SYMBOLIC DYNAMICS FOR LOW DIMENSIONAL NON-UNIFORMLY HYPERBOLIC SYSTEMS

YURI LIMA

Abstract. The goal of these notes is to construct symbolic models for uniformly hyperbolic systems and low-dimensional non-uniformly hyperbolic systems (surface diffeomorphisms and three dimensional flows). In the first part we focus on uniformly hyperbolic systems, and discuss the method of successive approximations of Sinai [Sin68] and the method of pseudo-orbits of Bowen [Bow08]. In the second part we discuss the recent method of Sarig for non-uniformly hyperbolic surface diffeomorphisms [Sar13], also implemented for non-uniformly hyperbolic three dimensional flows with positive speed [LS]. Applications are also discussed.

1. Introduction: uniformly hyperbolic systems

The prototypes of uniformly hyperbolic systems are:

- Hyperbolic toral automorphisms, e.g. $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ induced by $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$.
- Smale's horseshoe, see [Sma98].
- Geodesic flows on compact manifolds with negative sectional curvature.

See the appendix for the definitions of uniformly hyperbolic systems.

Introductory example: Smale's horseshoe. Let $g : K \to K$ be the Smale horseshoe map and $\sigma : \Sigma \to \Sigma$ the full shift, where $\Sigma = \{0,1\}^{\mathbb{Z}}$. There is a bijection $\pi : \Sigma \to K$ s.t. $\pi \circ \sigma = g \circ \pi$. Therefore g, from a dynamical point of view, can be analyzed via σ. Many properties of σ are easier to understand, such as:

- Easy iteration.
- Number of periodic orbits.
- Invariant measures.

The goal of these notes is to discuss other smooth systems with similar symbolic representations.

Symbolic models. The triple (Σ, σ, π) above is an example of a symbolic model. The general definition is as follows. Fix a countable oriented graph $\mathscr{G} = (V, E)$.

Topological Markov shift (TMS): Let Σ = $\mathbb{Z}$–indexed paths on $\mathscr{G}$, and let $\sigma : \Sigma \to \Sigma$ be the left shift. The pair (Σ, σ) is called a *topological Markov shift*.

An element of Σ is denoted by $\underline{v} = \{v_n\}_{n\in\mathbb{Z}}$. Let $f : M \to M$ be an Axiom A diffeomorphism.

Date: July 7, 2016.

Symbolic model for diffeomorphisms: A *symbolic model* for f is a triple (Σ,σ,π) where (Σ,σ) is a TMS and $\pi:\Sigma\to\Omega(f)$ is a finite-to-one Hölder continuous map s.t. $f\restriction_{\Omega(f)}\circ\pi=\pi\circ\sigma$.

To define a symbolic model for flows, we add the flow direction to the TMS.

Topological Markov flow (TMF) [LS]: Given a TMS (Σ,σ) and a Hölder continuous function $r:\Sigma\to\mathbb{R}$ with $0<\inf r\leq\sup r<\infty$, define the *topological Markov flow* (Σ_r,σ_r) by:

- $\Sigma_r=\{(\underline{v},t):\underline{v}\in\Sigma,0\leq t\leq r(\underline{v})\}$ with the identification $(\underline{v},r(\underline{v}))\sim(\sigma(\underline{v}),0)$.
- $\sigma_r:\Sigma_r\to\Sigma_r$ the unit speed vertical flow on Σ_r.

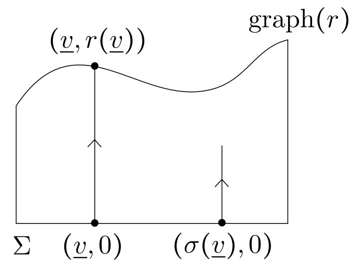


Figure 1. The suspension flow σ_r: starting at $(\underline{v},0)$, flow at unit speed until hit the graph of r, then return to the basis via the identification $(\underline{v},r(\underline{v}))\sim(\sigma(\underline{v}),0)$ and continue flowing.

Let $\varphi:M\to M$ be an Axiom A flow.

Symbolic model for flows: A *symbolic model* for φ is a triple $(\Sigma_r,\sigma_r,\pi_r)$ where (Σ_r,σ_r) is a TMF and $\pi_r:\Sigma_r\to\Omega(\varphi)$ is a finite-to-one Hölder continuous map s.t. $\varphi^t\restriction_{\Omega(\varphi)}\circ\pi_r=\pi_r\circ\sigma_r^t$ for all $t\in\mathbb{R}$.

Markov partitions. Let $f:M\to M$ be an Axiom A diffeomorphism, and let $x,y\in\Omega(f)$. The following are classical:

- $W^s_\varepsilon(x)=\{y\in M:\text{dist}(f^n(x),f^n(y))\leq\varepsilon,\forall n\geq 0\}$ = local stable manifold.
- $W^u_\varepsilon(x)=\{y\in M:\text{dist}(f^n(x),f^n(y))\leq\varepsilon,\forall n\leq 0\}$ = local unstable manifold.
- $\{[x,y]\}=W^s_\varepsilon(x)\cap W^u_\varepsilon(y)$ = Smale product. It exists whenever $\text{dist}(x,y)\ll 1$.

Fix $\delta\ll\varepsilon$ s.t. $[x,y]$ is well-defined whenever $\text{dist}(x,y)<\delta$. Given a subset $R\subset\Omega(f)$, let $R^*\subset\Omega(f)$ denote its interior in the induced topology of $\Omega(f)$. We call R a *rectangle* if:

- Regularity: $R=\overline{R^*}$ and $\text{diam}(R)<\delta$.
- Product structure: $x,y\in R\Rightarrow[x,y]\in R$.

In this case, let $W^s(x,R):=W^s_\varepsilon(x)\cap R$ and $W^u(x,R):=W^u_\varepsilon(x)\cap R$.

It is easy to construct rectangles: given $x\in\Omega(f)$, the set

$$[W^u_\rho(x)\cap\Omega(f),W^s_\rho(x)\cap\Omega(f)]$$

is a rectangle for all $\rho>0$ sufficiently small. Let $\mathscr{R}$ be a cover of $\Omega(f)$ by rectangles.

