Consider the probability measure $\mu = f\nu$ (i.e. $\mu(g) = \nu(fg)$ for all $g \in \mathcal{C}(\Sigma_A^+)$). Let $\sigma^*\mu$ be the image of μ under σ. Then $\sigma^*\mu = (\lambda^{-1}\mathcal{L}f)\nu$, i.e. $\lambda^{-1}\mathcal{L}f$ is the density of the measure $\sigma^*\mu$. (Note, in particular, that $\nu(\lambda^{-1}\mathcal{L}f) = \nu(f) = 1$.) Thus, the operator $\lambda^{-1}\mathcal{L}$ transforms the densities of measures with respect to ν under the shift σ, just like the regular Perron-Frobenius operator for interval maps transforms the densities with respect to the Lebesgue measure. To prove this, note: $(\sigma^*\mu)(g) = \mu(g\circ\sigma) = \nu(f\cdot(g\circ\sigma)) = \lambda^{-1}(\mathcal{L}^*\nu)(f\cdot(g\circ\sigma)) = \lambda^{-1}\nu(\mathcal{L}(f\cdot(g\circ\sigma)))$, then use Remark 2.14.

Now we have to find a function $h > 0$ such that $\mathcal{L}h = \lambda h$. It suffices to find a nonempty compact convex set $\mathcal{D} \subset \mathcal{C}(\Sigma_A^+)$ invariant under the linear continuous operator $\lambda^{-1}\mathcal{L}$ and then employ the Schauder-Tychonoff fixed point theorem again. Since $\varphi \in \mathcal{H}(\Sigma_A^+)$, there are $\theta \in (0,1)$ and $b > 0$ such that $\text{var}_n\varphi \leq b\theta^n$ for all $n \geq 0$. Put

$$B_n = \exp\left[\sum_{i=n+1}^{\infty} 2b\theta^i\right]$$

We will write $\underline{x} \overset{n}{\sim} \underline{y}$ if the points $\underline{x}, \underline{y}$ belong in one cylinder $C^+(\omega_{[0,n]}) \subset \Sigma_A^+$, i.e. if the first $n+1$ symbols of $\underline{x}$ and $\underline{y}$ coincide. In particular, $\underline{x} \overset{0}{\sim} \underline{y}$ means $x_0 = y_0$. We define $\mathcal{D}$ to be the set of functions $f \in \mathcal{C}(\Sigma_A^+)$ such that
($\mathcal{D}1$) $f \geq 0$;
($\mathcal{D}2$) $\nu(f) = 1$;
($\mathcal{D}3$) $f(\underline{x}) \leq B_n f(\underline{x}')$ for $\underline{x} \overset{n}{\sim} \underline{x}'$.

Claim 1. $\mathcal{D}$ is a nonempty compact convex set. Clearly, $f \equiv 1$ is in $\mathcal{D}$. The convexity is obvious. For compactness, we first prove that $\mathcal{D}$ is uniformly bounded:

$$||f|| \leq K := B_0 \lambda^k e^{k||\varphi||} \qquad \text{for all} \quad f \in \mathcal{D} \tag{2.40}$$

To prove this, note that for any $\underline{x} \overset{0}{\sim} \underline{x}'$ we have $f(\underline{x}) \geq B_0^{-1} f(\underline{x}')$ by ($\mathcal{D}3$), hence $f(\underline{x}) \geq B_0^{-1} f(\underline{x}') \chi_{C^+(x_0')}$ for all $\underline{x}'$, and so $1 = \nu(f) \geq B_0^{-1} f(\underline{x}') \nu(C^+(x_0'))$, and then we use (2.39). So, $\mathcal{D}$ is uniformly bounded. Furthermore, for any $\underline{x} \overset{n}{\sim} \underline{x}'$ and $f \in \mathcal{D}$ we have

$$|f(\underline{x}) - f(\underline{x}')| \leq (B_n - 1)K \to 0 \quad \text{as} \quad n \to \infty$$

by ($\mathcal{D}3$) and (2.40), so that $\mathcal{D}$ is equicontinuous. Thus, $\mathcal{D}$ is compact by the Arzela-Ascoli theorem.

Claim 2. $(\lambda^{-1}\mathcal{L})\mathcal{D} \subset \mathcal{D}$. The preservation of ($\mathcal{D}1$) and ($\mathcal{D}2$) by the operator $\lambda^{-1}\mathcal{L}$ follows from Remark 2.17. The preservation of ($\mathcal{D}3$) is verified by direct calculation: for $\underline{x} \overset{n}{\sim} \underline{x}'$ we have

$$\begin{aligned}
\mathcal{L}f(\underline{x}) &= \sum_{\underline{y} \in \sigma^{-1}\underline{x}} e^{\varphi(\underline{y})} f(\underline{y}) \\
&\leq \sum_{\underline{y}' \in \sigma^{-1}\underline{x}'} e^{\varphi(\underline{y}') + b\theta^{n+1}} B_{n+1} f(\underline{y}') \\
&= e^{b\theta^{n+1}} B_{n+1} \mathcal{L}f(\underline{x}') && (2.41) \\
&\leq B_n \mathcal{L}f(\underline{x}') && (2.42)
\end{aligned}$$

(Note that the sequences $\underline{y} \in \sigma^{-1}\underline{x}$ start with symbols $\{i\}$ such that $A_{ix_0} = 1$, and the same is true for $\underline{y}' \in \sigma^{-1}\underline{x}'$ since $x_0 = x_0'$. Therefore, there is a one-to-one correspondence between points $\underline{y} \in \sigma^{-1}\underline{x}$ and $\underline{y}' \in \sigma^{-1}\underline{x}'$, and $\underline{y} \overset{n+1}{\sim} \underline{y}'$ for the corresponding $\underline{y}$, $\underline{y}'$).

Now we have a continuous map, $\lambda^{-1}\mathcal{L}$ By the Schauder-Tychonoff theorem, a function $h \in \mathcal{D}$ exists such that $\mathcal{L}h = \lambda h$.

To complete the proof of Theorem 2.24, we need to derive (2.38), which will be done by a series of claims.

Claim 3. For any $f \in \mathcal{D}$ we have $\min_{\underline{x}} \lambda^{-k}\mathcal{L}^k f(\underline{x}) \geq K^{-1}$. To prove this, find an $\underline{x}'$ such that $f(\underline{x}') \geq 1$ (one exists due to the property (ii)). Then for any point $\underline{x} \overset{0}{\sim} \underline{x}'$ we have $f(\underline{x}) \geq B_0^{-1}$, hence $f \geq B_0^{-1}\chi_{C^+(x_0)}$. Now the claim follows from (2.36).

Next, we want to show that $\lambda^{-n}\mathcal{L}^n f$ converges to h for any $f \in \mathcal{D}$. The fact that $\lambda^{-k}\mathcal{L}^k f$ is uniformly bounded below suggests the following trick. We want to find an $\eta > 0$ such that the function $g := \lambda^{-k}\mathcal{L}^k f - \eta h$ will belong in $\mathcal{D}$, up to a normalizing factor. Note that g is positive for any $\eta||h|| \leq K^{-1}$. Note that $\nu(g) = 1 - \eta$, so that $g/(1-\eta)$ satisfies ($\mathcal{D}2$).

Claim 4. There is an $\eta > 0$ such that for any $f \in \mathcal{D}$ we have $g/(1-\eta) \in \mathcal{D}$ for $g = \lambda^{-k}\mathcal{L}^k f - \eta h$. We need to verify ($\mathcal{D}3$), i.e. to show that $g(\underline{x}) \leq B_n g(\underline{x}')$ for all $\underline{x} \overset{n}{\sim} \underline{x}'$, or equivalently,

$$\eta(B_n h(\underline{x}') - h(\underline{x})) \leq B_n \lambda^{-k}\mathcal{L}^k f(\underline{x}') - \lambda^{-k}\mathcal{L}^k f(\underline{x}) \tag{2.43}$$

Applying (2.42) to the function $\lambda^{-k+1}\mathcal{L}^{k-1}f$ gives

$$\lambda^{-k}\mathcal{L}^k f(\underline{x}) \leq e^{b\theta^{n+1}} B_{n+1}\lambda^{-k}\mathcal{L}^k f(\underline{x}')$$

Note that, $h(\underline{x}) \geq B_n^{-1}h(\underline{x}')$, because $h \in \mathcal{D}$. To prove (2.43) it is therefore enough to have

$$\eta(B_n - B_n^{-1})h(\underline{x}') \leq (B_n - e^{b\theta^{n+1}}B_{n+1})\lambda^{-k}\mathcal{L}^k f(\underline{x}')$$

or

$$\eta(B_n - B_n^{-1})||h|| \leq (B_n - e^{b\theta^{n+1}}B_{n+1})K^{-1}$$

A direct inspection now shows that an $\eta > 0$ exists such that the above inequality holds uniformly in n. This proves the claim.

Claim 5. There are constants $c > 0$ and $\beta \in (0, 1)$ such that

$$||\lambda^{-n}\mathcal{L}^n f - h|| \leq c\beta^n$$

for all $f \in \mathcal{D}$, $n \geq 0$.

This is a clear implication of the previous claim. Every k interations of the operator $\lambda^{-1}\mathcal{L}$ allow us to subtract ηh from the function we have at that time. Then ηh will stay invariant under $\lambda^{-1}\mathcal{L}$, while the norm of the difference is reduced by the factor $1-\eta$. As a result, after kN iterations, $N \geq 1$,

$$\lambda^{-kN}\mathcal{L}^{kN}f = (1-(1-\eta)^N)h + (1-\eta)^N f'$$

for some $f' \in \mathcal{D}$. Then the claim follows for all $n = kN$. We leave it to the reader to verify the claim for $n \neq kN$.

Claim 6. Let $f \in \mathcal{FR}_r(\Sigma_A^+)$ for some $r \geq 0$, and $g \in \mathcal{D}$. Suppose that $f \geq 0$ and fg is not identically zero. Then $g_r := [\nu(fg)]^{-1}\lambda^{-r}\mathcal{L}^r(fg) \in \mathcal{D}$.

Note that fg is positive on some cylinder, hence $\nu(fg) > 0$ by Remark 2.16. The properties $(\mathcal{D}1)$ and $(\mathcal{D}2)$ for g_r follow from Remark 2.17. The proof of $(\mathcal{D}3)$ for g_r goes by a direct calculation that is based on Remark 2.13 and generalizes the proof of Claim 2. We leave it to the reader.

Claim 7. Let $f \in \mathcal{FR}_r(\Sigma_A^+)$ and $g \in \mathcal{D}$. Then for $n \geq 0$

$$||\lambda^{-n-r}\mathcal{L}^{n+r}(fg) - \nu(fg)h|| \leq c\nu(|fg|)\beta^n$$

To prove this, let $f = f^+ - f^-$, where $f^+, f^- \geq 0$ and $f^+, f^- \in \mathcal{FR}_r(\Sigma_A^+)$. Combining Claims 6 and 5, we prove the above bound for f^+ and f^- separately (note that in the case $f^\pm g \equiv 0$ the bound is trivial), then we add up the two bounds and complete the proof of the claim.

The last claim implies (2.38) by the standard approximation techniques. Theorem 2.24 is now proved. □

Let $\varphi = \psi + K - u + u \circ \sigma$ for some $\varphi, \psi, u \in \mathcal{H}(\Sigma_A^+)$ and a constant K, just as in (2.31). Then

$$\mathcal{L}_\psi(e^{-u}f) = e^{-K}e^{-u}\mathcal{L}_\varphi(f)$$

for all $f, g \in \mathcal{C}(\Sigma_A^+)$. In particular, choosing $K = \ln\lambda$ and $u = \ln h$ and setting $f = h$ gives $\mathcal{L}_\psi \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the function identically equal to one. Hence, for any potential φ there is a cohomologous potential $\psi \sim \varphi$ such that $\mathcal{L}_\psi$ has the leading eigenvalue $\lambda = 1$ with a constant eigenfunction. We say that $\mathcal{L}_\psi$ (or ψ) is *normalized.* Note that $\mathcal{L}_\psi$ is a partial inverse to the operator $\sigma^* : f \mapsto f \circ \sigma$, i.e. $\mathcal{L}_\psi \circ \sigma^* =$identity.

Remark 2.18. Let φ be a normalized potential for a Gibbs measure ρ. One can use (2.34) to prove that $\varphi \leq 0$ and $S_k\varphi < 0$.

The use of normalized potentials can simplify the proofs of many properties of Gibbs measures, this observation is used in [55].

Theorem 2.25 *With notation and assumptions of the previous theorem, the probability measure $\rho = h\nu$ on Σ_A^+ is σ-invariant. Its values on cylinders $C^+(\omega_\Lambda) \subset \Sigma_A^+$, $\Lambda = [0, n-1]$, satisfy*

$$0 < c_1 \leq \frac{\rho(C^+(\omega_\Lambda))}{\exp[-Pn + S_n\varphi(\underline{x})]} \leq c_2 < \infty \tag{2.44}$$

with any $\underline{x} \in C^+(\omega_\Lambda)$. Here $P = \ln\lambda$ and c_1, c_2 are independent of $\underline{x}$, $C(\omega_\Lambda)$ and n.

Proof. The σ-invariance of ρ follows from (2.37) and Remark 2.17. To prove (2.44), put $C = C^+(\omega_\Lambda)$ and write

$$\rho(C) = \rho(\chi_C) = \nu(h\chi_C) = \lambda^{-n}\mathcal{L}^{*n}\nu(h\chi_C) = \lambda^{-n}\nu(\mathcal{L}^n(h\chi_C))$$

Applying Remark 2.13 gives

$$\mathcal{L}^n(h\chi_C)(\underline{x}) = \sum_{\underline{y}\in\sigma^{-n}\underline{x}} e^{S_n\varphi(\underline{y})}h(\underline{y})\chi_C(\underline{y})$$

Note that for any $\underline{x} \in \Sigma_A^+$ there is at most one $\underline{y} \in \sigma^{-n}\underline{x}$ such that $\underline{y} \in C = C(\omega_\Lambda)$. This observation and Remark 2.7 give an upper bound,

$$\rho(C) \leq \text{const} \cdot \lambda^{-n}e^{S_n\varphi(\underline{x}')}||h|| = \text{const} \cdot \exp[-Pn + S_n\varphi(\underline{x}')]$$

again $\underline{x}' \in C$ is an arbitrary point and $P = \ln\lambda$.

To get a lower bound on $\rho(C)$, we write

$$\rho(C) = \nu(h\chi_C) = \lambda^{-n-k}\mathcal{L}^{*n+k}\nu(h\chi_C) = \lambda^{-n-k}\nu(\mathcal{L}^{n+k}(h\chi_C))$$

with

$$\mathcal{L}^{n+k}(h\chi_C)(\underline{x}) = \sum_{\underline{y}\in\sigma^{-n-k}\underline{x}} e^{S_{n+k}\varphi(\underline{y})}h(\underline{y})\chi_C(\underline{y})$$

Now, for any $\underline{x} \in \Sigma_A^+$ there is *at least* one $\underline{y} \in \sigma^{-n-k}\underline{x}$ such that $\underline{y} \in C$. This observation and Remark 2.7 give a lower bound,

$$\rho(C) \geq \text{const} \cdot \lambda^{-n-k}e^{S_{n+k}\varphi(\underline{x}')}||h|| = \text{const} \cdot \exp[-Pn + S_n\varphi(\underline{x}')]$$

where $\underline{x}' \in C$ is an arbitrary point and $P = \ln\lambda$. Here we used two facts: λ^{-k} =const (independent of C, n, ρ) and

$$\exp[S_{n+k}\varphi(\underline{x}) - S_n\varphi(\underline{x})] \leq e^{k||\varphi||} = \text{const}$$

Thus we proved (2.44) and Theorem 2.25. □

The measure ρ is not a Gibbs measure yet because it is defined on the "wrong" space, Σ_A^+. It can be naturally extended to a σ-invariant measure on Σ_A by setting $\rho(C(\omega_{[q,r]})) = \rho(C^+(\omega_{[q,r]}))$ for all $0 \le q \le r$, and then $\rho(C(\omega_{[q,r]})) = \rho(\sigma^q C(\omega_{[q,r]}))$ for all $q < 0$. A direct inspection shows that ρ is then a σ-invariant probability measure on Σ_A.

Corollary 2.26 *The measure ρ on Σ_A is a Gibbs measure with potential φ.*

Remark 2.19. The value λ in Theorem 2.24 is unique, for a given potential φ, since $\lambda = e^{P_\varphi}$. The uniqueness of the function h follows from (2.38). Now the measure $\nu = h^{-1}\rho$ is unique, since so is the Gibbs measure ρ.

2.3 Properties of Gibbs measures

We have already proved some very basic properties of Gibbs measures, e.g., ergodicity. We now turn to more advanced properties that are particularly important in many applications.

Theorem 2.27 (Exponential cluster property) *Let (Σ_A, σ) be mixing and ρ be the Gibbs measure for a potential $\varphi \in \mathcal{H}(\Sigma_A)$. There are constants $c > 0, \beta \in (0,1)$ such that for any two cylinders $C = C(\omega_{[0,s]})$ and $D = C(\omega_{[0,r]})$ we have*

$$|\rho(C \cap \sigma^{-n} D) - \rho(C)\rho(D)| \le c\rho(C)\rho(D)\beta^{n-s} \tag{2.45}$$

Note that $n - s$ is the gap between the intervals on which the cylinders C and $\sigma^{-n}D$ are based.

Proof. It is enough to work in Σ_A^+ and prove the theorem for $C = C^+(\omega_{[0,s]})$ and $D = C^+(\omega_{[0,r]})$. First, we have

$$\begin{aligned}
\rho(C \cap \sigma^{-n} D) &= \rho(\chi_C \cdot (\chi_D \circ \sigma^n)) \\
&= \nu(h\chi_C \cdot (\chi_D \circ \sigma^n)) \\
&= \lambda^{-n} \mathcal{L}^{*n} \nu(h\chi_C \cdot (\chi_D \circ \sigma^n)) \\
&= \nu(\lambda^{-n} \mathcal{L}^n (h\chi_C \cdot (\chi_D \circ \sigma^n))) \\
&= \nu(\lambda^{-n} \mathcal{L}^n (h\chi_C) \cdot \chi_D)
\end{aligned}$$

At the last step we used Remark 2.14. Now

$$\begin{aligned}|\rho(C\cap\sigma^{-n}D)-\rho(C)\rho(D)| &= |\rho(C\cap\sigma^{-n}D)-\nu(h\chi_C)\nu(h\chi_D)| \\ &= |\nu((\lambda^{-n}\mathcal{L}^n(h\chi_C)-\nu(h\chi_C)h)\chi_D)| \\ &\leq ||\lambda^{-n}\mathcal{L}^n(h\chi_C)-\nu(h\chi_C)h||\cdot\nu(\chi_D)\end{aligned}$$

We now apply the crucial Claim 7 in the proof of Theorem 2.24 to the function $\chi_C \in \mathcal{FR}_s(\Sigma_A^+)$ and get

$$||\lambda^{-n}\mathcal{L}^n(h\chi_C)-\nu(h\chi_C)h|| \leq c\rho(C)\beta^{n-s}$$

Note that $\nu(\chi_D) \leq (\min h)^{-1}\rho(D)$. Therefore,

$$|\rho(C\cap\sigma^{-n}D)-\rho(C)\rho(D)| \leq c'\rho(C)\rho(D)\beta^{n-s}$$

with $c' = cK$, because $\min h \geq K^{-1}$. □

Corollary 2.28 *Every Gibbs measure ρ is mixing.*

A dynamical system is said to be Bernoulli if it is isomorphic to a Bernoulli shift. The Bernoulli property is the highest one in the hierarchy of ergodic properties, it implies mixing of any order and K-mixing.

Theorem 2.29 (Bernoulli property) *Let (Σ_A, σ) be mixing. Then for any potential $\varphi \in \mathcal{H}(\Sigma_A)$ the Gibbs measure ρ_φ is Bernoulli.*

Proof. Two finite or countable partitions ξ, η of Σ_A are said to be ε-independent if

$$\sum_{A\in\xi, B\in\eta} |\rho(A\cap B)-\rho(A)\rho(B)| \leq \varepsilon$$

A partition ξ is said to be *weak Bernoulli* (for σ and ρ) if for any $\varepsilon > 0$ there is an $N \geq 1$ such that the partitions $\xi \vee \sigma^{-1}\xi \vee \cdots \vee \sigma^{-s}\xi$ and $\sigma^{-n}\xi \vee \sigma^{-n-1}\xi \vee \cdots \vee \sigma^{-n-r}\xi$ are ε-independent for all $s, r \geq 0$ and $n \geq s+N$. A theorem by Friedman and Ornstein [30] asserts that if there exists a weak Bernoulli, generating partition ξ, then the dynamical system is Bernoulli, cf. also [78].

In our case, the partition ξ into cylinders $C(i) = \{\underline{\omega} \in \Sigma_A : \omega_0 = i\}$ is a generating one. Theorem 2.27 immediately implies that the partitions

$\xi \vee \sigma^{-1}\xi \vee \cdots \vee \sigma^{-s}\xi$ and $\sigma^{-n}\xi \vee \sigma^{-n-1}\xi \vee \cdots \vee \sigma^{-n-r}\xi$ are $(c\beta^{n-s})$-independent, hence ξ is weak Bernoulli. $\square$

Statistical properties. Let $T : M \to M$ be a dynamical system preserving a probability measure ν and $f : M \to \mathbb{R}$ a measurable function. Then the sequence $\xi_n = f \circ T^n$ is a stationary stochastic process defined on the probability space (M, ν) (its stationarity follows from the invariance of the measure ν). Many basic results of probability theory can be carried over to stochastic processes generated by dynamical systems. The two most important ones are the asymptotics of the *correlation function*

$$C_f(n) = \nu(f \cdot (f \circ T^n)) - [\nu(f)]^2 \tag{2.46}$$

and the *central limit theorem*

$$\lim_{n\to\infty} \nu \left\{ \frac{S_n f(x) - n\nu(f)}{\sqrt{n}} < z \right\} = \frac{1}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^{z} e^{-\frac{x^2}{2\sigma_f^2}}\, dx \tag{2.47}$$

for all $-\infty < z < \infty$. Here $S_n f(x) = f(x) + \cdots + f(T^{n-1}x)$ and $\sigma_f \geq 0$ (in the case $\sigma_f = 0$ the integral on the right equals 0 for $z < 0$ and 1 for $z > 0$). The property (2.47) is equivalent, in the language of probability theory, to the convergence of $n^{-1/2}(S_n f(x) - n\nu(f))$ in distribution to the normal random variable $N(0, \sigma_f^2)$. Note that this property is a refinement of the Birkhoff ergodic theorem. It tells us that the typical deviations of the ergodic sum $S_n f$ from its average value, $n\nu(f)$, are of order $\sqrt{n}$, and their distribution is asymptotically Gaussian. The variance σ_f^2 is normally given by

$$\sigma_f^2 = C_f(0) + 2\sum_{n=1}^{\infty} C_f(n) \tag{2.48}$$

Furthermore, in typical cases

$$\sigma_f^2 = 0 \quad \Longleftrightarrow \quad f(x) = g(x) - g(Tx) + K \tag{2.49}$$

for some $g(x) \in L_2(M)$ and a constant K (such functions f are also called coboundary, compare to (2.31)). It is also common to study the asymptotics of more general correlation functions than (2.46):

$$C_{f,g}(n) = \nu(f \cdot (g \circ T^n)) - \nu(f)\nu(g) \tag{2.50}$$

where $g : M \to \mathbb{R}$ is another measurable function. We know that $C_{f,g}(n) \to 0$ as $n \to \infty$ for all $f, g \in L_2(M)$ if and only if the dynamical system (M, T, ν) is mixing, cf. 3.6.h in [44]. No degree of ergodicity (not even the Bernoulli property) can enforce any speed of convergence of $C_{f,g}(n)$ to zero for generic L_2-function f, g. Even for most chaotic dynamical systems (such as expanding interval maps or hyperbolic toral automorphisms) and generic continuous (!) functions f, g the convergence of $C_{f,g}(n)$ to zero appears to be arbitrarily slow, and, furthermore, the central limit theorem (2.47) fails. To obtain affirmative results, one has to assume certain degree of smoothness for the functions f, g, such as Hölder continuity.

Theorem 2.30 (Exponential decay of correlations) *Let (Σ_A, σ) be mixing and ρ be the Gibbs measure for a potential $\varphi \in \mathcal{H}_\theta(\Sigma_A)$. For any $f, g \in \mathcal{H}_{\theta'}(\Sigma_A)$ we have*

$$|\rho(f \cdot (g \circ \sigma^n)) - \rho(f)\rho(g)| \le c_{f,g} \cdot \gamma^n$$

with some $c_{f,g} > 0$ and $\gamma \in (0, 1)$ that depends only on φ and θ'.

Proof. For $j \ge 1$ and any $f \in \mathcal{H}_{\theta'}(\Sigma_A)$, denote by f_j the conditional expectation of f on the partition of Σ_A into cylinders $C(\omega_\Lambda)$, $\Lambda = [-j, j]$. This means that $f_j \equiv \rho(f\chi_C)/\rho(\chi_C)$ on every cylinder C in this partition. Then $f_j \in \mathcal{FR}_j(\Sigma_A)$, $\rho(f_j) = \rho(f)$ and $||f - f_j|| \le b_f \theta'^j$. Applying the triangle inequality gives

$$|\rho(f \cdot (g \circ \sigma^n)) - \rho(f)\rho(g)| \le |\rho(f_j \cdot (g_j \circ \sigma^n)) - \rho(f_j)\rho(g_j)| + 2b_f b_g \theta'^j$$

Now f_j and g_j are constant on cylinders $C(\omega_\Lambda)$, $\Lambda = [-j, j]$. A direct application of Theorem 2.27 then gives

$$|\rho(f_j \cdot (g_j \circ \sigma^n)) - \rho(f_j)\rho(g_j)| \le c||f|| \, ||g||\beta^{n-2j}$$

Letting $j = [n/3]$ we get the result with $\gamma = \max\{(\theta')^{1/3}, \beta^{1/3}\}$. □

Theorem 2.31 (Central limit theorem) *Let (Σ_A, σ) be mixing and ρ be the Gibbs measure for a potential $\varphi \in \mathcal{H}(\Sigma_A)$. For any $f \in \mathcal{H}(\Sigma_A)$ the central limit theorem holds in the sense of (2.47)-(2.49).*

There are several ways to prove this theorem, but they involve advanced methods of probability theory, such as martingales or characteristic functions and the continuity theorem. These are all beyond the scope of this survey, and we refer the reader to [64].

Remark 2.20. M. Denker extended [22] the central limit theorem from Hölder continuous functions to all functions $f \in L_2(\Sigma_A)$ such that

$$\sum_{j\geq 0} \left(\rho(f-f_j)^2\right)^{1/2} < \infty$$

where f_j is, as in the proof of Theorem 2.30, the conditional expectation of f on the partition into cylinders $C(\omega_{[-j,j]}) \subset \Sigma_A$. This class of functions is much larger than $\mathcal{H}(\Sigma_A)$. From Denker's arguments, it seems rather unlikely that this class could be enlarged much further.

Many more advanced statistical properties have been proven for Gibbs measures: the local central limit theorem, the functional central limit theorem (the convergence to the Wiener process), the law of iterated logarithms, the renewal theorems, etc. These are all beyond the scope of this survey, and we refer the reader to [36, 59] for exact statements and proofs.

While all these statistical properties are similar to the classical limit theorems in probability theory, the following one is unusual, it is specific for Gibbs measures.

Theorem 2.32 (Large deviations) *Let (Σ_A, σ) be mixing and ρ be the Gibbs measure for a potential $\varphi \in \mathcal{H}(\Sigma_A)$. For any $f \in \mathcal{H}(\Sigma_A)$ there is a real analytic strictly concave function $\eta(p)$ on an open interval (p_1^*, p_2^*) such that for every interval $I \subset \mathbb{R}$*

$$\lim_{n\to\infty} \frac{1}{n} \ln \rho \left\{ \frac{1}{n} S_n f(x) \in I \right\} = \sup_{p \in I \cap (p_1^*, p_2^*)} \eta(p)$$

provided $I \cap (p_1^*, p_2^*) \neq \emptyset$.

It is easy to see that the function $\eta(p)$ takes its maximum value $\eta(p) = 0$ at $p = \rho(f)$. In physics, the function $\eta(p)$ is called *free energy*. Note that if $f \sim$ const (f cohomologous to a constant), then $|S_n f - n\rho(f)| \leq$const, and the above theorem degenerates to $p_1^* = p_2^*$.

Both the central limit theorem 2.31 and the above theorem 2.32 describe the fluctuations of S_nf around its mean value $\rho(S_nf) = n\rho(f)$. The difference is that the CLT applies when $S_nf - n\rho(f) = O(\sqrt{n})$ and 2.32 applies to much larger deviations, when $S_nf - n\rho(f) = O(n)$.

Equilibrium states and Gibbs measures. An important property of Gibbs measures is that they are (unique) *equilibrium states* for their potential functions. Let $\varphi \in \mathcal{C}(\Sigma_A)$ and P_φ denote the topological pressure of the function φ. The variational principle, cf. 4.4.d in [44], states that

$$P_\varphi = \sup_\mu [h_\mu(\sigma) + \mu(\varphi)] \tag{2.51}$$

where the supremum is taken over all σ-invariant measures μ on Σ_A. Here $h_\mu(\sigma)$ is the measure-theoretic entropy (also called Kolmogorov-Sinai entropy) of the measure μ. Any measure μ which maximizes the right hand side of (2.51) is called an *equilibrium state* for the function φ.

Theorem 2.33 *Let (Σ_A, σ) be mixing and $\varphi \in \mathcal{H}(\Sigma_A)$. Then the Gibbs measure ρ_φ is the only equilibrium state for the function φ.*

Proof. Theorem 2.18 immediately implies that ρ is an equilibrium state. The uniqueness follows from a general theorem saying that equilibrium states are unique for expansive homeomorphisms of compact metric spaces provided they sartisfy the so called specification property, see 4.4.f in [44]. The verification of the expansiveness and specification property is an easy exercise that we leave to the reader. Note that there is also a direct proof of the uniqueness, see Bowen's book [11]. □

The topological pressure of the zero function $\varphi \equiv 0$ is the *topological entropy* of σ, i.e. $P_0 = h_{\rm top}(\sigma)$, cf. 2.5.k in [44]. Then (2.51) becomes

$$P_0 = h_{\rm top}(\sigma) = \sup_\mu h_\mu(\sigma) \tag{2.52}$$

Hence, the Gibbs measure ρ_0 for the potential $\varphi \equiv 0$ is the only measure of maximal entropy for the TMC (Σ_A, σ). We have seen in Example 2.8 that the measure ρ_0 is a Markov measure, called also the Parry measure.

Remark 2.21. Let (Σ_A, σ) be mixing and $\lambda_A > 0$ the largest eigenvalue of the matrix A. Then $h_{\rm top}(\sigma) = \ln \lambda_A$. Indeed, one can use the explicit description of the measure ρ_0 in Examples 2.7 and 2.8, and a remark about λ in Example 2.8.

The Parry measure ρ_0 is, in a sense, the most "chaotic" invariant measure on Σ_A, if one takes the entropy as the measure of chaoticity. Its other important property is that it describes the distribution of periodic points in Σ_A, as we see below.

Statistics of periodic orbits. Consider

$$\mathrm{Fix}(\sigma^n, \Sigma_A) = \{\underline{\omega} \in \Sigma_A : \sigma^n \underline{\omega} = \underline{\omega}\}$$

the set of periodic points of period n in Σ_A. Each point $\underline{\omega} \in \mathrm{Fix}(\sigma^n, \Sigma_A)$ is, of course, a periodic symbolic sequence with period n. The number of such sequences starting with $\omega_0 = i$ coincides with the number of allowable words of length $n+1$ such that $\omega_0 = \omega_{n+1} = i$. One can easily see that this number is equal to the i-th diagonal entry of the matrix A^n. Hence, the total number of periodic sequences of period n is

$$\mathrm{Per}_n(\Sigma_A) := \mathrm{Card}\Big[\mathrm{Fix}(\sigma^n, \Sigma_A)\Big] = \mathrm{tr}\, A^n$$

Then we have a simple and exact formula

$$\mathrm{Per}_n(\Sigma_A) = \sum_{i=1}^{m} \lambda_i^n \tag{2.53}$$

where $\lambda_1, \ldots, \lambda_m$ are all the eigenvalues of A. (Note that for a typical matrix A, the eigenvalues are irrational or even complex, still the above sum is a nonnegative integer for each n.) Asymptotically, of course, $\mathrm{Per}_n(\Sigma_A) = \lambda_+^n + O(|\lambda'|^n)$, where λ' is the second largest (in absolute value) eigenvalue of A.

The complete information about the sequence $\{\mathrm{Per}_n(\Sigma_A)\}$ is encoded in the so-called *dynamical zeta-function*

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{Per}_n(\Sigma_A) \tag{2.54}$$

(similarly, the dynamical zeta-function can be defined for continuous time dynamical systems, see 2.5.d in [44]). It easily follows from (2.53) that

$$\zeta(z) = \frac{1}{\det(I - zA)}$$

so that the zeta-function $\zeta(z)$ is rational on $\mathbb{C}$ and analytic in the open disc $|z| < 1/\lambda_+$.

For the proofs of the following two theorems we refer the reader to [68], see also Theorem 20.3.7 in [43].

Theorem 2.34 *Let (Σ_A, σ) be mixing and ρ_0 the (unique) measure of maximal entropy, i.e. the Parry measure. For any $B \subset \Sigma_A$, let $\mathrm{Per}_n(B)$ denote the number of periodic points of period n contained in B. If $\rho_0(\partial B) = 0$, then*

$$\lim_{n\to\infty} \frac{\mathrm{Per}_n(B)}{\mathrm{Per}_n(\Sigma_A)} = \rho_0(B) \tag{2.55}$$

(in particular, this holds if B is any cylinder, in this case $\partial B = \emptyset$).

This theorem allows us to approximate the measure ρ_0 by finite measures supported on periodic orbits. Let ν_n be a finite atomic measure that assigns equal weights, $1/\mathrm{Per}_n(\Sigma_A)$ to each periodic point $x \in \mathrm{Fix}(\sigma^n, \Sigma_A)$. Then the measure ν_n weakly converges to ρ_0, as $n \to \infty$. (This is exactly what the above proposition states!) Note that every measure ν_n has zero entropy, while $h_{\rho_0}(\sigma) = \ln \lambda_+ > 0$, so the entropy is not continuous on the space of invariant measures. It is, however, upper-semicontinuous.

It is possible to approximate *any* Gibbs measure ρ_φ on Σ_A in a similar way, by finite atomic measures on periodic orbits. One only has to assign weights properly.

Theorem 2.35 *Let (Σ_A, σ) be mixing and $\varphi \in \mathcal{H}$ a Hölder continuous function. For $n \geq 1$, let ν_n be the finite atomic measure concentrated on $\mathrm{Fix}(\sigma^n, \Sigma_A)$ that assigns weight*

$$\nu_n(\underline{\omega}) = Z^{-1} \exp\left[\varphi(\underline{\omega}) + \cdots + \varphi(\sigma^{n-1}\underline{\omega})\right] \tag{2.56}$$

to each point $\underline{\omega} \in \mathrm{Fix}(\sigma^n, \Sigma_A)$ (here Z is the normalizing factor). Then ν_n weakly converges to ρ_φ, as $n \to \infty$.

Note that Theorem 2.34 is a particular case of 2.35, since for $\varphi \equiv 0$ we have $\nu_n(\underline{\omega}) = Z^{-1} = 1/\mathrm{Per}_n(\Sigma_A)$ in (2.56).

A few additional properties. We have seen that Gibbs measures are positive on all cylinders in Σ_+. In fact, there is an exponential lower bound

$$\rho(C) \geq c_1\theta_1^n$$

for every cylinder $C = C(\omega_{[0,n]}) \subset \Sigma_A$, where $c_1 > 0$ and $\theta_1 \in (0,1)$ only depend on ρ (cf. Remark 2.16). Furthermore, if $C' = C(\omega_{[-p,n+q]}) \subset C = C(\omega_{[0,n]})$ with some $p, q \geq 0$, then

$$\rho(C') \geq c_1 \theta_1^{p+q} \rho(C)$$

There are similar exponential upper bounds:

$$\rho(C) \leq c_2 \theta_2^n$$

and

$$\rho(C') \leq c_2 \theta_2^{p+q} \rho(C)$$

in the same notation, with some $c_2 > 0$ and $\theta_2 \in (0,1)$ only depending on ρ. These bounds can be obtained with the help of normalized potentials, see Remark 2.18, we leave the proofs to the reader.

Based on the above bounds, for any two Gibbs measures ρ_1 and ρ_2, there are $c > 1$ and $a > 1$ such that for all cylinders $C \subset \Sigma_A$

$$c^{-1}[\rho_1(C)]^a \leq \rho_2(C) \leq c[\rho_1(C)]^{1/a}$$

Hence, even though distinct Gibbs measures are mutually singular, their mutual singularity is under certain control.

Another useful property of Gibbs measures is their *almost direct product structure*. Fix an $\underline{\omega} \in \Sigma_A$. Let $\Lambda = [n', n'']$ be an interval in $\mathbb{Z}$ and $C = C(\underline{\omega}|_\Lambda) \subset \Sigma_A$ a cylinder. Clearly, C has a topological product structure: $C = C^+ \times C^-$, where $C^+ = C(\underline{\omega}|_{(-\infty,n'']})$ and $C^- = C(\underline{\omega}|_{[n',\infty)})$. Now, for any Gibbs measure ρ there are two Borel probability measures $\rho^\pm$ defined on $C^\pm$, respectively, such that

$$\left| \frac{1}{\rho(C)} \cdot \frac{d\rho}{d(\rho^+ \times \rho^-)} - 1 \right| \leq c\,\theta^{n''-n'}$$

where $c > 0$ and $\theta < 1$ only depend on ρ. Note that $1/\rho(C)$ is just a normalizing factor, because $\rho(C) < 1$ and $[\rho^+ \times \rho^-](C) = 1$. This bound follows from the definition of Gibbs states through interactions, see more details in [39, 17].

Moreover, let $n' \to -\infty$, so that C will shrink to C^+. Then the measure ρ^+ will weakly converge to a limit measure on C^+. This is the conditional measure induced by ρ on C^+. Similarly, as $n'' \to \infty$, the measure ρ^- will weakly converge to the conditional measure on C^- induced by ρ.

3 Sinai-Ruelle-Bowen measures

We turn back to the smooth hyperbolic systems – Anosov diffeomorphisms $T : M \to M$ and Axiom A maps on basic sets $T : \Omega_r \to \Omega_r$ (as we remarked in Section 1, the latter includes compact locally maximal transitive hyperbolic sets of more general diffeomorphisms). We assume topological mixing for T in all cases.

Let $\mathcal{R}$ be a Markov partition of M (resp., Ω_r). It produces a topological Markov chain (Σ_A, σ). Now let $\varphi : \Sigma_A \to \mathbb{R}$ be a Hölder continuous function. It produces a Gibbs measure ρ_φ on Σ_A. The projection of ρ_φ under $\pi : \Sigma_A \to M$ is a T-invariant measure on M. First we check that this measure inherits the properties of the measure ρ_φ.

Proposition 3.1 *The projection $\pi : \Sigma_A \to M$ is ρ_φ-almost everywhere one-to-one. Hence, the ergodic and statistical properties of the measure ρ_φ and those of its image on M are the same.*

Proof. Recall that π^{-1} fails to be one-to-one on the subset $M^\# = \cup_n T^n(\partial\mathcal{R})$, where $\partial\mathcal{R} = \partial^s\mathcal{R} \cup \partial^u\mathcal{R}$. Now let $D^{u,s} = \pi^{-1}(\partial^{u,s}\mathcal{R})$. The sets D^u, D^s are closed subsets of Σ_A such that $\sigma(D^s) \subset D^s$ and $\sigma^{-1}(D^u) \subset D^u$ (this follows from the Markov property). Since the measure ρ_φ is σ-invariant, we have $\rho_\varphi(\sigma D^s) = \rho_\varphi(D^s)$, i.e. the set D^s is σ-invariant (mod 0). Since ρ_φ is ergodic, then D^s has measure 0 or 1. The latter is impossible since D^s is nowhere dense, according to Remark 2.5. Similarly, $\rho_\varphi(D^u) = 0$, hence we have $\rho_\varphi(\pi^{-1}M^\#) = 0$. □

Corollary 3.2 *Let $\Sigma_A^* = \pi^{-1}(M \setminus M^\#)$. Then $\rho_\varphi(\Sigma_A^*) = 1$ for all Gibbs measures ρ_φ.*

Remark 3.1. It is interesting that any T-invariant measure on M (resp., Ω_r) can be lifted to a (not necessarily unique) σ-invariant measure on Σ_A. The proof involves the Hahn-Banach theorem, i.e. is based on the Axiom of choice, see 4.3 in [11].