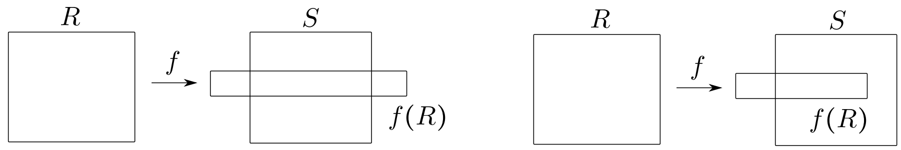
Markov partition: $\mathscr{R}$ is called a *Markov partition* for f if:

(1) Disjointness: The elements of $\mathscr{R}$ can only intersect at their boundaries.

(2) MARKOV PROPERTY: If $x \in R^*$ and $f(x) \in S^*$, then

$$f(W^s(x,R)) \subset W^s(f(x),S) \quad \text{and} \quad f^{-1}(W^u(f(x),S)) \subset W^u(x,R).$$

If $\mathscr{R}$ only satisfies (2), we call it a *Markov cover*.



☺ Allowed intersection ☺ ☹ Not allowed interserction ☹

FIGURE 2. The Markov property: if $f(R)$ intersects S non-trivially, then $f(R)$ crosses S completely all the way from one side to the other.

Markov sections. Let $\varphi : M \to M$ be an Axiom A flow.

PROPER SECTION [Bow73]: A finite family $\mathscr{M} = \{B_1, \ldots, B_n\}$ is a *proper section of size* α if there are closed differentiable discs $D_1, \ldots, D_n$ s.t.:

(1) CLOSEDNESS: Each B_i is a closed subset of $\Omega(\varphi)$.
(2) COVER: $\Omega(\varphi) = \bigcup_{i=1}^n \varphi^{[0,\alpha]} B_i$.
(3) REGULARITY: $B_i \subset \mathrm{int}(D_i)$ and $\overline{B_i^*} = B_i$, where B_i^* is the interior of B_i in the induced topology of $D_i \cap \Omega(\varphi)$.
(4) PARTIAL ORDER: For $i \neq j$, at least one of the sets $D_i \cap \varphi^{[0,4\alpha]} D_j$ and $D_j \cap \varphi^{[0,4\alpha]} D_i$ is empty; in particular $D_i \cap D_j = \varnothing$.

For simplicity, denote $B_1 \cup \cdots \cup B_n$ also by $\mathscr{M}$. Let $H = H_{\mathscr{M}} : \mathscr{M} \to \mathscr{M}$ be the *Poincaré return map*, i.e. $H(x)$ is the first time the forward orbit of x hits $\mathscr{M}$ again. Let also $\mathfrak{t} = \mathfrak{t}_{\mathscr{M}} : \mathscr{M} \to (0,\infty)$ be the *return time function*, i.e. $\mathfrak{t}(x)$ is the smallest $t > 0$ s.t. $H(x) = \varphi^t(x)$. By properties (2) and (4), $0 < \inf \mathfrak{t} \le \sup \mathfrak{t} \le \alpha$.

The maps $H, \mathfrak{t}$ are not continuous, but they are continuous on the subset

$$\mathscr{M}' := \left\{ x \in \mathscr{M} : H^k(x) \in \bigcup B_i^*, \forall k \in \mathbb{Z} \right\}.$$

Most of the arguments we will explain below will consider points in $\mathscr{M}'$. This avoids many problems, the first being the definition of the Markov property. We do not want to consider a transition from B_i to B_j when $H(B_i) \cap B_j$ is a subset of ∂B_j.

TRANSITIONS: We say that $B_i \to B_j$ if there exists $x \in \mathscr{M}'$ s.t. $x \in B_i$, $H(x) \in B_j$. When this happens, define $\mathscr{T}^s(B_i, B_j) := \overline{\{x \in \mathscr{M}' : x \in B_i, H(x) \in B_j\}}$ and $\mathscr{T}^u(B_i, B_j) := \overline{\{y \in \mathscr{M}' : y \in B_j, H^{-1}(y) \in B_i\}}$.

MARKOV SECTION [Bow73]: $\mathscr{M}$ is called a *Markov section of size* α if it is a proper section of size α with the following additional properties:

(5) PRODUCT STRUCTURE: Each B_i is a rectangle.
(6) MARKOV PROPERTY: If $B_i \to B_j$, then

$$x \in \mathscr{T}^s(B_i, B_j) \Rightarrow W^s(x, B_i) \subset \mathscr{T}^s(B_i, B_j)$$
$$y \in \mathscr{T}^u(B_i, B_j) \Rightarrow W^u(y, B_j) \subset \mathscr{T}^u(B_i, B_j).$$

Markov partitions/sections generate symbolic models. If $\mathscr{R}$ is a Markov partition for f, then f has a symbolic model:

- $\mathscr{G} = (V, E)$ with $V = \mathscr{R}$ and $E = \{R \to S : f(R^*) \cap S^* \neq \varnothing\}$.
- $\pi : \Sigma \to \Omega(f)$ is defined for $\underline{R} = \{R_n\} \in \Sigma$ by
$$\{\pi(\underline{R})\} := \bigcap_{n\in\mathbb{Z}} f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n).$$
Alternatively, $\pi(\underline{R})$ is the unique x s.t. $f^n(x) \in R_n$, $\forall n \in \mathbb{Z}$. The map π is well-defined because of the Markov property and uniform hyperbolicity.

If $\mathscr{M}$ is a Markov section for φ, then φ has a symbolic model:

- $\mathscr{G} = (V, E)$ with $V = \mathscr{M}$ and $E = \{B_i \to B_j : \exists x \in \mathscr{M}' \text{ s.t. } x \in B_i^*, H(x) \in B_j^*\}$.
- $\pi : \Sigma \to \mathscr{M} \cap \Omega(f)$ is defined for $\underline{B} = \{B_n\} \in \Sigma$ by
$$\{\pi(\underline{B})\} := \bigcap_{n\in\mathbb{Z}} H^{-n}(B_n).$$
- $r : \Sigma \to \mathbb{R}$ is defined by $r := \mathfrak{t} \circ \pi$.
- $\pi_r : \Sigma_r \to \Omega(\varphi)$ is defined by $\pi_r(\underline{B}, t) := \varphi^t[\pi(\underline{B})]$.