Now we have a class of T-invariant measures associated with Hölder continuous functions on Σ_A. It would be more natural to associate measures with functions on M (or just Ω_r) rather than with functions on Σ_A.

Let $\psi : M \to \mathbb{R}$ be a Hölder continuous function. Then for any Markov partition $\mathcal{R}$ the function $\varphi = \psi \circ \pi$ is a Hölder continuous function on Σ_A (its Hölder continuity follows by direct calculation based on the hyperbolicity of T, it is left as an exercise). Now one can associate the measure $\mu_\psi := \pi_* \rho_\varphi$ with the 'potential' function ψ and the Markov partition $\mathcal{R}$.

Proposition 3.3 *The measure μ_ψ does not depend on the choice of the Markov partition $\mathcal{R}$. Moreover, μ_ψ is the unique equilibrium state for the function ψ, i.e. the unique measure on which the following supremum is attained*

$$P_\psi = \sup_\mu \left(h_\mu(T) + \mu(\psi) \right) \tag{3.1}$$

where the supremum is taken over all T-invariant measures on Ω_r, $h_\mu(T)$ is the measure-theoretic entropy of the system (Ω_r, T, μ) and P_ψ is the topological pressure of the function ψ with respect to the map $T : \Omega_r \to \Omega_r$.

The first claim follows from the second. The second basically follows from Theorem 2.33, with some little extra work. We refer the reader to [11], Sect. 4A.

Definition. The measure μ_ψ is called the Gibbs measure corresponding to the potential function ψ on M (resp. Ω_r).

The following is a direct analogue of Theorem 2.23 and can be proved similarly:

Proposition 3.4 (see [11]) *Let T be an Anosov diffeomorphism on M or an Axiom A diffeomorphism restricted to a basic set Ω_r. Let φ, ψ be Hölder continuous on M, resp. Ω_r. The following are equivalent:*
(i) $\mu_\varphi = \mu_\psi$.
(ii) *There is a Hölder continuous function u on M, resp. Ω_r, and a constant K such that $\varphi - \psi = K + u \circ T - u$.*
(iii) *There is a constant K such that $S_n\varphi(x) - S_n\psi(x) = nK$ whenever $T^n x = x$.*
If these conditions hold, then $K = P_\varphi - P_\psi$.

One can now see that there are uncountably many Gibbs measures for any Anosov and Axiom A diffeomorphism. There is one special measure, however, associated with the the Riemannian volume on the manifold M, on which we concentrate next.

The function φ^u. Let J^u_x be the Jacobian of the linear map

$$DT : E^u_x \to E^u_{Tx}$$

Since the map $x \mapsto E^u_x$ is Hölder continuous (see 6.4.a in [44]) and the map $E^u_x \to J^u_x$ is differentiable, then the function J^u_x is Hölder continuous. We denote by

$$\mu^+ = \mu_{\varphi^u}$$

the Gibbs measure for the potential $\varphi^u = -\ln J^u_x$. It is called a *u-Gibbs measure*, now it is often referred to as *generalized Sinai-Ruelle-Bowen measure*. Below we study its properties.

We note that the measure μ^+ and the value of P_{φ^u} are independent of the choice of an equivalent Riemanian metric on M. Indeed, for every periodic point $x = T^n x$ the value of

$$S_n\varphi^u(x) = -\ln \mathrm{Jac}(DT^n : E^u_x \to E^u_x)$$

does not depend on the metric, and then we can use Proposition 3.4. (Here Jac is the absolute value of the Jacobian determinant.)

The potential φ^u and the measure μ^+ are intimately connected with the Riemannian volume Vol $(\cdot)$ on the manifold M. In particular, they can be used to estimate the volume of the so-called (ε, n)-balls. For $x \in \Omega_r$ let

$$B_x(\varepsilon, n) = \{y \in M : d(T^i x, T^i y) \le \varepsilon \;\; \forall i \in [0, n)\}$$

where $d(\cdot, \cdot)$ is the Riemannian distance on M. Also, let $B(\varepsilon, n) = \cup_x B_x(\varepsilon, n)$, where the union is taken over $x \in M$ (resp., $x \in \Omega_r$). We state a few technical lemmas without proofs.

Lemma 3.5 (First volume lemma) *For any small $\varepsilon > 0$, there is a $C_\varepsilon > 1$ such that for all $x \in M$*

$$C_\varepsilon^{-1} \exp[S_n\varphi^u(x)] \le \mathrm{Vol}\, B_x(\varepsilon, n) \le C_\varepsilon \exp[S_n\varphi^u(x)]$$

We will denote this by

$$\mathrm{Vol}\, B_x(\varepsilon, n) \asymp_\varepsilon \exp[S_n\varphi^u(x)]$$

Lemma 3.6 (Second volume lemma) *For any small $\varepsilon, \delta > 0$ there is a $C = C(\varepsilon, \delta) > 0$ such that for all $x \in \Omega_r$ and $y \in B_x(\varepsilon, n)$*

$$\mathrm{Vol}\, B_y(\delta, n) \geq C \cdot \mathrm{Vol}\, B_x(\varepsilon, n)$$

Lemma 3.7 (Pressure) *For small $\varepsilon > 0$, we have on any basic set Ω_r*

$$P_{\varphi^u} = \lim_{n\to\infty} \frac{1}{n} \ln \,\mathrm{Vol}\, B(\varepsilon, n) \leq 0$$

In particular, for an Anosov diffeomorphism, $B(\varepsilon, n) = M$ and $P_{\varphi^u} = 0$.

Corollary 3.8 *Let T be an Anosov diffeomorphism, $\mathcal{R}$ a Markov partition. For $x \in M$ and $i \geq 0$, denote by $R_{x,i} \in \mathcal{R}$ the rectangle containing $T^i x$. Then*

$$\mathrm{Vol}\, B_x(\varepsilon, n) \asymp_\varepsilon \mu^+(R_{x,0} \cap T^{-1} R_{x,1} \cap \cdots \cap T^{-n+1} R_{x,n-1})$$

This shows certain 'equivalence' of the volume on M and the measure μ^+. In the above formula, $\asymp_\varepsilon$ means that the ratio of the two quantities is bounded by some $C_\varepsilon < \infty$.

The above lemmas can be proved by somewhat involved calculations, we refer the reader to [11]. Note that, by 3.7, P_{φ^u} is the exponential rate of decrease of the volume of $B(\varepsilon, n)$, i.e. the exponential rate of 'escape' of the mass from the vicinity of the basic set Ω_r.

Attractors. The measure μ^+ has especially remarkable properties for Anosov diffeomorphisms and Axiom A attractors. By definition, an Axiom A basic set Ω_r is an *attractor* if there is an open set $U \supset X_r$ such that $\cap_{n\geq 0} T^n U = X_r$. Equivalently, for any $\varepsilon > 0$ there is an open neighborhood U of Ω_r such that $U \subset B_\varepsilon(X_r)$ and $TU \subset U$. Here we denote by $B_\varepsilon(x)$ an open ball of radius $\varepsilon > 0$ centered at $x \in M$ and set $B_\varepsilon(X) = \cup_{x\in X} B_\varepsilon(x)$ for $X \subset M$.

Before we study Axiom A attractors, we prove some related properties. We set $W^{u,s}_\varepsilon(x) = W^{u,s}(x) \cap B_\varepsilon(x)$. We agree to use the so-called Lyapunov (or adapted) metric in M, see [11] and 6.4.a in [44]. In particular, $TW^s_\varepsilon(x) \subset W^s_\varepsilon(Tx)$, etc.

Lemma 3.9 *Let Ω_r be a basic set. There is an $\varepsilon > 0$ such that $\cap_{n=-\infty}^{\infty} T^n B_\varepsilon(\Omega_r) = \Omega_r$.*

Proof. Let $y \in \cap_{n=-\infty}^{\infty} T^n B_\varepsilon(\Omega_r)$. For $n \in \mathbb{Z}$, pick an $x_n \in \Omega_s$ so that $d(x_n, T^n y) < \varepsilon$. Recall that the shadowing lemma says that every α-pseudo-orbit is β-shadowed by an orbit, cf. 6.6.c in [44]. If ε is small, then $\{x_n\}$ is an α-pseudo-orbit that is β-shadowed by an orbit $\{T^n x\}$ in Ω_s. Now, $d(T^n y, T^n x) < \varepsilon + \beta$ for all $n \in \mathbb{Z}$, hence $y = x$. □

Similarly, there is an $\varepsilon > 0$ such that $\cap_{n=-\infty}^{\infty} T^n B_\varepsilon(\Omega) = \Omega$ for the entire set of nonwandering points. The open set $B_\varepsilon(\Omega)$ is called a *fundamental neighborhood* of Ω.

Let

$$W^s(\Omega_r) = \{y \in M : d(T^n y, \Omega_r) \to 0 \quad \text{as} \quad n \to \infty\}$$

$$W^u(\Omega_r) = \{y \in M : d(T^{-n} y, \Omega_r) \to 0 \quad \text{as} \quad n \to \infty\}$$

Using the definition of nonwandering set it is easy to check that $T^n x \to \Omega$ and $T^{-n} x \to \Omega$ as $n \to \infty$ for every $x \in M$. As $\Omega = \cup \Omega_r$ is a disjoint union of basic sets, then

$$M = \cup_r W^s(\Omega_r) = \cup_r W^u(\Omega_r)$$

and these are disjoint unions.

Lemma 3.10 *We have*

$$W^{u,s}(\Omega_r) = \cup_{x \in \Omega_r} W^{u,s}(x)$$

For $\varepsilon > 0$, there is a neighborhood U_r of Ω_r such that

$$\cap_{n \geq 0} T^{-n} U_r \subset W^s_\varepsilon(\Omega_r) := \cup_{x \in \Omega_r} W^s_\varepsilon(x)$$

$$\cap_{n \geq 0} T^{n} U_r \subset W^u_\varepsilon(\Omega_r) := \cup_{x \in \Omega_r} W^u_\varepsilon(x)$$

Proof. Just like in the previous proof, let $T^n y \to \Omega_r$, i.e. $d(T^n y, \Omega_r) < \varepsilon$ for all $n \geq N$. Pick $x_n \in \Omega_r$ so that $d(x_n, T^n y) < \varepsilon$ for $n \geq N$. For $n < N$, let $x_n = T^{n-N} x_N$. Then $\{x_n\}$ is a pseudo-orbit shadowed by an orbit $\{T^n x\}$, and then $y \in W^s(x)$. □

Lemma 3.11 *A basic set Ω_r is an attractor if and only if $\Omega_r = \cup_{x \in \Omega_r} W^u(x)$, i.e. Ω_r is a union of (global) unstable manifolds of its points.*

Proof. If Ω_r is an attractor, then for every $x \in \Omega_r$ the subset $W^u(x) \cap \Omega_r$ is both open and closed in the interior topology of $W^u(x)$ induced by the Riemannian metric on $W^u(x)$. The closedness follows from that of Ω_r. The openness follows from the definition of attractor and Lemma 3.9. Hence, $W^u(x) \subset \Omega_r$.

Conversely, if $W^u(x) \subset \Omega_r$ for all $x \in \Omega_r$, then obviously the set $U = \cup_{y \in \Omega_r} W^s_\varepsilon(y)$ is an open neighborhood of Ω_r, and $TU \subset U$. □

Lemma 3.12 (i) *If $W^u_\varepsilon(x) \subset \Omega_r$ for some $x \in \Omega_r$ and $\varepsilon > 0$, then Ω_r is an attractor.*
(ii) *If Ω_r is not an attractor, then there is a $\delta > 0$ such that for every $x \in \Omega_r$ there is a $y \in W^u_\varepsilon(x)$ with $d(y, \Omega_r) > \delta$ (i.e., every unstable manifold of size ε 'sticks out' of Ω_r by at least δ).*

Proof. Let $W^u_\varepsilon(x) \subset \Omega_r$. Obviously, $W^u_\varepsilon(T^n x) \subset \Omega_r$ for all $n \geq 0$. It is enough to prove that $W^u_{\varepsilon/4}(y) \subset \Omega_r$ for all $y \in \Omega_r$. For each $n \geq 0$, due to the local hyperbolic product structure in Ω_r, there is a small ball $B_\delta(T^n x)$ with $\delta \ll \varepsilon$ about $T^n x$ such that for all $y \in B_\delta(T^n x) \cap \Omega_r$ we have $W^u_{\varepsilon/2}(y) \subset \Omega_r$. If the semiorbit $\{T^n x\}_0^\infty$ is dense in Ω_r, we are done. If not, there is a point $y \in B(x)\Omega_r$ with a dense future semiorbit, and we apply the same argument to it.

To prove (ii), assume that for every $\delta = n^{-1}$ there is an $x_n \in \Omega_r$ such that $W^u_\varepsilon(x_n) \cap \Omega_r$ is n^{-1}-dense in $W^u_\varepsilon(x_n)$. Take a limit point $x_{n_k} \to x \in \Omega_r$ and get $W^u_\varepsilon(x) \subset \Omega_r$, then apply (i). □

The following theorem describes attractors in terms of the measure μ^+:

Theorem 3.13 *Let Ω_r be a basic set. The following are equivalent:*
(a) *Ω_r is an attractor*
(b) $\mathrm{Vol}\,(W^s(\Omega_r)) > 0$
(c) *$P_{\varphi^u} = 0$ and $h_{\mu^+}(T) = -\mu^+(\varphi^u)$, with respect to $T : \Omega_r \to \Omega_r$.*

Proof. The implication (a)⇒(b) follows from Lemma 3.11. The implication (b)⇒(c) follows from Lemma 3.7 and (3.1). The proof of the implication (c)⇒(a) is based on the following observation: if Ω_r is not an attractor, then the mass escapes from any vicinity of Ω_r exponentially fast (due to part (ii) of Lemma 3.12), and so $P_{\varphi^u} < 0$ (Lemma 3.7). The exact argument involves some calculations and the second volume lemma, and we omit it, see [11], Sect. 4B. □

Since $M = \cup_r W^s(\Omega_r)$, the previous theorem implies that the future semiorbit of almost every point $x \in M$ (with respect to the Riemannian volume) approaches some attractor. For any attractor Ω_r, the open set $W^s(\Omega_r)$ is called the *basin of attraction*. We will see now that the measure μ^+ on Ω_r describes the asymptotic distribution of the semiorbits of points $x \in W^s(\Omega_r)$. Let δ_x be the δ-measure concentrated at the point x, and

$$\delta_{x,n} = \frac{1}{n}(\delta_x + \delta_{Tx} + \cdots + \delta_{T^{n-1}x})$$

Theorem 3.14 *Let Ω_r be an attractor. For almost all points $x \in W^s(\Omega_r)$ in the basin of attraction (with respect to the Riemannian volume on M) the measure $\delta_{x,n}$ weakly converges to μ^+ (one says in this case that x is a μ^+-generic point). In other words, for any continuous function $f : M \to \mathbb{R}$ we have*

$$\lim_{n\to\infty} \frac{f(x) + f(Tx) + \cdots + f(T^{n-1}x)}{n} = \int f\, d\mu^+ \tag{3.2}$$

We only outline the proof. It is enough to prove the theorem for a.e. $x \in W^s_\varepsilon(\Omega_r)$, since the orbit of any other point x will eventually come to $W^s_\varepsilon(\Omega_r)$, and the preimages of sets of zero volume under T are sets of zero volume. Next, since μ^+ is an ergodic measure, the ergodic theorem guarantees the weak convergence of $\delta_{x,n}$ to μ^+ for μ^+-almost all $x \in W^u_\varepsilon(\Omega_r)$. It remains to prove that the Riemanniam volume is 'similar enough' to the measure μ^+ within the set W^u_ε. This can be done with the help of First Volume Lemma 3.5, see also Corollary 3.8. The exact argument is quite involved, though, we refer the reader to [11], Sect. 4C.

Corollary 3.15 *Let ν be an absolutely continuous probability measure on the basin of attraction $W^s(\Omega_r)$. Then the measure $n^{-1}(\nu + T_*\nu + \cdots + T_*^{n-1}\nu)$ weakly converges to μ^+.*

Remark 3.2. In the case of topologically mixing basic sets Ω_r, as well as Anosov diffeomorphisms, the measure $T_*^n\nu$ weakly converges to μ^+. This can be proved based on two facts: (i) the 'equivalence' of the Lebesgue volume and the measure μ^+ proclaimed by Corollary 3.8 and (ii) the Bernoulli property of the measure μ^+. We refer the reader to 5.3 in [13], where a similar statement was proved for Axiom A flows.

Definition. The measure μ^+ on an Axiom A attractor or a transitive Anosov diffeomorphism is called a *Sinai-Ruelle-Bowen (SRB) measure*. (See also a more general definition of SRB measures in Section 6.)

The above theorem shows that the SRB measure μ^+ is the only *physically observable* measure. In numerical experiments with physical models, one picks an initial point $x \in M$ at random (with respect to the volume, of course) and follows its orbit $T^n x$, $n \geq 0$. The identity between the time averages and space averages proclaimed by (3.2) is the basic property that characterizes physically meaningful measures, *steady states*, in statistical mechanics. Note that any absolutely continuous ergodic measure is physically observable.

Lyapunov exponents. Let $T : M \to M$ be a diffeomorphism. Oseledec' theorem [54] implies that for any T-invariant probability measure μ and for μ-almost every point $x \in M$ there are Lyapunov exponents $\lambda_1 \leq \cdots \leq \lambda_{\rm dimM}$ at x. They describe exponential expansion rates of tangent vectors for $\lambda_i > 0$, contraction rates for $\lambda_i < 0$. Zero exponents correspond to neutral vectors (neither expanding nor contracting exponentially). If μ is ergodic, then the Lyapunov spectrum $\{\lambda_1, \ldots, \lambda_{\rm dimM}\}$ is the same at μ-almost every $x \in M$.

In the case of Anosov and Axiom A diffeomorphisms there are no zero Lyapunov exponents. All positive Lyapunov exponents correspond to unstable tangent vectors $v \in E_x^u$ and all negative Lyapunov exponents to stable tangent vectors $v \in E_x^s$. While different points may have different Lyapunov exponents (or no Lyapunov exponents at all), for every Gibbs measure μ the Lyapunov exponents exist almost everywhere, and are a.e. constant on M.

For an ergodic measure μ, let

$$\Lambda_\mu^+ = \Sigma^+ \lambda_i \qquad \mu\text{-a.e.}$$

be the sum of positive Lyapunov exponents (which are μ-almost everywhere constant). We need the following technical lemma whose proof is left to the reader.

Lemma 3.16 *For every ergodic measure μ of an Axiom A or Anosov diffeomorphism we have $\Lambda_\mu^+ = -\mu(\varphi^u)$.*

The variational principle (2.51) applies to the present context, cf. 4.4.d in [44], and so we get $P_{\varphi^u} = \sup_\mu [h_\mu(T) - \Lambda_\mu^+]$. Now Lemma 3.7 and Theorem 3.13 imply

Theorem 3.17 *Let Ω_r be a basic set and μ an ergodic measure on Ω_r. Then $h_\mu(T) \le \Lambda_\mu^+$. We have*

$$h_\mu(T) = \Lambda_\mu^+ \tag{3.3}$$

if and only if Ω_r is an attractor and $\mu = \mu^+$ the SRB measure.

Hence, the identity (3.3), known as Pesin's formula, is another characteristic property of SRB measures.

Smooth and SRB measures. We discuss here Anosov diffeomorphisms that preserve a smooth measure. It follows immediately from Corollary 3.15 that if T preserves an absolutely continuous probability measure ν, then $\nu = \mu^+$ is the SRB measure.

Theorem 3.18 *Let $T : M \to M$ be a transitive Anosov diffeomorphism. The following are equivalent:*
(a) *T preserves an absolutely continuous measure μ*
(b) *T admits an invariant measure whose density is Hölder continuous*
(c) *for every periodic point $x = T^n x$ the map $DT^n : \mathcal{T}_x M \to \mathcal{T}_x M$ has determinant ± 1, i.e. the volume is (locally) preserved along every periodic orbit.*

Proof. Clearly, (b) implies (a). Assume that (a) holds, then $\mu = \mu^+$. Note that T^{-1} is also an Anosov diffeomorphism, with $E^u_{x,T^{-1}} = E^s_{x,T}$. Consider the function $\varphi^s(x) = \ln J^s_x$ where J^s_x is the Jacobian of the linear map $DT : E^s_x \to E^s_{Tx}$. That function is cohomologous to the function $\varphi^u_{T^{-1}}(x)$. Denote by μ^- the Gibbs measure for the function φ^s, it corresponds to the μ^+ measure for T^{-1}. Thus, $\mu^+ = \mu = \mu^-$ and $P_{\varphi^u} = 0 = P_{\varphi^s}$. By Proposition 3.4, for $T^n x = x$ we have $S_n\varphi^u(x) - S_n\varphi^s(x) = 0$. Exponentiating gives (c).

Lastly, let (c) hold. Consider the smooth function $\psi(x) = \ln \mathrm{Jac}(DT : \mathcal{T}_x M \to \mathcal{T}_{Tx} M)$. Then for every periodic point $T^n x = x$ we have $S_n\psi(x) = 0$. Proposition 3.4 implies that $\psi = u \circ T - u$ for some Hölder continuous function u on M. Then any measure whose density is proportional to $e^{u(x)}$ is T-invariant.

Remark 3.3. In fact, the density of the invariant measure μ in the above theorem is C^1 rather than just Hölder continuous, see [51, 52].

Corollary 3.19 *In the space of transitive C^2 Anosov diffeomorphisms, those without absolutely continuous invariant measures form an open dense subset.*

Proof. To prove the corollary one needs the structural stability of Anosov diffeomorphisms: any C^2 perturbation T_1 of an Anosov diffeomorphism T is also an Anosov diffeomorphism, whose trajectories are close to those of T, see [2, 43]. In particular, if $T^n x = x$, then there is a point x_1 close to x such that $T_1^n x_1 = x_1$. Now, if $\text{Jac}DT^n(x) = 1$, then there is a small perturbation T_1 such that $\text{Jac}DT_1^n(x_1) \neq 1$, which proves the density claim. If $\text{Jac}DT^n(x) \neq 1$ for some $n \geq 1$ and $T^n x = x$, then for any small perturbation T_1 we have $\text{Jac}DT_1^n(x_1) \neq 1$. This proves the openness.

In some models of mathematical physics, smooth invariant measures correspond to the so called equilibrium states, which are usually established in classical Hamiltonian systems in the absence of external forces. On the contrary, small external forces or other perturbations of Hamiltonian equations of motion often destroy smooth invariant measures and create *nonequilibrium dynamics*. In that case natural measures describing the dynamics are those satisfying (3.2) for typical points x in the phase space. Such measures are called *nonequilibrium steady states*. We have seen that for Anosov diffeomorphisms and Axiom A attractors without smooth invariant measures, nonequilibrium steady states are SRB measures.

Smoothness of SRB measures. The SRB measures have another remarkable property. Their conditional distributions on unstable manifolds are smooth with respect to the interior Lebesgue measure on those manifolds. This was first observed by Sinai for Anosov systems. He constructed invariant measures with smooth distributions on unstable manifolds in [74] and later proved that those measures are Gibbs measures with potential φ^u, i.e. SRB measures, see [76]. Sinai's original proofs are quite involved, we outline a simpler but less rigorous argument.

Let $W^u \subset M$ be an unstable manifold and $\mathcal{R}$ a Markov partition. Take a rectangle $R \in \mathcal{R}$, a point $x \in W^u \cap R$ and consider $W_0^u := W^u(x) \cap R$. Note that $W_0^u = \overline{\text{int}\, W_0^u}$. The conditional measure induced by the SRB measure μ^+ on W_0^u (let us call that measure ν_0^u) is the weak limit of the conditional measures induced by μ^+ on thin u-subrectangles of R converging to W_0^u. The existence of that limit was shown in the end of Section 2.

Symbolically, W_0^u corresponds to a one-sided sequence $\xi \in \Sigma_{(-\infty,0]}$. Then the measure ν_0^u on W_0^u can be lifted to a measure on the 'cylinder' $C(\xi) = \pi^{-1}_{(-\infty,0]}\xi \subset \Sigma_A$, that new measure will be denoted by the same symbol, ν_0^u. It is characterized by its values

$$\nu_0^u(C(\xi) \cap C(\omega_{[0,m]})) = \lim_{n\to\infty} \mu^+(C(\xi|_{[-n,0]}) \cap C(\omega_{[0,m]}))/\mu^+(C(\xi|_{[-n,0]}))$$

on subsets $C(\xi) \cap C(\omega_{[0,m]})$ of $C(\xi)$. The Eq. (2.25) and the fact that, on attractors, $P_{\varphi^u} = 0$ imply that

$$\nu_0^u(C(\xi) \cap C(\omega_{[0,m]})) \asymp \exp\left[S_m\varphi^u(\underline{\omega})\right] \tag{3.4}$$

for an arbitrary $\underline{\omega} \in C(\xi) \cap C(\omega_{[0,m]})$. On the other hand, the set $T^m\pi(C(\xi) \cap C(\omega_{[0,m]}))$ is a closed domain in the unstable manifold $W^u(T^m x)$ that stretches across some rectangle of the partition $\mathcal{R}$. Hence, its Lebesgue volume is of order one. Recall that $\exp\left[S_m\varphi^u(\underline{\omega})\right] = 1/\mathrm{Jac}(DT^m : E_x^u \to E_{T^m x}^u)$ whenever $x = \pi(\underline{\omega})$, and the same holds for all points $x \in W_0^u$. Applying Remark 2.7 to get the uniformity in x, we arrive at

$$\mathrm{Vol}^u(\pi(C(\xi) \cap C(\omega_{[0,m]}))) \asymp \exp\left[S_m\varphi^u(\underline{\omega})\right] \tag{3.5}$$

where Vol^u is the interior Lebesgue volume on W_0^u. Comparing (3.4) and (3.5) shows that the measure ν_0^u on W_0^u is equivalent to the Lebesgue volume.

u-SRB measures. One can actually compute the density of the measure ν_0^u with respect to the Lebesgue volume on W_0^u, let us denote it by $f_0(x)$. Let ν be any measure on M equivalent to the Lebesgue volume. Its conditional distributions on unstable manifolds are continuous and have positive densities. Let $x, y \in W_0^u$. According to Corollary 3.15,

$$\frac{f_0(x)}{f_0(y)} = \lim_{n\to\infty} \frac{g_0(x) + \cdots + g_{n-1}(x)}{g_0(y) + \cdots + g_{n-1}(y)}$$

where g_i is the conditional density of the measure $T_*^i\nu$ on W_0^u. Clearly,

$$\frac{g_i(x)}{g_i(y)} = \frac{g(T^{-i}x)}{g(T^{-i}y)} \cdot \frac{J^u_{T^{-i}y} \cdots J^u_{T^{-1}y}}{J^u_{T^{-i}x} \cdots J^u_{T^{-1}x}}$$

where g stands for the conditional density of the measure ν on the corresponding unstable manifolds. Since the function J^u_x is uniformly smooth on unstable manifolds and the past trajectories $\{T^{-i}x\}$, $\{T^{-i}y\}$ converge to each other exponentially as $i \to \infty$, there is a limit

$$F(x,y) := \lim_{n\to\infty} \frac{J^u_{T^{-n}y} \cdots J^u_{T^{-1}y}}{J^u_{T^{-n}x} \cdots J^u_{T^{-1}x}} \tag{3.6}$$

It then follows that $f_0(x)/f_0(y) = F(x,y)$ for all $x, y \in W^u_0$. This, together with the normalization of the measure ν^u_0, determines its density f_0 on W^u_0 completely.

Remark 3.4. It is easy to see from the above argument that the density f_0 is Lipschitz continuous. A more careful analysis shows that if T is C^r, then actually f_0 is C^{r-1} smooth (at the same time the unstable and stable manifolds are as smooth as T, i.e. C^r).

The formula (3.6) for smooth hyperbolic maps was first found by Anosov and Sinai [4].

Definition. Let W^u be a local unstable manifold, i.e. a compact domain on a global unstable manifold. Then the unique probability measure ν^u on W^u whose density $f(x)$ with respect to the induced Lebesgue volume on W^u satisfies the equation

$$f(x)/f(y) = F(x,y) \qquad \forall x, y \in W^u$$

is called the u-SRB measure (on W^u).

4 Gibbs measures for Anosov and Axiom A flows

This section is devoted to hyperbolic systems with continuous time – flows. We will explain how the theory of Gibbs measures and particularly SRB measures extends to smooth hyperbolic flows. First we define appropriate classes of flows.

Let $\varphi^t : M \to M$ be a C^2 smooth flow on a C^∞ smooth compact Riemannian manifold M. This means that the trajectories of φ^t are defined by ordinary differential equations $d\varphi^t/dt = v(x)$, where $v(x)$ is a C^2 vector field on M. Fixed points of φ^t correspond to zeroes of the field v.

A φ^t-invariant subset $X \subset M$ without fixed points is said to be *(uniformly) hyperbolic* if for all $x \in X$ there exists a $D\varphi^t$-invariant decomposition

$$\mathcal{T}_x M = E_x^\varphi \oplus E_x^u \oplus E_x^s \tag{4.1}$$

such that
(i) E_x^φ is a one-dimensional subspace spanned by the vector $v(x)$;
(ii) $\dim E_x^{u,s} \neq 0$ and there are $C > 0$ and $\lambda < 1$ such that

$$||D\varphi^t(w)|| \leq C\lambda^t ||w|| \qquad \text{for } \; w \in E_x^s,\; t \geq 0$$

$$||D\varphi^{-t}(w)|| \leq C\lambda^t ||w|| \qquad \text{for } \; w \in E_x^u,\; t \geq 0$$

There is always a metric on M (called Lyapunov or adapted metric) in which $C = 1$, i.e. the above contraction of tangent vectors is monotone in time.

It is easy to see that the splitting (4.1) depends on x continuously. The space E_x^φ depends on x smoothly, but this is not true for $E_x^{u,s}$. In fact, just like for Anosov and Axiom A diffeomorphisms, the spaces $E_x^{u,s}$ depend on x Hölder continuously, see [3, 43]. Anosov [2] showed that $E_x^{u,s}$ may not be differentiable, even if $v(x)$ is real analytic. The original Anosov's argument worked in high dimensions only, but later Plante [60] showed that $E_x^{u,s}$ were not necessarily differentiable even in the simplest case $\dim E_x^u = \dim E_x^s = 1$. In this last case, however, the two-dimensional families $E_x^u \oplus E_x^\varphi$ and $E_x^s \oplus E_x^\varphi$ are C^1 smooth, see [38] for more.

Definition. A flow $\varphi^t : M \to M$ is called an *Anosov flow* if the entire manifold M is a uniformly hyperbolic set (in particular, the flow has no fixed points in M).

Definition. A flow $\varphi^t : M \to M$ satisfies Smale's *Axiom A* if the set of nonwandering points $\Omega \subset M$ is a disjoint union $\Omega = X \cup F$, where X is a closed uniformly hyperbolic set in which periodic orbits of the flow are dense, and F is a finite set consisting of isolated hyperbolic fixed points.

Theorem 4.1 (Smale's spectral decomposition) *For an Axiom A flow, we have $X = X_1 \cup \cdots \cup X_m$, where X_i are disjoint closed φ^t-invariant subsets such that*

(a) *the flow* $\varphi^t : X_i \to X_i$ *is transitive (i.e. there is a dense orbit in each* X_i*);*
(b) *there is an open set* $U_i \supset X_i$ *such that* $X_i = \cap_{t=-\infty}^{\infty} \varphi^t(U_i)$.
The sets X_i *are called the* basic sets *of the flow.*

An Anosov flow is transitive if there is a dense orbit in M. In that case 4.1 implies that the entire manifold M is a single basic set of the flow.

Recall that the transitivity of all Anosov diffeomorphisms is still an open problem at present, and we had to assume the transitivity for Anosov diffeomorphisms in Sections 1–3 to obtain reasonable results. The corresponding open problem for Anosov flows was solved by J. Franks and R. Williams in the late 70-s [29]: they found examples of nontransitive Anosov flows! So, they justifiably called such Anosov flows 'anomalous'. To avoid such anomalities, we will always assume transitivity for Anosov flows discussed in this section.

Gibbs measures can be defined for Anosov and Axiom A flows. The constructions of Gibbs measures for these two classes of flows are almost identical, the two cases only differ in relatively minor technical details. For brevity, we often mention just Anosov flows, but everything goes through for Axiom A flows as well. Furthermore, compact locally maximal topologically transitive hyperbolic sets of any smooth flow have all the characteristic properties of Axiom A flow basic sets, so our constructions apply as well.

The first step in the construction of Gibbs measures is to define a version of Markov partitions and symbolic dynamics for Anosov flows.

Markov sections. For every $x \in M$ and $\delta > 0$ denote by $W_\delta^s(x)$ and $W_\delta^u(x)$ the local stable and unstable manifolds (of size δ) through x. For small $\varepsilon > 0$, consider

$$W^{wu}(x) := \varphi^{[-\varepsilon,\varepsilon]} W_\delta^u(x) \quad \text{and} \quad W^{ws}(x) := \varphi^{[-\varepsilon,\varepsilon]} W_\delta^s(x)$$

the so called local *weakly* unstable and *weakly* stable manifolds, respectively. Here and further on we adopt Bowen's notation

$$\varphi^{[a,b]} A := \cup_{a \le t \le b} \varphi^t A$$

for any $A \subset M$. Weakly unstable and stable manifolds have one extra dimension as compared to $W^u(x)$ and $W^s(x)$. The manifolds $W^u(x)$ and $W^s(x)$ are often called, accordingly, *strongly* unstable and stable manifolds.

A closed subset $R \subset M$ is called a *rectangle* if there is a small closed codimension one smooth disk $D \subset M$ transversal to the flow φ^t, such that $R \subset D$, and for any $x, y \in R$ the point

$$[x, y]_R := D \cap W^{ws}(x) \cap W^{wu}(y)$$

exists and also belongs to R. A rectangle R is said to be *proper* if $R = \overline{\operatorname{int} R}$ in the internal topology of D. For any rectangle R and $x \in R$ we put

$$W^u(x, R) := R \cap W^{wu}(x) \quad \text{and} \quad W^s(x, R) := R \cap W^{ws}(x)$$

Then R is a direct product of the sets $W^u(x, R)$ and $W^s(x, R)$ in the sense of the map $(y, z) \mapsto [y, z]_R$ for $y, z \in R$.

Definition (cf. [10]). A finite collection of closed sets $\mathcal{R} = \{R_1, \dots, R_m\}$ is said to be a *proper family of size* $\alpha > 0$ if
(i) $M = \varphi^{[-\alpha,0]}(\mathcal{S})$, where $\mathcal{S} = R_1 \cup \dots \cup R_m$;
and there are disks $D_1, \dots, D_m$ containing these sets such that for every i
(ii) $\operatorname{diam} D_i < \alpha$;
(iii) $R_i = \overline{\operatorname{int} R_i}$;
(iv) for any $i \neq j$ at least one of the two sets $D_i \cap \varphi^{[0,\alpha]} D_j$ and $D_j \cap \varphi^{[0,\alpha]} D_i$ is empty; in particular, $D_i \cap D_j = \emptyset$.

It follows from (i) that for any $x \in \mathcal{S}$ there is a smallest positive return time $l(x)$ such that $\varphi^{l(x)}(x) \in \mathcal{S}$. According to (i) and (iv), the function $l(x)$ is bounded from above and below: $0 < l_{\min} \le l(x) \le l_{\max} < \infty$. The set $\mathcal{S}$ is called a cross-section of the flow φ^t. It generates the first return map on $\mathcal{S}$ (Poincaré map) defined by $T(x) = \varphi^{l(x)}(x)$. Note that $T : \mathcal{S} \to \mathcal{S}$ is a one-to-one map. It is easy to check that the functions $l(x)$ and $T(x)$ are locally as smooth as the flow φ^t, i.e. of class C^2. They are not continuous globally on $\mathcal{S}$, see below.

We note that $l(x)$ and $T(x)$ are smooth at all $x \in (\cup_i \operatorname{int} R_i) \setminus \cup_i (T^{-1} \partial R_i)$. All the iterations of T are continuous on the T-invariant set

$$\mathcal{S}^* = \left\{ x \in \mathcal{S} : T^k x \in \cup_{i=1}^m \operatorname{int} R_i \quad \text{for all} \quad k \in \mathbb{Z} \right\}$$

Definition (cf. [10]). A proper family of sets $\mathcal{R} = \{R_1, \dots, R_m\}$ of a small size α is said to be a *Markov section* (or a *Markov family*), if
(i) every $R_i \in \mathcal{R}$ is a (proper) rectangle;
(ii) we have the Markov property: if $x \in \operatorname{int} R_i \cap T^{-1}(\operatorname{int} R_j)$, then

$$W^s(x, R_i) \subset \overline{T^{-1}(W^s(Tx, R_j))} \quad \text{and} \quad \overline{T(W^u(x, R_i))} \supset W^u(Tx, R_j)$$

Theorem 4.2 *Any transitive Anosov flow $\varphi^t : M \to M$ has Markov sections of arbitrary small size. The same is true for the restriction of an Axiom A flow to any basic set.*

One can modify the construction of Markov partitions for Anosov and Axiom A diffeomorphisms to prove this theorem. This was done by Bowen, and we refer the reader to his paper [10] for details.

We will explain next that the theory of Anosov diffeomorphisms, at least in its essential parts, applies to the map $T : \mathcal{S} \to \mathcal{S}$.

Just like for Anosov diffeomorphisms, we introduce the following notation for flows. For any rectangle R put $\partial R = \partial^u R \cup \partial^s R$, where $\partial^u R = \cup_{x\in R} \partial W^s(x, R)$ and $\partial^s R = \cup_{x\in R} \partial W^u(x, R)$. We put $\partial^u \mathcal{S} = \cup_i \partial^u R_i$ and $\partial^s \mathcal{S} = \cup_i \partial^s R_i$, then $\partial \mathcal{S} = \partial^u \mathcal{S} \cup \partial^s \mathcal{S}$. Now we have, as in (1.7), that $T\,(\partial^s \mathcal{S}) \subset \partial^s \mathcal{S}$ and $T^{-1}\,(\partial^u \mathcal{S}) \subset \partial^u \mathcal{S}$. Note that $\partial S \cup T^{\pm 1}\partial \mathcal{S}$ is a closed nowhere dense subset of $\mathcal{S}$, so the set $\mathcal{S}^*$ above is a residual set. The set $\mathcal{S}^*$ will be essential for us, while what happens in the complement $\mathcal{S} \setminus \mathcal{S}^*$ will not be so essential, see Remark 4.1 below.

The cross-section $\mathcal{S}$ is now a submanifold of codimension one in M, it is a finite union of connected submanifolds with boundary. The function $l(x)$ and the Poincaré map T are piecewise smooth (of class at least C^2) on $\mathcal{S}$. The discontinuities of $l(x)$ and T are in $T^{-1}(\partial^s \mathcal{S})$.

For every $R_i \in \mathcal{R}$ and $x \in R_i$ let $E_x^{s,u}$ be the tangent planes to $W^{s,u}(x, R_i)$, respectively. The splitting

$$\mathcal{T}_x \mathcal{S} = E_x^s \oplus E_x^u \tag{4.2}$$

is DT-invariant and there are constants $C_T > 0$ and $\lambda_T \in (0, 1)$ such that

$$\begin{aligned} ||DT^n(v)|| &\leq C_T \lambda_T^n ||v|| \quad \text{for } v \in E_x^s,\ n \geq 0 \\ ||DT^{-n}(v)|| &\leq C_T \lambda_T^n ||v|| \quad \text{for } v \in E_x^u,\ n \geq 0. \end{aligned} \tag{4.3}$$

(at singular points $x \in \partial \mathcal{S} \cup T^{-1}(\partial^s \mathcal{S})$ a one-sided derivative DT can be used in (4.3)). The splitting (4.2) is Hölder continuous in x, since so is the splitting (4.1). The sets $W^{s,u}(x, R_i)$ are stable and unstable manifolds for T, respectively.

It is important to note that the map $T : \mathcal{S} \to \mathcal{S}$ is not exactly an Anosov diffeomorphism, since $\mathcal{S}$ is disconnected and has boundary, and T has singularities on $\mathcal{S} \cap T^{-1}(\partial^s \mathcal{S})$. Nonetheless, the restriction of T to the set $\mathcal{S}^*$

behaves exactly like the restriction of an Anosov diffeomorphism $T : M \to M$ to the essential set $M \setminus M^{\#}$ defined in Section 1, cf. Theorem 1.3. Thus, the partition $\mathcal{R}$ of $\mathcal{S}$ has all essential properties of a Markov partition for an Anosov diffeomorphism. Some minor technical differences can be easily taken care of, see also below.