Therefore it is sufficient, in the uniformly hyperbolic case, to construct Markov partitions/sections.

Markov partitions for hyperbolic toral automorphisms [AW67]**.** This method, developed by Adler and Weiss [AW67], constructs Markov partitions for two dimensional hyperbolic toral automorphisms. Fix $f = f_A$ as in the introduction, and let W^s = contracting eigendirection at $(0,0)$, W^u = expanding eigendirection at $(0,0)$. Since f is linear, W^s and W^u are the stable and unstable manifolds of $(0,0)$, respectively. Construct a Markov partition for f following the steps below:

- Take a cover $\mathscr{R}$ of $\mathbb{T}^2$ by finitely many rectangles whose sides belong to W^s and W^u s.t. that every non-trivial intersection $f(R^*) \cap S^*$ is connected, i.e. $f(R^*)$ does not intersect S^* "twice".
- Since f contracts W^s, the stable boundary of $f(\mathscr{R})$ is contained in W^s, while its unstable boundary contains W^u. Partition $\mathscr{R}$ further by adding the pre-image of the unstable segments of $f(\mathscr{R})$.
- The final cover $\mathscr{R}$ is a Markov partition.

The projection map $\pi : \Sigma \to \mathbb{T}^2$ is surjective and injective on $\{x \in \mathbb{T}^2 : f^n(x) \in \bigcup_{R\in\mathscr{R}} R^*, \forall n \in \mathbb{Z}\}$.

For higher-dimensional toral automorphisms, a similar construction works, but there is an important difference from the two dimensional case: the boundary of any Markov partition is never smooth [Bow78]. Here is a heuristic explanation in three dimensions: if the matrix has one contracting and two expanding eigenvalues, then the stable boundary of $\mathscr{R}$ cannot be fully contained in E^s, therefore it has components on E^u and so it expands.

2. The method of successive approximations

The method of successive approximations for diffeomorphisms [Sin68]**.** This method, due to Sinai [Sin68], builds Markov partitions for Anosov diffeomorphisms. It was also works, under some modifications, for Axiom A diffeomorphisms [Bow70]. The construction consists of three main steps.

STEP 1. Let $\mathscr{T} = \{T_i\}$ be a finite cover of M by rectangles (as we have seen before, it is easy to build one such cover).

STEP 2 (SUCCESSIVE APPROXIMATIONS). Recursively define families $\mathscr{S}_k = \{S_{i,k}\}$ and $\mathscr{U}_k = \{U_{i,k}\}$ of rectangles as follows:

- $S_{i,0} = U_{i,0} = T_i$.
- If $\mathscr{S}_k, \mathscr{U}_k$ are defined, let

$$S_{i,k+1} := \bigcup_{x \in S_{i,k}} \{[y,z] : y \in S_{i,k}, z \in f(W^s(f^{-1}(x), S_{j,k})) \text{ for } f^{-1}(x) \in S_{j,k}\}$$

$$U_{i,k+1} := \bigcup_{x \in U_{i,k}} \{[z,y] : y \in U_{i,k}, z \in f^{-1}(W^u(f(x), U_{j,k})) \text{ for } f(x) \in U_{j,k}\}.$$

Let $S_i := \bigcup_{k \geq 0} S_{i,k}$, $U_i := \bigcup_{k \geq 0} U_{i,k}$, and $Z_i := [\overline{U_i}, \overline{S_i}]$. Then $\mathscr{Z} = \{Z_i\}$ is a finite cover of M by rectangles satisfying the Markov property.

STEP 3 (BOWEN-SINAI REFINEMENT). To destroy non-trivial intersections, refine $\mathscr{Z}$ as follows. For Z_i, let $I_i = \{j : Z_i^* \cap Z_j^* \neq \varnothing\}$. For $j \in I_i$, let $\mathscr{E}_{ij}$ = cover of Z_i by rectangles:

$$E_{ij}^{su} := \overline{\{x \in Z_i^* : W^s(x, Z_i) \cap Z_j^* \neq \varnothing, W^u(x, Z_i) \cap Z_j^* \neq \varnothing\}}$$
$$E_{ij}^{s\varnothing} := \overline{\{x \in Z_i^* : W^s(x, Z_i) \cap Z_j^* \neq \varnothing, W^u(x, Z_i) \cap Z_j = \varnothing\}}$$
$$E_{ij}^{\varnothing u} := \overline{\{x \in Z_i^* : W^s(x, Z_i) \cap Z_j = \varnothing, W^u(x, Z_i) \cap Z_j^* \neq \varnothing\}}$$
$$E_{ij}^{\varnothing\varnothing} := \overline{\{x \in Z_i^* : W^s(x, Z_i) \cap Z_j = \varnothing, W^u(x, Z_i) \cap Z_j = \varnothing\}}.$$

Hence $\mathscr{R}$:= cover defined by $\{\mathscr{E}_{ij} : Z_i \in \mathscr{Z}, j \in I_i\}$ is a Markov partition, and π is 1–1 on $\{x \in M : f^n(x) \in \bigcup_{R \in \mathscr{R}} R^*, \forall n \in \mathbb{Z}\}$.

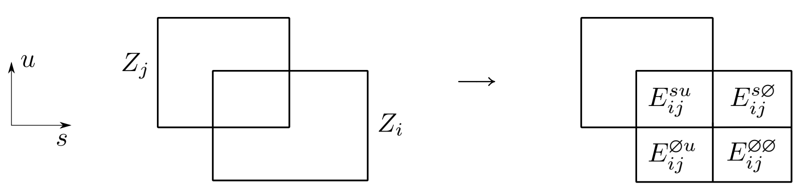


FIGURE 3. $\mathscr{E}_{ij} = \{E_{ij}^{su}, E_{ij}^{s\varnothing}, E_{ij}^{\varnothing u}, E_{ij}^{\varnothing\varnothing}\}$ is a cover of Z_i by rectangles.