Symbolic dynamics. The Markov partition $\mathcal{R}$ of $\mathcal{S}$ produces a topological Markov chain (Σ_A, σ) with a transition matrix A of size $m \times m$ defined by (1.8).

Some constructions of Section 1 have to be slightly modified, this concerns the boundary $\partial\mathcal{S}$, which looks and behaves differently from $\partial\mathcal{R}$ in Sect. 1. In particular, the projection $\pi : \Sigma_A \to \mathcal{S}$ must now be defined as follows:
(i) for every $x \in \mathcal{S}^*$, let $\underline{\omega} = \{\omega_n\} \in \Sigma_A$ be defined by $T^n x \in R_{\omega_n}$ for all $n \in \mathbb{Z}$, then we set $\pi(\underline{\omega}) = x$. This defines π continuously on a dense residual subset

$$\Sigma_A^* := \pi^{-1}\mathcal{S}^* \subset \Sigma_A$$

(ii) on the set $\Sigma_A \setminus \Sigma_A^*$, the map π is simply extended by continuity.

Lemma 4.3 *(i) The map π is surjective and one-to-one on $\Sigma_A^* = \pi^{-1}(\mathcal{S}^*)$. (ii) Also, π is Hölder continuous on Σ_A in any metric d_α, see Sect. 2. (iii) Lastly, $\pi(\underline{\omega}) = \cap_{n=-\infty}^{\infty} T^{-n} R_{\omega_n}$ for all $\underline{\omega} \in \Sigma_A^*$. (This does not necessarily hold for $\underline{\omega} \in \Sigma_A \setminus \Sigma_A^*$.)*

The proof is left as an exercise.

Remark 4.1. We have $\rho(\Sigma_A^*) = 1$ for every Gibbs measure ρ on Σ_A (recall Corollary 3.2).

The last remark indeed shows that the set $\mathcal{S} \setminus \mathcal{S}^*$ is not so essential and can be neglected in measure-theoretic considerations.

Remark 4.2. The identity $T \circ \pi = \pi \circ \sigma$ holds on Σ_A^* but may fail on $\Sigma_A \setminus \Sigma_A^*$. However, it is possible to redefine slightly the function $l(x)$ (and hence the map $T(x) = \varphi^{l(x)}(x)$) on $\partial\mathcal{S}$ so that the identity $T \circ \pi = \pi \circ \sigma$ will hold on the entire set Σ_A. This was observed by Bowen [10].

Since the flow $\varphi^t : M \to M$ (or the restriction of φ^t to X_r in the Axiom A case) is topologically transitive, then the map $T : \mathcal{S} \to \mathcal{S}$, and hence the TMC (Σ_A, σ), are transitive as well. However, the TMC need not be mixing (even if the flow φ^t is) for the reason being clear from the following argument. We claim that it is always possible to find a Markov section $\mathcal{R}$ that produces a mixing TMC (Σ_A, σ), even if the flow φ^t is not topologically mixing. Indeed, let σ, and hence the map $T : \mathcal{S}^* \to \mathcal{S}^*$ be not topologically mixing. Then $\mathcal{S} = \mathcal{S}^{(1)} \cup \cdots \cup \mathcal{S}^{(k)}$, with $\mathcal{S}^{(i)} \cap \mathcal{S}^{(j)} = \emptyset$ for $i \neq j$, and $T(\mathcal{S}^{(i)}) = \mathcal{S}^{(i+1)}$ (with $\mathcal{S}^{(k+1)} := \mathcal{S}^{(1)}$), and every $\mathcal{S}^{(i)}$ is a union of some rectangles of $\mathcal{R}$, and the map T^k is mixing on $\mathcal{S}^{(1)}$. Then we replace $\mathcal{S}$ by $\mathcal{S}^{(1)}$, the function $l(x)$ by $l(x) + \cdots + l(T^{k-1}x)$ and the map T by T^k. Now the new map T, and hence the corresponding TMC (Σ_A, σ), are topologically mixing. The above trick was described by Bowen and Ruelle [13].

Now we define symbolic dynamics for the flow φ^t. Consider the function $\tau : \Sigma_A \to \mathbb{R}$ defined by $\tau(\underline{\omega}) := l(\pi(\underline{\omega}))$ on Σ_A^* and extended by continuity to the entire set Σ_A. Since $l(x)$ is smooth at all $x \in \mathcal{S}^*$ and π is Hölder continuous, the function $\tau(x)$ is Hölder continuous on Σ_A^*, hence also on Σ_A. It is also bounded above and below just because so is $l(x)$, i.e. $0 < l_{\min} \leq \tau(\underline{\omega}) \leq l_{\max} < \infty$.

Note that $\varphi^{\tau(\underline{\omega})}\pi(\underline{\omega}) = \pi(\sigma(\underline{\omega}))$ for all $\underline{\omega} \in \Sigma_A$. To check this, first consider $\underline{\omega} \in \Sigma_A^*$ and then use the continuity of both τ and π.

Definition. We say that $\psi^t : Y \to Y$ is a *suspension flow* with the *base transformation* $\sigma : \Sigma_A \to \Sigma_A$ and the *ceiling function* $\tau(x)$, if

$$Y = \{(\underline{\omega}, s) : \underline{\omega} \in \Sigma_A, 0 \leq s \leq \tau(\underline{\omega})\}$$

with identification of points $(\underline{\omega}, \tau(\underline{\omega}))$ and $(\sigma(\underline{\omega}), 0)$, and ψ^t acts as

$$\psi^t(\underline{\omega}, s) = (\underline{\omega}, s + t)$$

(with the above identification). Note that the space Y is compact, and the flow ψ^t is continuous.

Define a projection $\tilde{\pi} : Y \to M$ by

$$\tilde{\pi} : (\underline{\omega}, s) \mapsto \varphi^s \pi(\underline{\omega})$$

The map $\tilde{\pi}$ is well defined on Y for all s due to the identification $(\underline{\omega}, \tau(\underline{\omega})) = (\sigma(\underline{\omega}), 0)$ and the identity $\varphi^{\tau(\underline{\omega})}\pi(\underline{\omega}) = \pi(\sigma(\underline{\omega}))$ proved above. One can verify by direct inspection that $\tilde{\pi}$ is continuous and surjective and commutes with our flows, i.e.

$$\varphi^t \circ \tilde{\pi} = \tilde{\pi} \circ \psi^t$$

The suspension flow $\psi^t : Y \to Y$ is a symbolic representation of the flow $\varphi^t : M \to M$. It has many other nice properties that we will not need here. For example, it is at most k-to-one for some finite k, any orbit of ψ^t is periodic if and only if so is its image under $\tilde{\pi}$, etc. We refer the reader to [10] for proofs of these and further properties.

Gibbs measures. First, we define Gibbs measures for the symbolic flow $\psi^t : Y \to Y$.

We start with an important observation: there is a natural one-to-one correspondence between ψ^t-invariant probability measures on Y and σ-invariant probability measures on the base Σ_A. Let a measures ν on Σ_A be σ-invariant and denote by m the Lebesgue measure on $\mathbb{R}$. Then the measure μ_ν first defined by $\nu \times m$, then restricted to Y and normalized, is φ^t-invariant (note that the identification $(\underline{\omega}, \tau(\underline{\omega})) = (\sigma(\underline{\omega}), 0)$ only affects a subset of μ_ν-measure zero). One can easily check that $\nu \mapsto \mu_\nu$ is a one-to-one correspondence between the space of σ-invariant probability measures, call it $\mathcal{M}(\Sigma_A)$, and that of ψ^t-invariant probability measures, call it $\mathcal{M}(Y)$. It is also easy to check that ν is ergodic if and only if so is μ_ν. This correspondence between $\mathcal{M}(\Sigma_A)$ and $\mathcal{M}(Y)$ will be very helpful.

Now, for any Hölder continuous function $g : \Sigma_A \to \mathbb{R}$ consider the Gibbs measure ρ_g on Σ_A. By the above correspondence, ρ_g produces a ψ^t-invariant measure, μ_{ρ_g}, on Y. We can construct Gibbs measures on Y in this way. It will be, however, more natural to associate Gibbs measures on Y with functions on Y rather than with functions on Σ_A.

Definition. Assume that the function $f : Y \to \mathbb{R}$ is continuous and the function

$$F(\underline{\omega}) = \int_0^{\tau(\underline{\omega})} f(\underline{\omega}, s)\, ds \tag{4.4}$$

is Hölder continuous on Σ_A. Then the function

$$g(\underline{\omega}) = F(\underline{\omega}) - P_f \cdot \tau(\underline{\omega}) \tag{4.5}$$

is also Hölder continuous on Σ_A. Here P_f is the topological pressure of the function f for the flow ψ^t, cf. 2.5.k in [44]. Now, the measure $\mu_f := \mu_{\rho_g}$ is called the *Gibbs measure* for the flow ψ^t associated with the potential f.

Note that the function $f(x)$ on Y has to be converted into a potential $g(\underline{\omega})$ on Σ_A by the rules (4.4)-(4.5) in order to define the measure μ_f.

Theorem 4.4 (Equilibrium states) (i) *The Gibbs measure* μ_f *is the unique equilibrium state for the function* f*, i.e. the unique measure on which the following supremum is attained:*

$$P_f = \sup_{\mu \in \mathcal{M}(Y)} \left(h_\mu(\psi^t) + \int_Y f \, d\mu \right)$$

(ii) *The measure* μ_f *is ergodic and positive on every open set.*

Proof. Recall that every measure $\mu \in \mathcal{M}(Y)$ is $\mu = \mu_\nu$ for some $\nu \in \mathcal{M}(\Sigma_A)$. Now

$$\int_Y f \, d\mu_\nu = \frac{\int_{\Sigma_A} F \, d\nu}{\int_{\Sigma_A} \tau \, d\nu}$$

and, according to Abramov's formula [1],

$$h_{\mu_\nu}(\psi^t) = \frac{h_\nu(\sigma)}{\int_{\Sigma_A} \tau \, d\nu}$$

Therefore,

$$\begin{aligned} P_f &= \sup_{\mu \in \mathcal{M}(Y)} \left(h_\mu(\psi^t) + \int_Y f \, d\mu \right) \\ &= \sup_{\nu \in \mathcal{M}(\Sigma_A)} \frac{h_\nu(\sigma) + \int_{\Sigma_A} F \, d\nu}{\int_{\Sigma_A} \tau \, d\nu} \end{aligned}$$

This implies that

$$P_g = \sup_{\nu \in \mathcal{M}(\Sigma_A)} \left(h_\nu(\sigma) + \int_{\Sigma_A} (F - P_f \tau) \, d\nu \right) = 0$$

We know from Section 3 that the above supremum is attained on a unique measure, the Gibbs measure $\nu = \rho_g$. This proves (i). The properties of μ_f claimed in (ii) immediately follow from similar properties of the Gibbs measure ρ_g on Σ_A. $\square$

Theorem 4.5 *For any Gibbs measure* μ_f *on* Y *the projection* $\tilde{\pi} : Y \to M$ *is* μ_f*-almost everywhere one-to-one. Therefore, the ergodic and statistical properties of the measure* μ_f *and its image* $\tilde{\pi}_* \mu_f$ *on* M *are identical. In particular, the measure* $\tilde{\pi}_* \mu_f$ *is ergodic and positive on open subsets (of* M *for Anosov flows and of* X_r *for Axiom A flows).*

The proof is based on Remark 4.1. See also a more direct proof in [13].

Let $p : M \to \mathbb{R}$ be a Hölder continuous function. Then the function $f = p \circ \tilde{\pi}$ is continuous on Y and the function $F(\underline{\omega}) = \int_0^{\tau(\underline{\omega})} f(\underline{\omega}, s)\, ds$ is Hölder continuous on Σ_A. The Hölder continuity of F follows by direct calculation based on the hyperbolicity of φ^t and the Hölder continuity of $\tau(\underline{\omega})$. We leave it as an exercise, see also 3.3 in [13]. Now one can associate the measure

$$\mu_p := \tilde{\pi}_* \mu_f$$

with the potential function $p(x)$ and the Markov section $\mathcal{R}$.

Proposition 4.6 *The measure μ_p does not depend on the choice of the Markov section $\mathcal{R}$. Furthermore, μ_p is the unique equilibrium state for the function $p(x)$, i.e. the unique measure on which the following supremum is attained*

$$P_p = \sup_{\mu} \left(h_\mu(\varphi^t) + \int_M p\, d\mu \right) \tag{4.6}$$

where the supremum is taken over all φ^t-invariant measures on M and P_p is the topological pressure of the function p with respect to the flow φ^t.

This is a direct analogue of Proposition 3.3, and can be proved similarly, see details in [13].

Remark 4.3. It can be also shown that $P_p = P_f$ where $f = p \circ \tilde{\pi}$.

Definition. The measure μ_p is called the *Gibbs measure* for an Anosov (Axiom A) flow corresponding to the potential function $p(x)$ on M (resp., X_r).

We emphasize the correspondence between the measures

$$\mu_p \text{ on } M \quad \leftrightarrow \quad \mu_f \text{ on } Y \quad \leftrightarrow \quad \rho_g \text{ on } \Sigma_A \tag{4.7}$$

with

$$f = p \circ \tilde{\pi} \quad \text{and} \quad g = \int_0^\tau f\, ds - P_f \cdot \tau$$

All the measures in (4.7) are Gibbs measures.

Let p_1, p_2 be two Hölder continuous functions on M, and $\mathcal{R}$ a Markov section. Obviously, the two Gibbs measures μ_{p_1}, μ_{p_2} on M coincide iff the measures μ_{f_1} and μ_{f_2} on Y coincide, with $f_i = p_i \circ \tilde{\pi}$. The latter happens iff the measures ρ_{g_1}, ρ_{g_2} on Σ_A coincide, with $g_i = \int_0^\tau f_i\, ds - \tau P_{f_i}$. Those two are Gibbs measures on a mixing TMC, and we can apply Theorem 2.23. Thus, we arrive at the following criterion.

Proposition 4.7 *The two Gibbs measures μ_{p_1} and μ_{p_2} coincide if and only if for any (and thus for every) Markov section $\mathcal{R}$ the functions*

$$G_1(x) = \int_0^{l(x)} p_1(\varphi^t x)\, dt - P_{p_1} l(x) \quad \text{and} \quad G_2(x) = \int_0^{l(x)} p_2(\varphi^t x)\, dt - P_{p_2} l(x)$$

are cohomologous on $\mathcal{S}$, i.e.

$$G_1(x) - G_2(x) = u(Tx) - u(x)$$

for some Hölder continuous function $u(x)$ on $\mathcal{S}$. (Note: there is no constant K in the above cohomological equation, since $K = P_{G_1} - P_{G_2}$ and $P_{G_1} = P_{G_2} = 0$, see the proof of 4.4).

Hence, there are uncountably many Gibbs measures for any Anosov and Axiom A flow.

A critical reader may notice that our discussion starts repeating that in Section 3, in particular Propositions 3.3 and 3.4. The only difference is the object of the discussion – now it is flows rather than diffeomorphisms. Essentially, though, the situation is very much the same, and the rest of the Section 3 – the theory of SRB measures – carries over to flows with very few changes. Below we summarize the main results of this theory in the case of flows, emphasizing the differences and connection between flows and diffeomorphisms. We do not provide the proofs which differ little from those in Section 3, the reader can adapt them easily.

The potential function p^u and SRB measures. Let $\Lambda_x^u(t)$ be the Jacobian of the linear map $D\varphi^t : E_x^u \to E_{\varphi^t x}^u$ and

$$p^u(x) = -\frac{d \ln \Lambda_x^u(t)}{dt}|_{t=0} = -\frac{d\Lambda_x^u(t)}{dt}|_{t=0}$$

Since the map $x \mapsto E_x^u$ is Hölder continuous and the map $E_x^u \to p^u(x)$ is differentiable, then the function $p^u(x)$ is Hölder continuous.

Definition. The Gibbs measure $\mu^+ = \mu_{p^u}$ is called a *u-Gibbs* or a *generalized Sinai-Ruelle-Bowen* measure for the flow φ^t.

Proposition 4.8 *For any $t > 0$ we have $\int_0^t p^u(\varphi^s x)\, ds = -\ln \Lambda_x^u(t)$.*

The proof is left as an exercise.

Attractors. A basic set X_r of an Axiom A flow is called an *attractor* if there is an open neighborhood U of X_r such that $\cap_{t>0}\varphi^t U = X_r$. In particular, for transitive Anosov flows, the entire manifold M is an attractor.

A basic set X_r is an attractor if and only if $X_r = \cup_{x\in X_r} W^u(x)$, i.e. X_r is the union of (global) unstable manifolds of its points. Equivalently, $X_r = \cup_{x\in X_r} W^{wu}(x)$, i.e. X_r is the union of weakly unstable manifolds of its points. Furthermore, X_r is an attractor if $W_\varepsilon^u(x) \subset X_r$ for some $x \in X_r$ and $\varepsilon > 0$, i.e. X_r contains a small open ball of an unstable manifold.

Let

$$W^s(X_r) = \{y \in M : d(\varphi^t y, X_r) \to 0 \quad \text{as} \quad t \to \infty\}$$

Then $M = \cup_r W^s(X_r)$, and this is a disjoint union (here the hyperbolic fixed points $x \in F$ are also considered as basic sets). It is also true that

$$W^s(X_r) = \cup_{x\in X_r} W^s(x) = \cup_{x\in X_r} W^{ws}(x)$$

The following are equivalent:
(a) X_r is an attractor
(b) $\mathrm{Vol}\,(W^s(X_r)) > 0$
(c) The set $W^s(X_r)$ is open
(d) $P_{p^u} = 0$ with respect to the flow $\varphi^t : X_r \to X_r$.

Let X_r be an attractor and $\mathcal{R}$ a Markov section in X_r. We denote by J_x^u the absolute value of the Jacobian of the linear map $DT : E_x^u \to E_{Tx}^u$ and put $\varphi^u(x) = -\ln J_x^u$ on $\mathcal{S}$. Consider the measure

$$\rho^+ := \rho_{\varphi^u \circ \pi}$$

on Σ_A. In a sense, ρ^+ is the SRB measure for the map $T : \mathcal{S} \to \mathcal{S}$ (which is not exactly an Axiom A attractor, but possesses all essential properties of Axiom A attractors as explained after Theorem 4.2).

Proposition 4.9 *Let X_r be an attractor and $\mathcal{R}$ a Markov section in X_r. The generalized SRB measure μ^+ on X_r corresponds to the measure ρ^+, in the sense of (4.7). This also holds for transitive Anosov flows.*

Proof. Proposition 4.8 implies that

$$\int_0^{\tau(\underline{\omega})} f^u \circ \tilde{\pi}(\underline{\omega}, s)\, ds = -\ln \Lambda^u_{\pi(\underline{\omega})}(\tau(\underline{\omega})) =: g^+(\underline{\omega})$$

Since $P_{p^u} = P_{p^u \circ \tilde{\pi}} = 0$ (cf. Remark 4.3), the measure μ^+ corresponds to the Gibbs measure ρ_{g^+} on Σ_A. One can easily verify that the function g^+ is cohomologous to $\varphi^u \circ \pi$ on Σ_A^*. Hence, $\rho_{g^+} = \rho^+$. □

Definition. The measure μ^+ on an attractor X_r of an Axiom A flow (respectively, on M for a transitive Anosov flow) is called the *Sinai-Ruelle-Bowen (SRB) measure*.

Since $M = \cup_r W^s(X_r)$, the future semiorbit of almost every point $x \in M$ (with respect to the Riemannian volume) approaches some attractor. For any attractor X_r, the open set $W^s(X_r)$ is called the *basin of attraction*. The measure μ^+ on the attractor X_r describes the asymptotic distribution of the semiorbits of generic points $x \in W^s(X_r)$ in the following sense. For almost every point $x \in W^s(X_r)$ and any continuous function $q : M \to \mathbb{R}$ we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T q(\varphi^t x)\, dt = \int q\, d\mu^+ \tag{4.8}$$

Furthermore, if the restriction of the flow φ^t to X_r is topologically mixing, then for any absolutely continuous probability measure ν on the basin of attraction $W^s(X_r)$ the measure $\varphi^t_* \nu$ weakly converges to μ^+, see a proof in [13]. In that case μ^+ is K-mixing and Bernoulli, see also Theorem 4.11 below.

In the sense of (4.8), the SRB measures on attractors are the only physically observable measures. Adopting physical terminology, one can call them steady states for Anosov and Axiom A flows.

Anosov flows and SRB measures. Let $\varphi^t : M \to M$ be a transitive Anosov flow. If φ^t preserves an absolutely continuous probability measure μ, then certainly $\mu = \mu^+$. In this case the density of μ with respect to the Riemannian volume is C^1 smooth. The Anosov flow φ^t preserves a smooth measure if and only if for every periodic orbit $\varphi^t x = x$ (of period t) we have $D\varphi^t(x) = 1$, i.e. the volume is locally preserved along that orbit. In the space of transitive C^2 Anosov flows, those without absolutely continuous invariant measures form an open dense subset.

The SRB measures for Anosov flows and Axiom A attractors have smooth conditional distributions on unstable manifolds. For any local unstable manifold $W^u \subset M$ (resp., $W^u \subset X_r$) the density $f(x)$ of the conditional distribution of μ^+ on W^u satisfies

$$\frac{f(x)}{f(y)} = \lim_{t\to-\infty} \frac{\Lambda^u_x(t)}{\Lambda^u_y(t)}$$

We can also call the unique probability measure on W^u satisfying the above condition the u-SRB measure on W^u.

All the above properties can be proved in the same way as those of Anosov diffeomorphisms.

This concludes the extension of the results of Section 3 to hyperbolic flows.

Mixing flows versus cycles. There is a serious difference between hyperbolic flows and diffeomorphisms: some flows are not topologically mixing for one reason that is specific for flows, as illustrated below.

Example 4.1. Let $T : M_1 \to M_1$ be a transitive Anosov diffeomorphism and $c > 0$. Consider the suspension flow over the base M_1 under a constant ceiling function $\tau(x) \equiv c$. It is defined on the compact space $M = M_1 \times [0, c]$ with the usual identification of points $(x, c) = (Tx, 0)$. The space M can be endowed with a metric in which it will be a smooth compact manifold without boundary, and the suspension flow φ^t will be Anosov. It is clearly transitive but not mixing. No φ^t-invariant measure can be mixing either. In the t-direction, the flow is just a circle rotation.

Definition. We call a smooth flow $\varphi^t : M \to M$ a *cycle* if it admits a smooth cross-section with a constant return function. Cycles can never be topologically mixing.

Example 4.2. In Example 4.1, one can slightly refine the ceiling function, retaining the nontransitivity. Assume that $\tau(x)$ takes values *commensurate to a constant*, i.e. $\tau(x) = a\, n(x)$, where $a > 0$ is a constant and $n(x) > 0$ is an integer-valued function. In this case one can easily find another Markov section in M so that the suspension flow will again have a constant return function, so φ^t is a cycle.

Remark 4.4. Let a flow φ^t have a cross-section $\mathcal{S}$ with return function $l(x)$ and the return map $T(x) = \varphi^{l(x)}(x)$ on $\mathcal{S}$. Assume that $l(x) = l_0(x)+u(Tx)-$

$u(x)$, where l_0 is commensurate to a constant and $u(x)$ is a smooth function on $\mathcal{S}$. Then φ^t is a cycle. To show this, just consider new cross-section $\mathcal{S}_1 = \{\varphi^{u(x)}(x) : x \in \mathcal{S}\}$.

Theorem 4.10 (Anosov alternative) *Let $\varphi^t : M \to M$ be a transitive Anosov flow. Then either*
(i) *φ^t is a cycle, or*
(ii) *the global unstable manifold $W^u(x)$ and the global stable manifold $W^s(x)$ are both dense in M for every $x \in M$. In this case the flow is topologically mixing.*

The proof of this theorem involves topological considerations and resembles that of Smale's spectral decomposition 4.1. We refer the reader to a well written Plante's paper [60].

Remark 4.5. In the cycle case (i) above, for any Markov section $\mathcal{R}$ in M the return function $l(x)$ on $\mathcal{S}$ is cohomologous to a function whose values are commensurate to a constant. This follows from some general results by Gurevich [37] on suspension flows. So, Examples 4.1, 4.2 and Remark 4.4 essentially describe all types of cycles.

It is interesting to note that if an Anosov flow has two periodic orbits with incommensurate periods, then it is topologically mixing.

Remark 4.6. The property claimed in (ii) of Theorem 4.10 is often taken as a definition of topological mixing for Anosov flows.

Geodesic flows. There is a well studied class of Anosov flows. Every geodesic flow on a smooth compact connected Riemannian manifold with negative sectional curvatures is a topologically mixing Anosov flow. Hyperbolic properties of these geodesic flows were observed by Hopf [41], who also proved the ergodicity for these flows. Actually, motivated by that, Anosov introduced his class of flows, which he originally called C-flows, they are now called Anosov flows. Arnold proved [6, 7] that the above geodesic flows are always topologically mixing, see also Anosov [2]. In addition, they are K-mixing and Bernoulli with respect to the natural smooth invariant measures.

Theorem 4.11 ([76, 53, 65]) *Let φ^t be a topologically mixing Anosov or Axiom A flow on a basic set X_r. Then every Gibbs measure μ_p on M (resp., X_r) is mixing, K-mixing and Bernoulli.*

Joint integrability. Let $\varphi^t : M \to M$ be an Anosov flow. The spaces E_x^u, E_x^s are said to be *jointly integrable* if the family of hyperplanes $E_x^u \oplus E_x^s$ in $\mathcal{T}M$ is integrable, i.e. if there is a C^1 foliation of M by codimension one submanifolds tangent to $E_x^u \oplus E_x^s$. In this case every unstable and stable manifold wholly lies in one leaf of that foliation. Clearly, this is the case for cycles. Indeed, if $\mathcal{S}$ is a cross-section with a constant return function $l(x)$, then all the hyperplanes $E_x^u \oplus E_x^s$ are tangent to level surfaces $l(x)$ =const.

A simple geometric construction can be used to detect the joint integrability of $E_x^{u,s}$, see Fig. 3. Let $x \in M$ and $\varepsilon > 0$. Take a point $y \in W_\varepsilon^u(x)$ and a point $z \in W_\varepsilon^s(x)$, both at distance ε from x. Consider

$$y' = W_\delta^s(y) \cap W_\delta^{wu}(z) \quad \text{and} \quad z' = W_\delta^{ws}(y) \cap W_\delta^u(z)$$

If $\delta > 2\varepsilon$, but small enough, each of the above intersections consists of one point. One can see that the points y', z' belong in the same orbit, i.e. $z' = \varphi^\tau y'$ for some small τ. Clearly, for jointly integrable spaces E_x^u, E_x^s we have $\tau \equiv 0$ for all x and $\varepsilon > 0$. As Plante showed for Anosov flows [60], if only $\tau \equiv 0$ in a small open set and all small $\varepsilon > 0$, then the flow φ^t is a cycle. On the other hand, if $\tau \neq 0$ for at least one x, then φ^t is topological mixing. Thus, even local joint integrability of E_x^u, E_x^s is equivalent to the cycle property for Anosov flows (but not for Axiom A flows, see below).

The joint nonintegrability of stable and unstable subspaces for a geodesic flow on a manifold M_0 of negative sectional curvature can now be verified as follows. It is known that each unstable manifold W^u is a smooth strictly convex hypersurface in M_0 equipped with outward unit normal vectors. Likewise, each stable manifold W^s is a smooth strictly convex hypersurface in M_0 equipped with inward unit normal vectors. By using the convexity, one easily observes that $\tau \neq 0$ in a neighborhood of any point. It is interesting that a similar convexity argument works for some mechanical systems with hyperbolic behavior, in particular for dispersing billiards as was noted by Sinai.

Central Limit Theorem (CLT) for our flows says that for every Gibbs measure μ and every Hölder continuous function $q(x)$ on M

$$\lim_{T\to\infty} \mu \left\{ \frac{\int_0^T q(\varphi^t x)\, dt - T\mu(q)}{\sqrt{T}} < z \right\} = \frac{1}{\sqrt{2\pi}\sigma_q} \int_{-\infty}^z e^{-\frac{x^2}{2\sigma_q^2}}\, dx \tag{4.9}$$

for all $-\infty < z < \infty$ and some constant $\sigma_q \geq 0$.

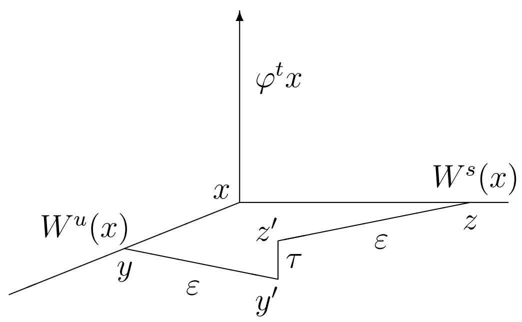

Figure 3: Joint nonintegrability of stable and unstable foliations.

It is interesting that the CLT holds for all hyperbolic flows, including cycles, it is even easier to prove in the case of cycles. Let φ^t be a cycle and $\mathcal{S}$ a Markov section with a constant return function $l(x) \equiv c$. Then (4.9) easily reduces to the CLT for the Gibbs measure ρ_g on $\mathcal{S}$ corresponding to μ and the Hölder continuous function $\int_0^c q(\varphi^t x)\, dt$ on $\mathcal{S}$. Note that the Gibbs measure μ is not even mixing in this case, still the CLT holds! Strangely enough, for mixing flows the proof of the CLT is somewhat more complicated. We refer the reader to [64].

Decay of correlations. Let φ^t be an Anosov or Axiom A flow and μ a Gibbs measure. The *correlation function* for a Hölder continuous function $q(x)$ is defined by

$$C_q(t) = \mu(q \cdot (q \circ \varphi^t)) - [\mu(q)]^2 \tag{4.10}$$

It is standard that $C_q(t) \to 0$ as $t \to \infty$ for all $q(x) \in L^2(\mu)$ if and only if the measure μ is mixing, or equivalently the flow φ^t is topologically mixing. If φ^t is a cycle, then $C_q(t)$ is easily seen to be a periodic function in t. In 1975 Bowen and Ruelle [13, 68] raised a question:

What is the speed of convergence of $C_q(t)$ to zero for mixing Axiom A flows? Is it exponential in t?

A negative answer was given by Ruelle [70] and strengthened by Pollicott [62] who found mixing Axiom A flows with correlations decaying arbitrarily slowly. Their example is quite instructive, we describe it next.

Consider the baker's transformation $T(x,y)$ on the unit square $[0,1]\times[0,1]$ defined by $T(x,y) = (2x, y/2)$ for $x \le 1/2$ and $T(x,y) = (2x-1, (y+1)/2)$ for $x > 1/2$. It has a finite Markov partition $\mathcal{R} = \{R_1, R_2\}$ with $R_1 = [0,1/2] \times [0,1]$ and $R_2 = [1/2,1] \times [0,1]$. Since T is only discontinuous on $\partial R_1 \cup \partial R_2$, the map T has all the essential properties of Anosov and Axiom A diffeomorphisms (the situation here is very similar to that for the return map on a Markov section of an Anosov flow, as we explained after Theorem 4.2). Now consider the ceiling function $l(x) \equiv 1$ on R_1 and $l(x) \equiv \alpha$ on R_2. Of course, for any rational α the suspension flow is a cycle. Pollicott showed that if α is irrational but allows good approximations by rationals, then the correlation function $C_q(t)$ will decrease as slowly as one wishes.

Note that the ceiling function $l(x)$ in the above example is locally constant, so locally (but not globally) the spaces E_x^u, E_x^s are jointly integrable.

Hence, one can have local joint integrability for topologically mixing Axiom A flows (this is impossible for Anosov flows, as we said above).

Recall that local joint integrability of E^u_x, E^s_x is characterized by $\tau = 0$, where τ is introduced in the construction illustrated on Fig. 3. It is unknown if Axiom A or Anosov flows with $\tau \neq 0$ have exponential decay of correlations. At least, by requiring a certain lower bound on τ/ε^2 (again in terms of the construction shown on Fig. 3), one can ensure a rapid decay of correlations:

Theorem 4.12 (Chernov [17]) *Let $\varphi^t : M^3 \to M^3$ be an Anosov flow and $|\tau| \geq c\varepsilon^2$ where $c > 0$ is a constant, then $|C_q(t)| \leq Ce^{-at^{1/2}}$ with some constants $C > 0$ and $a > 0$ for the SRB measure $\mu = \mu^+$.*

Alternatively, by requiring an extra smoothness of the flow φ^t and the spaces $E^{u,s}_x$ (as functions of x) one can obtain even an exponential decay of correlations:

Theorem 4.13 (Dolgopyat [26]) *Let $\varphi^t : M \to M$ be a $C^{2+\varepsilon}$ mixing Anosov flow, for which the spaces E^u_x, E^s_x are C^1 functions of x. Then $|C_q(t)| \leq Ce^{-at}$ with some constants $C > 0$ and $a > 0$ for the SRB measure $\mu = \mu^+$.*

Remark 4.7. The last theorem covers geodesic flows on compact surfaces of negative curvature, because the spaces $E^{u,s}_x$ are known to be C^1 for such geodesic flows [40].

For generic Anosov flows, not much is known, but we have the following mild bound on correlations:

Theorem 4.14 (Dolgopyat [26]) *Let $\varphi^t : M \to M$ be a mixing Anosov flow. Then $|C_q(t)|$ decays rapidly in the sense of Schwartz. In particular, it decays faster than any power function t^{-a}, $a > 0$. This is true for any Gibbs measure μ.*

More discussions of this issue can be found in the survey [63]. It is still unknown if $|C_q(t)|$ decays exponentially for any mixing Anosov flow. It remains yet to explore the speed of the decay of correlations for generic Anosov and Axiom A flows.

5 Volume compression

Here we concentrate on hyperbolic diffeomorphisms that do not preserve absolutely continuous measures.

Let $T : M \to M$ be an Anosov or Axiom A diffeomorphism. For $x \in M$ and $n \geq 1$ the map T^n compresses the phase volume (locally) at x at the exponential rate

$$R_n(x) = -\ln |DT^n(x)| = -\sum_{i=0}^{n-1} \ln |DT(T^i x)|$$

Consider again the functions $\varphi^u(x) = -\ln J^u_x$ and $\varphi^s(x) = \ln J^s_x$, where J^u_x and J^s_x are the Jacobians of the linear maps $DT : E^u_x \to E^u_{Tx}$ and $DT : E^s_x \to E^s_{Tx}$, respectively. Note that $\varphi^{u,s} < 0$ in the Lyapunov metric on M. In addition, let α_x be the angle between E^u_x and E^s_x in the tangent space $\mathcal{T}_x M$, and put $u(x) = -\ln \sin \alpha_x$. Then

$$-\ln |DT(x)| = \varphi^u(x) - \varphi^s(x) + u(Tx) - u(x)$$

Hence,

$$R_n(x) = S_n\varphi^u(x) - S_n\varphi^s(x) + u(T^n x) - u(x) \tag{5.1}$$

Let δ_x be the δ-measure concentrated at the point x. Suppose that the sequence of measures $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i x}$ weakly converges to a T-invariant measure μ. Then we can define the *volume compression rate per unit time* by

$$\bar{R}(x) := \lim_{n\to\infty} \frac{1}{n} R_n(x) = \int [\varphi^u - \varphi^s]\, d\mu = -\int \ln |DT|\, d\mu$$

We now recall Theorem 3.14 and obtain

Proposition 5.1 *Let Ω be an Axiom A attractor. Then for almost every point x in the basin of attraction (with respect to the Riemannian volume) we have*

$$\bar{R}(x) = \bar{R} := \int_\Omega [\varphi^u - \varphi^s]\, d\mu^+$$

where μ^+ is the SRB measure on Ω. The same holds for almost every point $x \in M$ if T is a transitive Anosov diffeomorphism (in which case, of course, $\Omega = M$).

We call the above quantity $\bar{R}$ the *volume compression rate.*

Theorem 5.2 *Let Ω be an attractor. Then $\bar{R} \geq 0$. The following conditions are equivalent:*
(i) $\bar{R} = 0$
(ii) Ω *is an open set and the SRB measure μ^+ is absolutely continuous on Ω.*

Proof. Let $\mathcal{M}(\Omega)$ be the space of T-invariant probability measures on Ω. The variational principle says that

$$P_{\varphi^s} = \sup_{\mu \in \mathcal{M}(\Omega)} \left(h_\mu(T) + \int_\Omega \varphi^s \, d\mu \right) \tag{5.2}$$

and the supremum is attained on a unique measure, the Gibbs measure $\mu^- := \mu_{\varphi^s}$. We also have $P_{\varphi^s} \leq 0$ by Lemma 3.7. (Note that the function φ^s plays the same role for T^{-1} as the function φ^u for T, so Lemma 3.7 applies.) On the other hand, we know that

$$0 = P_{\varphi^u} = h_{\mu^+}(T) + \int_\Omega \varphi^u \, d\mu^+ \tag{5.3}$$

Substituting $\mu = \mu^+$ in (5.2) and subtracting (5.3) from (5.2) gives $\bar{R} \geq 0$. The case $\bar{R} = 0$ is only possible when $P_{\varphi^s} = 0$ and $\mu^+ = \mu^-$, i.e. the functions φ^u and φ^s are cohomologous. The fact $P_{\varphi^s} = 0$ implies that Ω is an attractor for T^{-1}, i.e. Ω is a union of stable manifolds (by Theorem 3.13 and Lemma 3.11 applied to T^{-1}). Since Ω is also a union of unstable manifolds by the same Lemma 3.11, it must be an open set. Next, the same argument as in the proof of Theorem 3.18 gives the absolute continuity of μ^+. One can also deduce the absolute continuity of μ^+ combining its absolute continuity on unstable manifolds and that on stable manifolds (the latter follows from the fact $\mu^+ = \mu^-$, but we omit details). □

Remark 5.1. Note that

$$-\bar{R} = \sum \lambda_i \qquad\qquad \mu^+\text{−a.e.}$$

the sum of all Lyapunov exponents, which is constant μ^+-almost everywhere. This follows from Lemma 3.16 and its counterpart for φ^s and negative exponents.

Entropy production. The volume compression rate $\bar{R}$ is closely related to entropy production in statistical mechanics. The latter is defined as follows [71]. Let ν be an absolutely continuous probability measure on the basin of attraction of an attractor Ω, with density $g(x)$. The *statistical mechanics entropy*, or *Gibbs entropy*, associated with $g(x)$ is

$$S(g) = -\int g(x) \ln g(x)\, dx = \nu(-\ln g)$$

The measure $T_*\nu$ has density $g_1(x) = g(T^{-1}x)|DT(T^{-1}x)|^{-1}$. Hence,

$$\begin{aligned} S(g_1) &= -\int g(T^{-1}x)|DT(T^{-1}x)|^{-1} \Big(\log g(T^{-1}x) - \ln|DT(T^{-1}x)| \Big)\, dx \\ &= -\int g(x) \Big(\log g(x) - \ln|DT(x)| \Big)\, dx \end{aligned}$$

Hence, the Gibbs entropy of the system decreases by the amount

$$S(g) - S(g_1) = -\int g(x) \ln|DT(x)|\, dx = \nu(-\ln|DT|)$$

In physical terms, this means that the system produces this much entropy and gives it to the outside world (or the entropy is pumped out of the system).

The entropy produced by the system during n iterations of T is then

$$S(g) - S(g_n) = -\int g(x) \ln|DT^n(x)|\, dx = \sum_{i=0}^{n-1} \nu_i(-\ln|DT|)$$

where $\nu_n = T_*^n \nu$ and $g_n(x)$ is the density of the measure ν_n.

We now combine the above equation with Proposition 5.1 and Corollary 3.15.

Corollary 5.3 *The asymptotic entropy production rate per unit time is*

$$\lim_{n\to\infty} \frac{1}{n}(S(g) - S(g_n)) = \bar{R}$$

where $\bar{R}$ is our volume compression rate.

We can interpret the results of Corollary 5.3 and Theorem 5.2 as follows. We see that Anosov diffeomorphisms and Axiom A attractors with absolutely continuous invariant measures do not produce Gibbs entropy ($\bar{R} = 0$).