The method of successive approximations for flows [Rat69, Rat73, Bow73]**.** Ratner applied the method of successive approximations for three dimensional Anosov flows [Rat69]. Later she used it for higher-dimensional Anosov flows [Rat73], and Bowen used it for Axiom A flows [Bow73]. We follow Bowen's construction. The main difficult is the presence of discontinuities for the Poincaré map.

Consider a proper section $\mathscr{C}$. By transversality, the stable/unstable directions of φ project to stable/unstable directions of the Poincaré map H, hence it is easy to construct rectangles inside $\mathscr{C}$. Let $\mathscr{R}$ be a cover of $\mathscr{C} \cap \Omega(\varphi)$ by rectangles. To apply successive approximations, proceed as follows:

- Take $L > 0$ large s.t. for every $x \in R \in \mathscr{R}$ there are $C^+, C^- \in \mathscr{C}$ s.t. $\varphi^L(W_\varepsilon^s(x)) \subset \varphi^{[-\alpha,\alpha]}(C^+)$ and $\varphi^{-L}(W_\varepsilon^u(x)) \subset \varphi^{[-\alpha,\alpha]}(C^-)$.

- For each such x, take a neighborhood $V \ni x$ small enough s.t. $\varphi^L(V) \subset \varphi^{[-\alpha,\alpha]}(C^+)$ and $\varphi^{-L}(V) \subset \varphi^{[-\alpha,\alpha]}(C^-)$, and define $H_V^+ : V \to C^+$ and $H_V^- : V \to C^-$ by:

$$H_V^+ := (\text{projection to } C^+) \circ \varphi^L, \quad H_V^- := (\text{projection to } C^-) \circ \varphi^{-L}.$$

- Pass to a finite collection of neighborhoods V as above, and apply the method of successive approximations to the maps H_V^+, H_V^-. The resulting cover by rectangles has a Markov property: for each $x \in \mathscr{R}$ there are $k, \ell > 0$ s.t. x satisfies a stable Markov property at $H^k(x)$ and an unstable Markov property at $H^{-\ell}(x)$.
- The values of k, ℓ are uniformly bounded by some $N > 0$.
- To get the Markov property for H, apply a refinement procedure along the iterates $-N, \ldots, N$ of H. The resulting partition is a Markov section.

Summary of results.

- Adler-Weiss 1967: two dimensional hyperbolic toral automorphisms [AW67].
- Sinai 1968: Anosov diffeomorphisms [Sin68].
- Bowen 1970: Axiom A diffeomorphisms [Bow70].
- Ratner 1973: Anosov flows [Rat69, Rat73].
- Bowen 1973: Axiom A flows [Bow73].

3. Applications

The main applications consist of pushing properties of the symbolic model to the smooth one. We focus on three of them:

- Lifting invariant measures.
- Counting periodic orbits.
- Ergodic properties of equilibrium states.

Lifting invariant measures. Assume $f : M \to M$ is an Axiom A diffeomorphism, and let (Σ, σ, π) be a symbolic model. It is easy to project measures: if $\widehat{\mu}$ is σ–invariant, then $\widehat{\mu} \circ \pi^{-1}$ is f–invariant. It is harder to lift measures without increasing entropy, but this is possible when π is finite-to-one. If μ is f–invariant, then $\widehat{\mu} = \int_M \frac{1}{|\pi^{-1}(x)|}(\sum_{y \in \pi^{-1}(x)} \delta_y) d\mu(x)$ is σ–invariant and satisfies $h_{\widehat{\mu}}(\sigma) = h_\mu(f)$, by the Abramov-Rokhlin formula. This is part of Sinai's program on statistical mechanics: first build a symbolic model, then use it to construct Gibbs measures [Sin72].

Another consequence of finiteness-to-one is that $h_{\rm top}(\sigma) = h_{\rm top}(f)$, and $\widehat{\mu}$ is a measure of maximal entropy for σ iff μ is for f. Hence every (transitive) Axiom A diffeomorphism has a unique measure of maximal entropy [Bow70].

Counting periodic orbits. Assume f has a measure of maximal entropy. Periodic orbits of σ project to periodic orbits of f. Reversely, a periodic orbit of f lifts to finitely many periodic orbits of σ, hence ${\rm Per}_n(f) \sim {\rm Per}_n(\sigma)$. If (Σ, σ) has period p then ${\rm Per}_{np}(\sigma) \sim e^{nph}$, where $h = h_{\rm top}(\sigma) = h_{\rm top}(f)$. Therefore ${\rm Per}_{np}(f) \sim e^{nph_{\rm top}(f)}$ for all $n \geq 0$.

For Axiom A flows, Parry and Pollicott proved that $\#\{$closed orbits of period $\leq T\} \sim \frac{e^{Th}}{T}$ [PP83]. The proof is much harder than for diffeomorphisms.

Ergodic properties of equilibrium measures. Gibbs measures, under regularity assumptions, are equilibrium measures. The definition of equilibrium measures is in the appendix. They are the measures that minimize the free energy of a fixed potential. Equilibrium measures of Axiom A systems are lifted to equilibrium measures of the symbolic model. This implies that every Hölder continuous function has a unique equilibrium measure [Bow75b]. Additionally, it is either Bernoulli or Bernoulli times a periodic measure [Bow75a].

4. The method of pseudo-orbits

Bowen developed an alternative method to build Markov partitions for Axiom A diffeomorphisms [Bow08]. The construction uses pseudo-orbits, graph transforms, and shadowing. For simplicity, let us assume that M is a surface and that its Riemannian metric is adapted to f (see the appendix).

Lyapunov charts. Let $e_1 = (1,0), e_2 = (0,1)$ be the canonical basis of $\mathbb{R}^2$. Given $x \in \Omega(f)$, let $e^s_x, e^u_x \in T_xM$ be unitary vectors in the directions of E^s_x, E^u_x respectively. Define $C(x) : \mathbb{R}^2 \to T_xM$ s.t.

$$C(x) : e_1 \mapsto e^s_x \,,\, e_2 \mapsto e^u_x.$$

Fix $\varepsilon > 0$ small enough so that the exponential map $\exp_x : [-\varepsilon,\varepsilon]^2 \subset T_xM \to M$ is a diffeomorphism onto its image for every $x \in M$.

Lyapunov chart: It is the map $\Psi_x : [-\varepsilon,\varepsilon]^2 \to M$, $\Psi_x := \exp_x \circ C(x)$.

In Lyapunov charts, f takes the form $f_x := \Psi^{-1}_{f(x)} \circ f \circ \Psi_x$.

Theorem 4.1. *The following hold for all $\varepsilon > 0$ small enough.*

(a) $d(f_x)_0 = C(f(x))^{-1} \circ df_x \circ C(x) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ *with* $|A| < \lambda$ *and* $|B| > \lambda^{-1}$.

(b) $f_x(u,v) = (Au + h_1(u,v), Bv + h_2(u,v))$, *where* $\|h_1\|_{C^1}, \|h_2\|_{C^1} < \varepsilon$.

When $f(x)$ is close to y, we can similarly define $f_{xy} := \Psi^{-1}_y \circ f \circ \Psi_x$. By continuity, if $\mathrm{dist}(f(x),y) \ll 1$ then f_{xy} satisfies similar conditions of Theorem 4.1. We fix $\delta > 0$ sufficiently small so that f_{xy} satisfies Theorem 4.1 for every $x,y \in \Omega(f)$ with $\mathrm{dist}(f(x),y) < \delta$, and f^{-1}_{xy} satisfies Theorem 4.1 for every $x,y \in \Omega(f)$ with $\mathrm{dist}(f^{-1}(y),x) < \delta$. When $\mathrm{dist}(f(x),y) < \delta$ and $\mathrm{dist}(f^{-1}(y),x) < \delta$, write $\Psi_x \to \Psi_y$ and call it an *edge*.

Graph transform. A consequence of the hyperbolic behavior of f_{xy} is that it sends curves almost parallel to e_1 to curves with the same property, and analogously its inverse sends curves almost parallel to e_2 to curves with the same property. These curves can be represented as graphs of real functions.

Admissible manifolds: An *s–admissible manifold* at Ψ_x is a set of the form $V^s = \Psi_x\{(t,F(t)) : |t| \le \varepsilon\}$, where $F : [-\varepsilon,\varepsilon] \to \mathbb{R}$ is a C^1 function s.t. $|F(0)| < \varepsilon$ and $\|F'\|_{C^0} \le 1$. Similarly, a *u–admissible manifold* at Ψ_x is a set of the form $V^u = \Psi_x\{(G(t),t) : |t| \le \varepsilon\}$, where $G : [-\varepsilon,\varepsilon] \to \mathbb{R}$ is a C^1 function s.t. $|G(0)| < \varepsilon$ and $\|G'\|_{C^0} \le 1$.

We call F,G the *representing functions*. Let $\mathscr{M}^s_x, \mathscr{M}^u_x$ be the space of all s,u–admissible manifolds at Ψ_x respectively. Introduce a metric on $\mathscr{M}^s_x$ as follows: for $V_1, V_2 \in \mathscr{M}^s_x$ with representing functions F_1, F_2, let

$$\mathrm{dist}(F_1,F_2) := \|F_1 - F_2\|_{C^0}.$$

A similar definition holds for $\mathscr{M}^u_x$. Assume that $\Psi_x \to \Psi_y$.

Graph transforms $\mathscr{F}^s_{x,y}, \mathscr{F}^u_{x,y}$: The *unstable graph transform* $\mathscr{F}^u_{x,y} : \mathscr{M}^u_x \to \mathscr{M}^u_y$ is the map that sends $V^u \in \mathscr{M}^u_x$ to the unique $\mathscr{F}^u_{x,y}[V^u] \in \mathscr{M}^u_y$ with representing function F s.t. $\Psi_y\{(F(t),t) : |t| \le \varepsilon\} \subset f(V^u)$. Similarly, the *stable graph transform* $\mathscr{F}^s_{x,y} : \mathscr{M}^s_y \to \mathscr{M}^s_x$ is the map that sends $V^s \in \mathscr{M}^s_y$ to the unique $\mathscr{F}^s_{x,y}[V^s] \in \mathscr{M}^s_x$ with representing function G s.t. $\Psi_y\{(t,G(t)) : |t| \le \varepsilon\} \subset f^{-1}(V^s)$.

In other words, $\mathscr{F}^u_{x,y}$ takes a u–admissible manifold at Ψ_x with representing function F to a u–admissible manifold at Ψ_y whose graph of the representing function is contained in the graph of $f_{x,y} \circ F$, and $\mathscr{F}^s_{x,y}$ takes an s–admissible manifold at Ψ_y with representing function G to an s–admissible manifold at Ψ_x whose graph of the representing function is contained in the graph of $f^{-1}_{x,y} \circ G$. See figure 4.

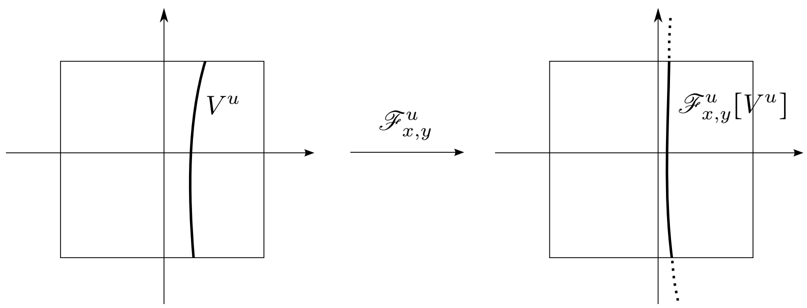


Figure 4. The graph transform $\mathscr{F}^u_{x,y}$: it sends a u–admissible manifold at Ψ_x to a u–admissible manifold at Ψ_y.

Theorem 4.2. *If $\Psi_x \to \Psi_y$, then $\mathscr{F}^s_{x,y}$ and $\mathscr{F}^u_{x,y}$ are well-defined contractions.*

The proof of the theorem follows from the hyperbolicity of f_{xy}.

Shadowing. Let $\underline{v} = \{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ be a sequence of Lyapunov charts s.t. $\Psi_{x_n} \to \Psi_{x_{n+1}}$ for all $n \in \mathbb{Z}$.