Recall that they serve as models of equilibrium dynamics in statistical mechanics. On the contrary, Anosov diffeomorphisms and Axiom A attractors without absolutely continuous invariant measures (whose SRB measure is singular) do produce Gibbs entropy ($\bar{R} > 0$). These serve as models of nonequilibrium dynamics in statistical mechanics. Therefore, positive entropy production characterizes nonequilibrium dynamics, as opposed to zero entropy production in equilibrium systems. This is a basic fact known in statistical mechanics [72].

Fluctuation theorem. The value of $\bar{R}$ gives the asymptotic volume compression rate. It is of interest, particularly in numerical experiments, to study finite time compression rates $\bar{R}_n(x) := \frac{1}{n}R_n(x)$. In view of (5.1), for large n we have

$$\bar{R}_n(x) = \frac{S_n\varphi^u(x)}{n} - \frac{S_n\varphi^s(x)}{n} + O(1/n)$$

Obviously, $\mu^+(\bar{R}_n(x)) = \bar{R}$ for any n. The deviations of $\bar{R}_n(x)$ from its mean value $\bar{R}$ are described by the central limit theorem and the large deviation theorem, see Section 2. There is a remarkable specification of the large deviation theorem for the quantity $\bar{R}_n(x)$ under an additional physically motivated assumption.

Let $T : M \to M$ be a transitive Anosov diffeomorphism. Assume that there is a smooth involution $\mathcal{I} : M \to M$ (i.e. $\mathcal{I}^2$ =identity) such that $\mathcal{I} \circ T = T^{-1} \circ \mathcal{I}$, i.e. $\mathcal{I}$ anticommutes with T. Applying $\mathcal{I}$ then amounts to conjugating T to T^{-1}, i.e. reversing the time. We will call such Anosov diffeomorphisms *time reversible*. We can also choose a Riemannian metric on M invariant under $\mathcal{I}$.

As a simple example, let $T : M \to M$ be an Anosov diffeomorphism, then the map $(x, y) \mapsto (Tx, T^{-1}y)$ is an Anosov diffeomorphism of the manifold $M \times M$. Then the involution $\mathcal{I}(x, y) = (y, x)$ will anticommute with this map.

In Hamiltonian mechanics, a natural involution on the phase space is defined by reversing the velocity vectors of all moving particles in the system. Then each particle will trace its past trajectory backwards under the same equations of motion, so the involution will anticommute with the dynamics. This property is called *time reversibility* in mechanics. Many physical models, both at equilibrium and nonequilibrium, naturally have the time reversibility feature, so the interest to time reversible Anosov diffeomorphism is well justified from physics point of view.

Theorem 5.4 (Gallavotti-Cohen Flutuation Theorem [32, 72]) *Let* $T : M \to M$ *be a transitive time reversible Anosov diffeomorphism. Assume that its SRB measure* μ^+ *is singular (hence* $\bar{R} > 0$*). Then there is a* $p^* > 0$ *such that*

$$p - \delta \leq \lim_{n\to\infty} \frac{1}{\bar{R}n} \ln \frac{\mu^+\{\bar{R}_n(x)/\bar{R} \in (p-\delta, p+\delta)\}}{\mu^+\{\bar{R}_n(x)/\bar{R} \in (-p-\delta, -p+\delta)\}} \leq p + \delta$$

for all p*,* $|p| \leq p^*$*, and* $\delta > 0$*.*

A loose interpretation of the theorem is the following. For $p > 0$, the probability of observing the value $\bar{R}_n \approx p\bar{R}$ is approximately $e^{p\bar{R}n}$ times higher than the probability of observing the value $\bar{R}_n \approx -p\bar{R}$. Let $f_n(p)$ be the probability density function of the variable $\bar{R}_n(x)/\bar{R}$, i.e. let

$$\mu^+\{p < \bar{R}_n(x)/\bar{R} < p + \delta\} = f_n(p)\delta + o(\delta)$$

The theorem can be, again loosely, stated as

$$f_n(p)/f_n(-p) \approx e^{p\bar{R}n} \tag{5.4}$$

This theorem can be verified by a numerical experiment as follows. One fixes n (e.g., $n = 50$ or $n = 100$, not too large) and picks a random point $x \in M$ (with respect to the Lebesgue measure on M) and computes its orbit $\{T^i x\}$, $0 \leq i \leq N$ for a very large $N \gg n$. One also computes the values $\bar{R}_n(i) := \bar{R}_n(T^{im}x)$ for some $m \geq n$ and $0 \leq i \leq J := [N/m] - 1$. This gives the values of the time n volume compression rate along $J + 1$ nonoverlapping segments of length n on the orbit $\{T^i x\}$. Note that time averages along the orbit $\{T^i x\}_{i=0}^N$ are very close to the space averages with respect to μ^+, if N is large, by Theorem 3.14. So we can estimate $\bar{R} = \mu^+(\bar{R}_n)$ by $[\bar{R}_n(0) + \cdots + \bar{R}_n(J)]/(J + 1)$. Then one can histogram the values of $\bar{R}_n(i)/\bar{R}$ to approximate the function $f_n(p)$ defined above. To verify (5.4) one fits a linear function $y = ap + b$ to the experimental values of the function $[\ln f_n(p) - \ln f_n(-p)]/\bar{R}n$ and checks that $a \approx 1$ and $b \approx 0$.

Such numerical experiments can be effectively done for many nonequilibrium mechanical models with time reversibility, which are far more general than Anosov diffeomorphisms. For many such systems a rigorous proof of the above theorem is lacking, even the existence of SRB measures or mere hyperbolicity cannot be rigorously established. On the other hand, a positive result of the test (i.e., if $a \approx 1$ and $b \approx 0$) can be then interpreted as a certain similarity of the underlying mechanical system and Anosov diffeomorphisms with a singular SRB measure. Such a similarity plays a crucial role in the modern understanding of nonequilibrium statistical mechanics, see the so called Axiom C and more extended discussions of the topic in [31, 72].

We provide a formal, heuristic argument supporting Theorem 5.4. A rigorous proof is based on the large deviation theorem 2.32, we refer the reader to [32, 72] for details.

Lemma 5.5 *The involution $\mathcal{I}$ maps unstable manifolds to stable manifolds and vice versa. As a result,*

$$S_n\varphi^u(x) = S_n\varphi^s(\mathcal{I}\circ T^n x) + O(1) \quad \text{and} \quad S_n\varphi^s(x) = S_n\varphi^u(\mathcal{I}\circ T^n x) + O(1)$$

and

$$\bar{R}_n(x) = -\bar{R}_n(\mathcal{I}\circ T^n x) + O(1/n)$$

for any $x \in M$. The terms $O(1)$ can be dropped if the Riemannian metric is chosen invariant under $\mathcal{I}$.

Lemma 5.6 *Let $\mathcal{R}_0$ be a Markov partition of M. Then $\mathcal{R} := \mathcal{R}_0 \vee \mathcal{I}\mathcal{R}_0$ is a Markov partition invariant under $\mathcal{I}$, i.e. for any $R \in \mathcal{R}$ we have $\mathcal{I}R \in \mathcal{R}$.*

The proofs of the lemmas are left as (easy) exercises.

Now, let $\mathcal{R}$ be a Markov partition from the last lemma. Let $\Lambda = [0, n]$ and $C(\omega_\Lambda) \subset \Sigma_A$ any admissible cylinder of length $n+1$ in the symbolic space Σ_A. The set $\pi(C(\omega_\Lambda))$ is a rectangle in M. The invariance of $\mathcal{R}$ under $\mathcal{I}$ implies that $\mathcal{I}\circ T^n \circ \pi(C(\omega_\Lambda)) = \pi(C(\omega'_\Lambda))$ for some other cylinder $C(\omega'_\Lambda) \subset \Sigma_A$ of length n. Recall that $|\bar{R}_n(x) - \bar{R}_n(y)| = O(1/n)$ for all $x, y \in \pi(C(\omega_\Lambda))$, and $|\bar{R}_n(x') - \bar{R}_n(y')| = O(1/n)$ for all $x', y' \in \pi(C(\omega'_\Lambda))$, see Remark 2.7. Then $\bar{R}_n(x) = -\bar{R}_n(x') + O(1/n)$ for all x, x' as above. Now, recall that

$$\mu^+(\pi(C(\omega_\Lambda))) \asymp \exp[S_n\varphi^u(x)]$$

and

$$\mu^+(\pi(C(\omega'_\Lambda))) \asymp \exp[S_n\varphi^u(x')] \asymp \exp[S_n\varphi^s(x)]$$

Hence,

$$\frac{\mu^+(\pi(C(\omega_\Lambda)))}{\mu^+(\pi(C(\omega'_\Lambda)))} \asymp \exp[S_n\varphi^u(x) - S_n\varphi^s(x)] \asymp \exp[n\bar{R}_n(x)] \qquad (5.5)$$

Now consider $C(\omega_\Lambda) \subset \Sigma_A$ such that $\bar{R}_n(x)/\bar{R} \in (p-\delta, p+\delta)$ for some $x \in \pi(C(\omega_\Lambda))$, and hence, $\bar{R}_n(x')/\bar{R} \in (-p-\delta-O(1/n), p+\delta+O(1/n))$ for any $x' \in \pi(C(\omega'_\Lambda))$. Summing up (5.5) over all such $C(\omega_\Lambda)$ completes our proof of the theorem. □

In terms of the large deviation theorem 2.32, the above result means that

$$\eta(p) + \eta(-p) = \bar{R}p$$

i.e. the odd part of the free energy function $\eta(p)$ is linear in p, with slope $\bar{R}$.

Theorem 5.4 is a specification of the large deviation theorem rather than the central limit theorem. One may mistakenly relate 5.4 to the CLT in the following way. By the CLT, $\sqrt{n}(\bar{R}_n - \bar{R})$ converges in distribution to a normal law $N(0, \sigma_0^2)$ with some $\sigma_0^2 \geq 0$. Hence, $\bar{R}_n/\bar{R}$ is approximately normal $N(1, \sigma_0^2/\bar{R}^2 n)$. Assuming the closeness of the corresponding densities, one can write

$$f_n(p) \approx \frac{\bar{R}\sqrt{n}}{\sqrt{2}\sigma_0} \exp\left[-\frac{(p-1)^2}{2\sigma_0^2}\bar{R}^2 n\right] \qquad (5.6)$$

Assuming now that (5.6) is exact, one obtains

$$f_n(p)/f_n(-p) = \exp[2\sigma_0^{-2}\bar{R}^2 np]$$

and Theorem 5.4 seems to follow provided we have $2\sigma_0^{-2}\bar{R} = 1$.

The above argument is, in fact, faulty. The CLT only applies to the values $\bar{R}_n = \bar{R} + O(1/\sqrt{n})$, i.e. $p = 1 + O(1/\sqrt{n})$. Theorem 5.4, on the contrary, is a large deviation result, it applies to all $|p| \leq p^*$, i.e. to all $|\bar{R}_n| \leq p^*\bar{R}$. It is logically unrelated to the CLT, in particular, it does not require that $2\sigma_0^{-2}\bar{R} = 1$.

6 SRB measures for general diffeomorphisms

Remarkably, the theory of SRB measures for Axiom A attractors, at least in its essential parts, extends to much more general diffeomorphisms. Here we describe these extensions, leaving out proofs, most of which are highly nontrivial. Our discussion is based on recent excellent survey articles by Ruelle [72] and Young [80].

Let $T : M \to M$ be a C^2 diffeomorphism of a smooth compact manifold M, and ρ a T-invariant ergodic measure on M. No assumptions on the hyperbolicity of T are necessary at this point.

The theorem of Oceledec [54] permits the definition of Lyapunov exponents $\lambda_i(x)$, $1 \le i \le \dim M$, at ρ-almost every point $x \in M$. The set of Lyapunov exponents $\{\lambda_i(x)\}$ (called sometimes the *Lyapunov spectrum*) is the same ρ-almost everywhere, due to the ergodicity of ρ. The tangent vectors with positive Lyapunov exponents span the *unstable subspace* E^u_x, those with negative Lyapunov exponents span the *stable subspace* E^s_x, and those with zero Lyapunov exponents span the *neutral subspace* E^0_x. The decomposition $\mathcal{T}_x M = E^u_x \oplus E^s_x \oplus E^0_x$ is defined ρ-almost everywhere and stays invariant under DT.

Pesin's theory [56, 57] allows the definition of stable and unstable manifolds ρ-almost everywhere. These are smooth submanifolds of M, the stable manifold $W^s(x)$ through x is tangent to E^s_x and the unstable manifold $W^u(x)$ through x is tangent to E^u_x.

We now make the first hyperbolicity assumption on T: let at least one Lyapunov exponent of T be positive ρ-almost everywhere. Then unstable manifolds $W^u(x)$ are not trivial.

Definition. An invariant ergodic measure ρ is called a *Sinai-Ruelle-Bowen (SRB) measure* if its conditional distributions on unstable manifolds are absolutely continuous with respect to the Lebesgue measure on those manifolds.

SRB measures have a remarkable property that generalizes Theorem 3.17. The following theorem was proved by Ledrappier, Strelcyn and Young:

Theorem 6.1 ([48, 50]) *An ergodic measure ρ is SRB if and only if*

$$h_\rho(T) = \Sigma^+ \lambda_i \qquad \rho\text{-a.e.} \tag{6.1}$$

where the right hand side is the sum of all positive Lyapunov exponents (note that it is ρ-almost everywhere constant).

The following inequality nicely complementing Theorem 6.1 was proved by Ruelle.

Theorem 6.2 ([69]) *For an arbitrary ergodic measure* ρ*, we only have*

$$h_\rho(T) \le \Sigma^+ \lambda_i \qquad \rho\text{-a.e.} \tag{6.2}$$

The identity (6.1) for SRB measures is known as *Pesin's formula*, while (6.2) for general ergodic measures is known as *Ruelle's inequality*.

Next, we make a stronger hyperbolicity assumption on T: let all its Lyapunov exponents be nonzero ρ-almost everywhere.

Definition. A diffeomorphism T is said to be *fully hyperbolic* with respect to an invariant measure ρ if all the Lyapunov exponents of T are nonzero almost everywhere.

The full hyperbolicity does not require uniform expansion of unstable vectors or uniform contraction of stable vectors (in the way it is required for Anosov and Axiom A systems). For this reason, full hyperbolicity is often referred to as *nonuniform hyperbolicity*.

Theorem 6.3 ([61]) *Let* T *be fully hyperbolic. If* ρ *is an SRB measure, then there is a subset* $B \subset M$ *of positive Lebesgue measure such that for each* $x \in B$ *the measure*

$$\delta_{x,n} := \frac{1}{n}(\delta_x + \delta_{Tx} + \cdots + \delta_{T^{n-1}x})$$

weakly converges to ρ *(here* δ_x *is the* δ*-measure concentrated at* x*). In this case,* x *is called a* ρ*-generic point. Equivalently, for any continuous function* $f : M \to \mathbb{R}$ *and every* $x \in B$ *we have*

$$\lim_{n\to\infty} \frac{f(x) + f(Tx) + \cdots + f(T^{n-1}x)}{n} = \int f\, d\rho$$

One often calls the set B the *basin of attraction* for the measure ρ.

Remark 6.1. If an ergodic measure ρ is absolutely continuous on M with density $g(x)$, then it is an SRB measure [57], and its basin of attraction contains the open set $\{g(x) > 0\}$ (possibly, minus a subset of Lebesgue measure 0).

According to our criteria in Section 3, absolutely continuous ergodic measures play the role of equilibrium states in statistical physics. If an SRB measure ρ is singular (i.e. there is a subset $A \subset M$ of zero Lebesgue measure such that $\rho(A) = 1$), then ρ plays the role of a nonequilibrium steady state.

The previous theorem easily implies that there are at most countably many distinct SRB measures. This is further specified by the following analogue of Smale's spectral decomposition:

Theorem 6.4 ([49, 57]) *Let $T : M \to M$ be a nonuniformly hyperbolic diffeomorphism. There are at most countably many distinct SRB measures. Let ρ be any SRB measure. Then either ρ is mixing and Bernoulli, or there is a decomposition of its support into finitely many disjoint measurable subsets $X_1, \ldots, X_n$ cyclically permuted by T, such that $T^n|X_i$ is mixing and Bernoulli for each $i = 1, \ldots, n$.*

Existence of SRB measures. The importance of SRB measures, especially in physical applications, is clear from our previous discussion. Theorem 6.4 establishes an upper bound on the number of SRB measures. However, nothing guarantees the shear existence of such measures. In fact, we only know that SRB measures exist for Anosov and Axiom A diffeomorphisms. For more general types of diffeomorphisms, very little is known. Just a few years ago, Benedicks and Young proved [9] the existence of SRB measures for Hénon attractors (and those are, probably, the only genuinely nonuniformly hyperbolic diffeomorphisms for which rigorous results exist). On the other hand, attractors without SRB measures have been observed by Hu and Young also recently, see [42]. Aside from these specific examples, there are no general mathematical theorems on the existence or nonexistence of SRB measures for nonuniformly hyperbolic diffeomorphisms.

There are hyperbolic maps other than diffeomorphisms for which SRB measures are rigorously constructed [58, 79], however. We only mention them briefly. Let T be defined on an open subset $U \subset M$ and $T : U \to T(U)$ a C^2 diffeomorphism. The set ∂U may be regarded as the singularity set for T. Now, assume a uniform hyperbolicity in the sense

$$||DT^n v|| \le c\lambda^n ||v|| \quad \text{for } v \in E^s_x, \ \ n \ge 0$$

$$||DT^n v|| \ge c^{-1}\lambda^{-n} ||v|| \quad \text{for } v \in E^u_x, \ \ n \ge 0$$

uniformly in x whenever T^n exists. Here $c > 0$ and $\lambda \in (0, 1)$ are constants. Many physically interesting models, e.g. billiards and Lorenz attractors, fit in this category, see [79, 80] for detail. For such models, SRB measures were constructed by Pesin [58] and under some extra assumptions statistical properties were studied by Young [79].

Numerical evidence is also overwhelmingly in favor of the existence of SRB measures in many hyperbolic models, whether smooth or containing singularities. For randomly chosen initial points in the basin of attraction, trajectory plots tend to produce pictures that are very much alike – the data points concentrate nicely along unstable manifolds and their distribution in the stable direction is very irregular (singular). Based on these observations and recent theoretical studies, it is currently assumed in statistical mechanics that nonequilibrium steady states in hyperbolic models typically are SRB measures. At the very least, they can be treated as SRB measures in Anosov and Axiom A attractors, for the purpose of computing time and space averages [31, 72].

Dimension of invariant measures. To better understand the difference between SRB measures and other ergodic measures, let us look at Pesin's formula (6.1) and Ruelle's inequality (6.2) as follows. It is our fundamental understanding that randomness (chaos) is created by the separation of nearby orbits. The asymptotic separation of nearby orbits is quantified by the sum of positive Lyapunov exponents. On the other hand, the *randomness* (in the sense of information) of a measure preserving transformation (M, T, ρ) is quantified by its entropy $h_\rho(T)$. Pesin's identity (6.1) clearly establishes the balance between the 'input' ("amount" of separation) and the 'output' ("amount" of randomness). Ruelle's inequality suggests that for non-SRB measures the separation is "wasted" on something other than randomness, i.e. there is a "leakage" of separated trajectories from the support of the invariant measure, so that the measure does not account for all trajectories. The separation occurs along unstable manifolds, so the leakage of separated trajectories can only happen if the invariant measure has "holes" on those manifolds, through which some trajectories leak out unaccounted.

This is exactly what happens to non-SRB measures, their conditional distributions on unstable manifolds are singular, so they only capture a fraction of those manifolds and leave out holes of some kind. To quantify the fraction of an unstable manifold captured by a measure we use the notion of *dimension*.

Definition. Let $B^u_\varepsilon(x)$ be a ball of radius ε on the unstable manifold $W^u(x)$, centered at x. If ρ_x^u is the conditional measure induced by ρ on $W^u(x)$, then we call the limit

$$d_x^u(\rho) = \lim_{\varepsilon\to 0} \frac{\log \rho_x^u(B_\varepsilon^u(x))}{\log \varepsilon}$$

(if one exists) the *dimension* of the measure ρ on the unstable manifold $W^u(x)$ (at the point x). This limit exists ρ^u-almost everywhere on $W^u(x)$. If ρ is ergodic, the limit is constant ρ-almost everywhere on M, and we denote it by $d^u(\rho)$.

We note that the above limit is the infimum of Hausdorff dimensions of subsets of full ρ_x^u-measure on $W^u(x)$. If the conditional measure is absolutely continuous on $W^u(x)$, then $d^u(\rho) = \dim W^u(x)$, otherwise $d^u(\rho) < \dim W^u(x)$. The deficiency of $d^u(\rho)$ signifies the presence of "holes" on $W^u(x)$, not filled properly by the measure ρ_x^u.