Stable/unstable manifolds: The *stable manifold* of $\{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ is the unique s–admissible manifold $V^s[\underline{v}] \in \mathscr{M}^s_{x_0}$ defined by

$$V^s[\underline{v}] := \lim_{n\to\infty} (\mathscr{F}^s_{x_0,x_1} \circ \cdots \circ \mathscr{F}^s_{x_{n-1},x_n})[V_n]$$

for some (any) sequence $\{V_n\}_{n\ge 0}$ with $V_n \in \mathscr{M}^s_{x_n}$. The *unstable manifold* of $\{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ is the unique u–admissible manifold $V^u[\underline{v}] \in \mathscr{M}^u_{x_0}$ defined by

$$V^u[\underline{v}] := \lim_{n\to-\infty} (\mathscr{F}^u_{x_{-1},x_0} \circ \cdots \circ \mathscr{F}^u_{x_n,x_{n+1}})[V_n]$$

for some (any) sequence $\{V_n\}_{n\le 0}$ with $V_n \in \mathscr{M}^u_{x_n}$.

$V^s[\underline{v}]$ and $V^u[\underline{v}]$ are well-defined because graph transforms are contractions (Theorem 4.2). Note that $V^s[\underline{v}]$ only depends on the future $\{\Psi_{x_n}\}_{n\ge 0}$, whereas $V^u[\underline{v}]$ only depends on the past $\{\Psi_{x_n}\}_{n\le 0}$.

Shadowing: We say that $\underline{v} = \{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ *shadows* x if $f^n(x) \in \Psi_{x_n}([-\varepsilon,\varepsilon]^2)$ for all $n \in \mathbb{Z}$. Necessarily $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$.

Construction of a Markov partition. We now use the above tools to construct a Markov partition for f. The proof is divided into three steps.

STEP 1 (COARSE GRAINING). Cover $\Omega(f)$ by a finite $\delta/4$–dense set of Lyapunov charts $\mathscr{A}$. The finite oriented graph with vertex set $\mathscr{A}$ and edge set $\Psi_x \to \Psi_y$ defines a TMS (Σ,σ) with finitely many symbols.

STEP 2 (INFINITE-TO-ONE EXTENSION). Define $\pi:\Sigma\to\Omega(f)$ by

$$\{\pi(\underline{v})\} := V^s[\underline{v}]\cap V^u[\underline{v}].$$

π is a surjective map with $\pi\circ\sigma = f\circ\pi$, but it is usually infinite-to-one.

STEP 3 (BOWEN-SINAI REFINEMENT). Let $\mathscr{Z}=\{Z_v : v\in\mathscr{A}\}$, where $Z_v=\{\pi(\underline{v}): v_0=v\}$. Each $Z\in\mathscr{Z}$ is a rectangle, and $\mathscr{Z}$ is a cover of $\Omega(f)$ with a symbolic Markov property: if $x=\pi(\underline{v})$ then

$$f(W^s(x,Z_{v_0}))\subset W^s(f(x),Z_{v_1}) \text{ and } f^{-1}(W^u(f(x),Z_{v_1}))\subset W^u(x,Z_{v_0}).$$

Apply the refinement as in the method of successive approximations. The resulting partition $\mathscr{R}$ is a Markov partition, and it generates a new TMS $(\widehat{\Sigma},\widehat{\sigma})$ and a new coding $\widehat{\pi}:\widehat{\Sigma}\to\Omega(f)$. If $\#\mathscr{R}=N$, then $\widehat{\pi}$ is at most N^2-to-one.

5. NON-UNIFORMLY HYPERBOLIC SURFACE DIFFEOMORPHISMS

Now we want to go beyond uniformly hyperbolic systems and construct symbolic models for non-uniformly hyperbolic systems. Many difficulties that arise, e.g. non-uniform hyperbolicity is an almost-everywhere statement, and the hyperbolicity parameters are usually not continuous. Here are examples of non-uniformly hyperbolic systems:

- The slow down of $f_A:\mathbb{T}^2\to\mathbb{T}^2$, see [Kat79].
- Geodesic flows on manifolds with nonpositive sectional curvature.

Introduction. The first construction of symbolic dynamics in this setting is due to Katok [Kat80]. He used Pesin theory and the method of pseudo-orbits to code sets of large (but not necessarily full) topological entropy. A consequence is that $C^{1+\beta}$ surface diffeomorphisms have horseshoes of large (but not necessarily full) topological entropy. Recently, Sarig introduced new ideas to the method of Katok and constructed, for $C^{1+\beta}$ surface diffeomorphisms, horseshoes[1] of full topological entropy [Sar13]. It is this result that we now discuss.

Let M be a closed smooth surface, and $f:M\to M$ a $C^{1+\beta}$ surface diffeomorphism. We want to code f with respect to hyperbolic measures. Let $\chi>0$.

χ–HYPERBOLIC MEASURE: A *χ–hyperbolic measure* is an f–invariant probability measure on M s.t. μ–a.e. point has one Lyapunov exponent $>\chi$ and one Lyapunov exponent $<-\chi$.

For example, every ergodic invariant measure with $h_\mu(f)>\chi$ is χ–hyperbolic. This follows from the Ruelle inequality and the identity $h_\mu(f)=h_\mu(f^{-1})$. For a fixed $\chi>0$, the method of Sarig codes *all* χ–hyperbolic measures simultaneously. Here is the precise statement.

---

[1] The terminology of horseshoes for non-uniformly hyperbolic diffeomorphisms we consider here is different from the classical one; for instance its alphabet is usually infinite.

Theorem 5.1 ([Sar13]). *Let $f : M \to M$ be a $C^{1+\beta}$ surface diffeomorphism on a closed smooth Riemannian manifold M. For every $\chi > 0$, there exists a TMS (Σ, σ) and a Hölder continuous map $\pi : \Sigma \to M$ s.t.:*

(1) $\pi \circ \sigma = f \circ \pi$.
(2) $\pi[\Sigma^\#]$ *has full μ–measure for every χ–hyperbolic measure μ.*
(3) π *is finite-to-one on* $\pi[\Sigma^\#]$.