An exact relation between the dimension and the entropy was established by Ledrappier and Young:

Theorem 6.5 ([50]) *Let ρ be an ergodic measure for a nonuniformly hyperbolic diffeomorphism. Then, corresponding to each Lyapunov exponent $\lambda_i > 0$, there is a number $\sigma_i \in [0,1]$ such that*
(a) $d^u(\rho) = \Sigma^+\sigma_i$;
(b) $h_\rho(T) = \Sigma^+\lambda_i\sigma_i$.

The number σ_i is essentially the dimension of ρ in the direction of the unstable tangent vector corresponding to the Lyapunov exponent $\lambda_i > 0$.

Remark 6.2. It is interesting that one can also define $d^s(\rho)$, the dimension of the conditional measures induced by ρ on stable manifolds. Then one would expect that

$$d^u(\rho) + d^s(\rho) = d(\rho) \tag{6.3}$$

where $d(\rho)$ is the dimension of ρ on M defined in a similar way:

$$d_x(\rho) = \lim_{\varepsilon\to 0} \frac{\log \rho(B_\varepsilon(x))}{\log \varepsilon}$$

The identity (6.3) has been known as the Eckmann-Ruelle conjecture for many years. It was proved very recently by Barreira, Pesin and Schmeling [8].

The leakage from the system can also be described in a different way. Let Ω be a compact T-invariant subset of M of zero Lebesgue measure, and we will study invariant measures on Ω. Let U be a small open neighborhood of Ω. For $m \geq 0$ define

$$U_n = \cap_{i=0}^n T^i U \qquad \text{and} \qquad U_{-n} = \cap_{i=0}^n T^{-i} U$$

If Ω is not an attractor, typical points $x \in U$ will sooner or later escape from U, i.e. generally, mass will leak out of U. In that case U_{-n} shrinks as $n \to \infty$ and its volume approaches zero. The limit

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \left(\mathrm{Vol}\, U_{-n}\right)^{-1} \geq 0 \tag{6.4}$$

(if one exists) is called the *escape rate.* The larger γ, the faster the mass leaks out of U.

Theorem 6.6 ([71]) *Let Ω be an Axiom A basic set. Then*
(a) *The limit (6.4) exists and*

$$\gamma = -P_{\varphi^u} \geq 0$$

where P_{φ^u} is the topological pressure of the function $\varphi^u = -\ln J_x^u$ defined in Section 3.
(b) *Let m_{-n} be the normalized Lebesgue measure on U_{-n}. Then the measure*

$$\frac{1}{n+1}\Big(m_{-n} + T_* m_{-n} + \cdots + T_*^n m_{-n}\Big)$$

weakly converges to the generalized SRB measure $\mu^+ = \mu_{\varphi^u}$ on the basic set Ω.

This is proved by Ruelle [71] and can be also derived from our results in Section 3, we omit details.

Recall that for any ergodic measure ρ on Ω we have

$$\rho(\varphi^u) = -\Sigma^+ \lambda_i \qquad \rho\text{-a.e.}$$

see Lemma 3.16. Now the variational principle (2.51) implies

Corollary 6.7 *For any ergodic measure ρ on Ω we have*

$$h_\rho(T) \le \Sigma^+\lambda_i - \gamma \qquad \rho\text{-a.e.}$$

and the equality occurs if and only if $\rho = \mu^+$, the generalized SRB measure on Ω. Hence,

$$h_{\mu^+}(T) = \Sigma^+\lambda_i - \gamma \qquad \mu^+\text{-a.e.} \tag{6.5}$$

Note that once again an SRB measure (this time a generalized SRB measure) is characterized by the equality in an entropy formula.

The identity (6.5) is known as the *escape rate formula*. It extends Pesin's formula from Axiom A attractors to arbitrary Axiom A basic sets. Note that it establishes an exact balance of randomness: it is the total amount of separation minus the amount spent on repelling trajectories away from Ω.

Theorem 6.6 also shows that the measure μ^+ describes the distribution of orbits of typical points $x \in U$ that stay in U for a long enough time (i.e., $x \in U_{-n}$ with a large n). This is a natural extension of the notion of nonequilibrium steady state to the present situation, where the mass is leaking out of the system. So, we will refer to μ^+ as a (generalized) nonequilibrium steady state on the set Ω.

If Ω is not an attractor, then almost every point $x \in U$ (with respect to the Lebesgue measure) eventually escapes from U. The basic set Ω only attracts the stable manifolds of its points and repels the rest of U. For this reason Ω is called a *semi-attractor* or a *repeller*, the latter term is now getting standard.

After the work of Gaspard and Nicolis [34], many physicists started extensive studies of chaotic systems with various openings or holes in phase space, and the corresponding repellers on which interesting invariant measures exist. This new branch of nonequilibrium statistical mechanics is now called *chaotic scattering theory.* We refer the reader to a survey by Gaspard and Dorfman [35] and references therein. The existence of (generalized) steady states on repellers and the escape rate formula (6.5) have been observed numerically and studied heuristically in a variety of models with nonuniform hyperbolicity. Very little has been proven mathematically, though.

We only mention one rigorous result obtained very recently [18, 19, 20, 21]. There, instead of taking a small open neighborhood of a repeller, the authors started with a completely hyperbolic (Anosov) diffeomorphism and cut some open holes in the manifold. The orbits that enter the holes are taken away from the system, thus simulating 'escape'.

Anosov diffeomorphisms with holes. Let $T : M \to M$ be a transitive Anosov diffeomorphism. Fix an open set $H \subset M$ and consider the dynamics on $U = M \setminus H$. For $n \geq 0$ put

$$U_n = \cap_{i=0}^n T^i U \quad \text{and} \quad U_{-n} = \cap_{i=0}^n T^{-i} U, \tag{6.6}$$

and also

$$U_+ = \cap_{n\geq 1} U_n, \quad U_- = \cap_{n\geq 1} U_{-n}, \quad \Omega = U_+ \cap U_- \tag{6.7}$$

Observe that all these sets are closed, $T^{-1}U_+ \subset U_+$, $TU_- \subset U_-$ and $T\Omega = T^{-1}\Omega = \Omega$. The set U_+ is a union of unstable manifolds, U_- is a union of stable manifolds, and Ω is a Cantor-like set of points. The trajectories of points $x \in \Omega$ (in the future and the past) always avoid H (the union of holes). Ω is also called a *repeller*.

Let m_{-n} be the normalized Lebesgue measure on U_{-n} and $m_n = T_*^n m_n$. Note that m_n can be obtained by iterating the Lebesgue measure on U, cutting out its fractions escaping through H and renormalizing the remainder.

Theorem 6.8 ([20, 21]) *Assume that* $\dim M = 2$ *and* H *satisfies some technical assumptions (see below). Then*
(a) *The measure*

$$\frac{1}{n+1}\Big(m_{-n} + T_* m_{-n} + \cdots + T_*^n m_{-n}\Big)$$

weakly converges to a T*-invariant measure* m^+ *on* Ω.
(b) *The measure* m^+ *is ergodic, K-mixing, and an equilibrium state for the function* $\varphi^u = -\ln J_x^u$ *on* Ω.
(c) *The escape rate (6.4) exists and the escape rate formula holds:*

$$h_{m^+}(T) = \Sigma^+ \lambda_i - \gamma \qquad m^+\text{-a.e.}$$

(d) *The measure* $m_n = T_*^n m_{-n}$ *weakly converges to a measure* $\tilde{m}^+$ *on* U_+ *which has two properties:* (i) *its conditional distributions on unstable manifolds* $W^u \subset U_+$ *are smooth (actually, they coincide with u-SRB measures defined in Section 3), and* (ii) *the measure* $\tilde{m}^+$ *is not* T*-invariant but satisfies*

$$(T_* \tilde{m}^+)(A) = e^{-\gamma} \cdot \tilde{m}^+(A) \qquad \text{for all} \quad A \subset U \tag{6.8}$$

i.e. the image of $\tilde{m}^+$ *restricted to* U *is proportional to* $\tilde{m}^+$*. Furthermore, a measure* $\tilde{m}^+$ *defined by the properties* (i) *and* (ii) *is unique.*
(d) *The measure* $T_*^{-n}\tilde{m}^+$ *weakly converges to* m^+ *as* $n \to \infty$*.*

The assumptions on H in this theorem are as follows. H is an open set consisting of a finite number, N, of connected components. The minimal distance between those components is some $d_0 > 0$. For any local unstable and stable manifold W of length $< d_0$ the intersection $W \cap H$ must consist of $\leq B$ intervals, where $B < \infty$ is some constant. The main assumption is that the connected components of H have diameter less than certain $d = d(T, N, d_0, B) > 0$.

The smallness of the components of H is necessary to prevent the decomposition of U into 'noninteracting' parts and hence the nonergodicity of the measure m^+ on Ω (such situations are described in [19]).

Remark 6.3. The measures m^+ and $\tilde{m}^+$ and the escape rate γ depend continuously on the open set H. In particular, if H shrinks to N isolated points, then $\gamma \to 0$ and both measures m^+, $\tilde{m}^+$ weakly converge to the SRB measure μ^+ of the Anosov diffeomorphism $T : M \to M$.

Remark 6.4. Since T^{-1} is also an Anosov diffeomorphism, one can construct a measure m^- on Ω and define the escape rate γ^- by using the map T^{-1} instead of T. Generally, $m^+ \neq m^-$ and $\gamma \neq \gamma^-$, see examples in [16]. It is proved in [21] that $m^+ = m^-$ and $\gamma = \gamma^-$ provided the Anosov diffeomorphism preserves an absolutely continuous measure. Again, physically, this corresponds to an equilibrium situation.

References

[1] L. M. Abramov, *On the entropy of a flow*, Amer. Math. Soc. Transl., Ser. 2, **49** (1966), 167–170.

[2] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds with negative curvature*, Proc. Steklov Inst. Math. **90** (1967).

[3] D. V. Anosov, *On tangent fields of transversal foliations in Y-systems*, Matem. Zametki **2** (1967), 539–548.

[4] D. V. Anosov and Ya. G. Sinai, *Some smooth ergodic systems*, Russ. Math. Surveys **22** (1967), 103–167.

[5] R. Adler and B. Weiss, *Entropy a complete metric invariant for automorphisms of the torus*, Proc. Nat. Acad. Sci. USA **57** (1967), 1573–1576.

[6] V. I. Arnold, *Some remarks on flows of line elements and frames*, Soviet Math. Dokl. **2** (1961), 562–564.

[7] V. I. Arnold, *Remarks on winding numbers*, Sibirsk. Mat. Z. (Russian) **2** (1961), 807–813.

[8] L. Barreira, Y. Pesin and J. Schmeling, *Dimension and product structure of hyperbolic measures*, Ann. of Math. **149** (1999), 755–783.

[9] M. Benedicks and L.-S. Young, *SBR measures for certain Henon maps*, Inventiones Math. **112** (1993), 541–576.

[10] R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–459.

[11] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes Math. **470**, Springer-Verlag, Berlin, 1975.

[12] R. Bowen, *Markov partitions are not smooth*, Proc. Amer. Math. Soc. **71** (1978), 130–132.

[13] R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Inventiones math. **29** (1975), 181–202.

[14] L. A. Bunimovich and Ya. G. Sinai, *Markov partitions for dispersed billiards*, Commun. Math. Phys. **73** (1980), 247–280.

[15] L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov, *Markov partitions for two-dimensional billiards*, Russ. Math. Surv. **45** (1990), 105–152.

[16] N.N. Čencova, *Statistical properties of smooth Smale horseshoes*, in: *Mathematical Problems of Statistical Mechanics and Dynamics*, R.L. Dobrushin, Editor, pp. 199–256, Reidel, Dordrecht, 1986.

[17] N. I. Chernov *Markov approximations and decay of correlations for Anosov flows*, Ann. Math. **147** (1998), 269–324.

[18] N. Chernov and R. Markarian, *Ergodic properties of Anosov maps with rectangular holes*, Bol. Soc. Bras. Mat. **28** (1997), 271–314.

[19] N. Chernov and R. Markarian, *Anosov maps with rectangular holes. Nonergodic cases*, Bol. Soc. Bras. Mat. **28** (1997), 315–342.

[20] N. Chernov, R. Markarian and S. Troubetzkoy, *Conditionally invariant measures for Anosov maps with small holes*, Ergod. Th. Dynam. Syst. **18** (1998), 1049–1073.

[21] N. Chernov, R. Markarian and S. Troubetzkoy, *Invariant measures for Anosov maps with small holes*, To appear in Ergod. Th. Dynam. Syst.

[22] M. Denker, *The central limit theorem for dynamical systems*, Dyn. Syst. Ergod. Th. Banach Center Publ., **23**, PWN–Polish Sci. Publ., Warsaw, 1989.

[23] R. L. Dobrushin, *The description of a random field by means of conditional probabilities and conditions of its regularity*, Th. Prob. Appl. **13** (1968), 197–224.

[24] R. L. Dobrushin, *Gibbsian random fields for lattice systems with pairwise interaction*, Funct. Anal. Appl. **2** (1968), 292–301.

[25] R. L. Dobrushin, *The problem of uniqueness of a Gibbsian random field and the problem of phase transitions*, Funct. Anal. Appl. **2** (1968), 302–312.

[26] D. Dolgopyat, *On decay of correlations in Anosov flows*, Ann. Math. **147** (1998), 357–390.

[27] N. Dunford and J. Schwartz, *Linear operators: General theory: Part I*, J. Wiley and Sons, 1988.

[28] J. Franks and M. Misiurewicz, *Topological methods in dynamics*, this Handbook.

[29] J. Franks and R. Williams, *Anomalous Anosov flows*, in: Global Theory of Dynamical Systems, Lect. Notes Math. **819**, Springer-Verlag, Berlin, 1980, pp. 158–174.

[30] N. Friedman and D. Ornstein, *On the isomorphism of weak Bernoulli transformations*, Adv. Math. **5** (1970), 365–394.

[31] G. Gallavotti and E.D.G. Cohen, *Dynamical ensembles in stationary states*, J. Stat. Phys. **80** (1995), 931–970.

[32] G. Gallavotti, *Reversible Anosov diffeomorphisms and large deviations*, Math. Phys. Electronic J. **1** (1995), 1–12.

[33] F. Gantmacher, *Applications of the theory of matrices*, Interscience Publishers, N.Y., 1959.

[34] P. Gaspard and G. Nicolis, *Transport properties, Lyapunov exponents, and entropy per unit time*, Phys. Rev. Lett. **65** (1990), 1693–1696.

[35] P. Gaspard and J. R. Dorfman, *Chaotic scattering theory, thermodynamic formalism, and transport coefficients*, Phys. Rev. E **52** (1995), 3525–3552.

[36] Y. Guivarc'h and J. Hardi, *Théorèmes limites pour une classe de chaines de Markov et applications aux difféomorphismes d'Anosov*, Ann. Inst. H. Poincaré, Probab. Statist. **24** (1988), 73–98.

[37] B. M. Gurevic, *Some conditions for the existence of K-partitions for special flows*, Trans. Moscow Math. Soc. **17** (1967), 99–128.

[38] B. Hasselblatt, *Hyperbolic dynamical systems*, this Handbook.

[39] N. T. Haydn, *Canonical product structure of equilibrium states*, Random Comp. Dynamics, **2** (1994), 79–96.

[40] M. Hirsch and C. Pugh, *Smoothness of horocycle foliations*, J. Diff. Geom. **10** (1975) 225–238.

[41] E. Hoph, *Statistik der Lösungen geodätischer Probleme vom unstabilen Typus, II*, Math. Annalen **117** (1940), 590–608.

[42] H. Hu and L.-S. Young, *Nonexistence of SRB measures for some systems that are "almost Anosov"*, Ergod. Th. Dynam. Syst. **15** (1995), 67–76.

[43] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge Univ. Press, 1995.

[44] A. Katok and B. Hasselblatt, *Principal structures*, this Handbook.

[45] T. Krüger and S. Troubetzkoy, *Markov partitions and shadowing for nonuniformly hyperbolic systems with singularities*, Ergod. Th. and Dynam. Sys. **12** 1992, 487–508.

[46] T. Krüger and S. Troubetzkoy, *Markov partitions via shadowing for diffeomorphisms with no zero exponents*, manuscript.

[47] O. E. Lanford and D. Ruelle, *Observables at infinity and states with short range correlations in statistical mechanics*, Commun. Math. Phys. **13** (1969), 194–215.

[48] F. Ledrappier and J.-M. Strelcyn, *A proof of the estimation from below in Pesin's entropy formula*, Ergod. Th. Dynam. Syst. **2** (1982), 203–219.

[49] F. Ledrappier, *Proprietes ergodiques des mesures de Sinai*, Publ. Math. IHES **59** (1984), 163–188.

[50] F. Ledrappier and L.-S. Young, *The metric entropy of diffeomorphisms: I. Characterization of measures satisfying Pesin's formula, II. Relations between entropy, exponents and dimension*, Ann. of Math. **122** (1985), 509–539, 540–574.

[51] A. N. Livsic, *Cohomology of dynamical systems*, Math. USSR Izvestija **6** (1972), 1276–1301.

[52] A. N. Livsic and Ya. G. Sinai, *On invariant measures compatible with the smooth structure for transitive U-systems*, Soviet Math. Dokl. **13** (1972), 1656-1659.

[53] D. S. Ornstein and B. Weiss, *Geodesic flows are Bernoullian*, Israel J. Math., **14** (1973), 184–198.

[54] V. I. Oseledec, *A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems*, Trans. Moscow Math. Soc. **19** (1968), 197–231.

[55] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187-188** (1990), 268pp.

[56] Ya. B. Pesin, *Families of invariant manifolds corresponding to non-zero characteristic exponents*, Math. USSR Izvestija **10** (1976), 1261–1305.

[57] Ya. B. Pesin, *Lyapunov characteristic exponents and smooth ergodic theory*, Russ. Math. Surv. **32** (1977), 55–114.

[58] Ya. B. Pesin, *Dynamical systems with generalized hyperbolic attractors*, Ergod. Th. Dynam. Sys. **12** (1992), 123–151.

[59] W. Philipp and W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Memoir. Amer. Math. Soc. **161** (1975).

[60] J. Plante, *Anosov flows*, Amer. J. Math. **94** (1972) 729–754.

[61] C. Pugh and M. Shub, *Ergodic attractors*, Trans. Amer. Math. Soc. **312** (1989), 1–54.

[62] M. Pollicott, *On the rate of mixing of Axiom A flows*, Invent. Math. **81** (1985), 413–426.

[63] M. Pollicott, *Periodic orbits and zeta functions*, this Handbook.

[64] M. Ratner, *The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature*, Israel J. Math. **16** (1973), 181–197.

[65] M. Ratner, *Anosov flows with Gibbs measures are also Bernoullian*, Israel J. Math. **17** (1974), 380–391.

[66] W. Rudin, *Real and complex analysis*, McGraw-Hill, 1987.

[67] D. Ruelle, *A measure associated with Axiom-A attractors*, Amer. J. Math. **98** (1976), 619–654.

[68] D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, Reading, Mass., 1978.

[69] D. Ruelle, *An inequality for the entropy of differentiable maps*, Bol. Soc. Bras. Mat. **9** (1978), 83–87.

[70] D. Ruelle, *Flots qui ne mélangent pas exponentiellement*, C. R. Acad. Sci. Paris **296** (1983), 191–193.

[71] D. Ruelle, *Positivity of entropy production in nonequilibrium statistical mechanics*, J. Statist. Phys. **85** (1996), 1–23.

[72] D. Ruelle, *Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics*, J. Statist. Phys. **94** (1999), 393–468.

[73] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[74] Ya. G. Sinai, *Markov partitions and C-diffeomorphisms*, Funct. Anal. Its Appl. **2** (1968), 61–82.

[75] Ya. G. Sinai, *Construction of Markov partitions*, Funct. Anal. Its Appl. **2** (1968), 245–253.

[76] Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surveys **27** (1972), 21–69.

[77] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.

[78] J.-P. Thouvenot, *Ergodic theory: Entropy, isomorphisms and Kakutani equivalence*, this Handbook.

[79] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Math. **147** (1998), 585–650.

[80] L.-S. Young, *Ergodic theory of chaotic dynamical systems*, XIIth International Congress of Mathematical Physics (ICMP'97) (Brisbane), 131–143, Internat. Press, Cambridge, MA, 1999.