Above, $\Sigma^\#$ is the *recurrent set* of Σ. It is the set of $\{v_n\} \in \Sigma$ s.t. $\exists v, w \in V$ s.t. $v_n = v$ for infinitely $n > 0$ and $v_n = w$ for infinitely many $n < 0$. Contrary to the uniformly hyperbolic case, Σ usually has infinitely many states.

The non-uniformly hyperbolic set NUH_χ**.** By the Oseledets theorem, every χ–hyperbolic measure is carried by the set NUH_χ of points $x \in M$ for which there are vectors $\{e^s_{f^n(x)}\}_{n\in\mathbb{Z}}$, $\{e^u_{f^n(x)}\}_{n\in\mathbb{Z}}$ s.t. for every $y = f^n(x)$ the following hold:

(1) $e^{s/u}_y \in T_yM$, $\|e^{s/u}_y\| = 1$.
(2) $\mathrm{span}(df^m e^{s/u}_y) = \mathrm{span}(e^{s/u}_{f^m(y)})$ for all $m \in \mathbb{Z}$.
(3) $\lim_{m\to\pm\infty} \frac{1}{m} \log \|df^m e^s_y\| < -\chi$ and $\lim_{m\to\pm\infty} \frac{1}{m} \log \|df^m e^u_y\| > \chi$.
(4) $\lim_{m\to\pm\infty} \frac{1}{m} \log |\sin \alpha(f^m(y))| = 0$, where $\alpha(f^m(y)) = \angle(e^s_{f^m(y)}, e^u_{f^m(y)})$.

We reinforce that, contrary to the uniformly hyperbolic case, the maps $x \in \mathrm{NUH}_\chi \mapsto e^s_x, e^u_x$ are usually not more than just measurable. Points of NUH_χ are just asymptotically hyperbolic. To measure the quality of hyperbolicity, we consider the parameters below.

Parameters $s(x), u(x)$: For $x \in \mathrm{NUH}_\chi$, define

$$s(x) := \sqrt{2}\left(\sum_{n\geq 0} e^{2n\chi}\|df^n e^s_x\|^2\right)^{1/2},\ u(x) := \sqrt{2}\left(\sum_{n\geq 0} e^{2n\chi}\|df^{-n} e^u_x\|^2\right)^{1/2}.$$

There are three cases when $x \in \mathrm{NUH}_\chi$ has bad uniform hyperbolicity:

- $s(x)$ is large: it takes a long time to see contraction along E^s.
- $u(x)$ is large: it takes a long time to see expansion along E^u.
- $\alpha(x)$ is small: it is hard to distinguish the stable and unstable directions.

Linear Pesin theory on NUH_χ**.** For $x \in \mathrm{NUH}_\chi$, consider the linear map $C_\chi(x) : \mathbb{R}^2 \to T_xM$ s.t.

$$C_\chi(x) : e_1 \mapsto \tfrac{e^s_x}{s(x)},\ e_2 \mapsto \tfrac{e^u_x}{u(x)}.$$

Theorem 5.2 ([Sar13]). *For all $x \in \mathrm{NUH}_\chi$, the following hold:*

(1) $\|C_\chi(x)\| \leq 1$ *and* $\|C_\chi(x)^{-1}\|_{\mathrm{Frob}} = \frac{\sqrt{s(x)^2+u(x)^2}}{|\sin\alpha(x)|}$.
(2) $C_\chi(f(x))^{-1} \circ df_x \circ C_\chi(x)$ *is a diagonal matrix with diagonal entries $A, B \in \mathbb{R}$ s.t. $|A| < e^{-\chi}$ and $|B| > e^\chi$.*

Above, $\|\cdot\|_{\mathrm{Frob}}$ represents the Frobenius norm[2].

[2]The Frobenius norm of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\|A\|_{\mathrm{Frob}} = \sqrt{a^2+b^2+c^2+d^2}$. It is equivalent to the usual sup norm.

Proof. (1) For $v = \alpha e_1 + \beta e_2$ with $\alpha^2 + \beta^2 \le 1$, $\|C_\chi(x)v\|^2 \le 2(\alpha^2\|C_\chi(x)e_1\|^2 + \beta^2\|C_\chi(x)e_2\|^2) \le 2(\frac{\alpha^2}{2}+\frac{\beta^2}{2}) \le 1$, hence $\|C_\chi(x)\| \le 1$. For the second part, we express $C_\chi(x)^{-1}$ in coordinates. In the basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ and the basis $\{e^s_x, (e^s_x)^\perp\}$ of T_xM, $C_\chi(x)$ takes the form $\begin{bmatrix} \frac{1}{s(x)} & \frac{\cos\alpha(x)}{u(x)} \\ 0 & \frac{\sin\alpha(x)}{u(x)} \end{bmatrix}$, whose inverse is $\begin{bmatrix} s(x) & -\frac{s(x)\cos\alpha(x)}{\sin\alpha(x)} \\ 0 & \frac{u(x)}{\sin\alpha(x)} \end{bmatrix}$. Hence $\|C_\chi(x)^{-1}\|_{\rm Frob} = \frac{\sqrt{s(x)^2+u(x)^2}}{|\sin\alpha(x)|}$.

(2) It is clear that e_1, e_2 are eigenvectors of $C_\chi(f(x))^{-1} \circ df_x \circ C_\chi(x)$. We calculate the eigenvalue of e_1 (the calculation of the eigenvalue of e_2 is similar, and is left to the reader). Since $df e^s_x = \pm\|df e^s_x\| e^s_{f(x)}$, $[df \circ C_\chi(x)](e_1) = \pm df\left[\frac{e^s_x}{s(x)}\right] = \pm\frac{\|df e^s_x\|}{s(x)} e^s_{f(x)}$, therefore $[C_\chi(f(x))^{-1} \circ df \circ C_\chi(x)](e_1) = \pm\|df e^s_x\|\frac{s(f(x))}{s(x)} e_1$. Then $|A| := \|df e^s_x\|\frac{s(f(x))}{s(x)}$ is the eigenvalue of e_1. Note that

$$s(f(x))^2 = 2\sum_{n\ge0} e^{2n\chi}\|df^n e^s_{f(x)}\|^2 = \frac{2}{e^{2\chi}\|df e^s_x\|^2}\sum_{n\ge1} e^{2n\chi}\|df^n e^s_x\|^2$$
$$= \frac{1}{e^{2\chi}\|df e^s_x\|^2}[s(x)^2 - 2] < \frac{s(x)^2}{e^{2\chi}\|df e^s_x\|^2},$$

hence $|A| < e^{-\chi}$. □

Note that the difference from C_χ to the matrix C in the uniformly hyperbolic case is the presence of the denominators $s(x), u(x)$: if the hyperbolicity is bad, then C_χ dilates e_1, e_2 accordingly. Alternatively, $s(x), u(x)$ boost the non-uniform hyperbolicity to a uniform one. This is what causes $\|C_\chi(x)^{-1}\|$ to be large, see Theorem 5.2(2).

Non-linear Pesin theory on NUH_χ. We now define the equivalent of Lyapunov charts for the non-uniformly hyperbolic case. These are the charts that make f close to a hyperbolic matrix. Fix $\varepsilon > 0$ small.

PESIN CHART Ψ_x: For $x \in \mathrm{NUH}_\chi$, let $\Psi_x : [-\varepsilon, \varepsilon] \to M$, $\Psi_x := \exp_x \circ C_\chi(x)$.

Let $f_x := \Psi^{-1}_{f(x)} \circ f \circ \Psi_x$. As before, we wish that f_x is close to a hyperbolic matrix. The main obstruction is when $\|C_\chi(f(x))^{-1}\|$ is large, hence we reduce the domain of Pesin charts. Since we want to construct a countable set of Pesin charts (coarse graining in the sequel), the sizes of Pesin charts will always belong to $I_\varepsilon := \{e^{-\frac{1}{3}\varepsilon n} : n \ge 0\}$.

PARAMETER $Q_\varepsilon(x)$: $Q_\varepsilon(x)$ = largest element of I_ε that is $\le e^{3/\beta}\|C_\chi(f(x))^{-1}\|^{-12/\beta}_{\rm Frob}$.

Lemma 5.3 ([Sar13]). *There is a set $\mathrm{NUH}^*_\chi \subset \mathrm{NUH}_\chi$ that carries all χ–hyperbolic measures s.t. $\lim_{n\to\pm\infty}\frac{1}{n}\log Q_\varepsilon(f^n(x)) = 0$ for all $x \in \mathrm{NUH}^*_\chi$.*

The lemma above folllows from recurrence and the Oseledets theorem.

Theorem 5.4 ([Sar13]). *The following holds for all $\varepsilon > 0$ small enough. If $x \in \mathrm{NUH}_\chi$ then f_x is well-defined on $[-10Q_\varepsilon(x), 10Q_\varepsilon(x)]^2$ and satisfies:*
(1) $d(f_x)_0 = C_\chi(f(x))^{-1} \circ df_x \circ C_\chi(x)$.
(2) $f_x(u,v) = (Au + h_1(u,v), Bv + h_2(u,v))$ *with:*
  (a) $|A| < e^{-\chi}$ *and* $|B| > e^\chi$.

(b) $h_1(0,0) = h_2(0,0) = 0$ *and* $\nabla h_1(0,0) = \nabla h_2(0,0) = 0$.

(c) $\|h_i\|_{C^{1+\beta/2}} < \varepsilon$*,* $i = 1,2$*, where the norm is taken in* $[-10Q_\varepsilon(x), 10Q_\varepsilon(x)]^2$.

A similar statement holds for $f_x^{-1} = \Psi_x^{-1} \circ f^{-1} \circ \Psi_{f(x)}$.

Proof. Part (1) is consequence of Theorem 5.2(1) and the identities $d(\exp_x)_0 = \mathrm{Id}$ and $d(\exp_{f(x)}^{-1})_{f(x)} = \mathrm{Id}$. This also implies (a)–(b) of part (2). We focus on (c). We have $(h_1, h_2) := f_x - d(f_x)_0$. Since f is $C^{1+\beta}$, there is a constant $\mathfrak{K} > 0$, depending only on M and f, s.t.

$$\|d(\exp_{f(x)}^{-1} \circ f \circ \exp_x)_z - d(\exp_{f(x)}^{-1} \circ f \circ \exp_x)_w\| \le \mathfrak{K}\|z - w\|^\beta$$

for all $x \in \mathrm{NUH}_\chi$ and $z, w \in [-10Q_\varepsilon(x), 10Q_\varepsilon(x)]^2$. Then

$$\begin{aligned}
&\|(h_1,h_2)(z) - (h_1,h_2)(w)\| = \|d(f_x)_z - d(f_x)_w\| \\
&= \|C_\chi(f(x))^{-1} \circ [d(\exp_{f(x)}^{-1} \circ f \circ \exp_x)_z - d(\exp_{f(x)}^{-1} \circ f \circ \exp_x)_w] \circ C_\chi(x)\| \\
&\le \mathfrak{K}\|z-w\|^\beta \|C_\chi(f(x))^{-1}\| \le \left(40^{\beta/2}\mathfrak{K}Q_\varepsilon(x)^{\beta/2}\|C_\chi(f(x))^{-1}\|\right)\|z-w\|^{\beta/2},
\end{aligned}$$

since $\|z - w\| < 40Q_\varepsilon(x)$ and $\|C_\chi(x)\| \le 1$. By the definition of $Q_\varepsilon(x)$, we have

$$40^{\beta/2}\mathfrak{K}Q_\varepsilon(x)^{\beta/2}\|C_\chi(f(x))^{-1}\| \le 40^{\beta/2}\mathfrak{K}\varepsilon^{3/2},$$

which is $< \varepsilon$ if $\varepsilon > 0$ is sufficiently small. □

We will eventually consider Pesin charts with different domains. For $\eta \in I_\varepsilon$, let $\Psi_x^\eta := \Psi_x \restriction_{[-\eta,\eta]^2}$.

ε–overlap of Pesin charts. For $\mathrm{dist}(f(x), y) \ll 1$, let $f_{xy} = \Psi_y^{-1} \circ f \circ \Psi_x$. We want to have an analogue of Theorem 5.4 for f_{xy}. The idea is to write f_{xy} as $(\Psi_y^{-1} \circ \Psi_{f(x)}) \circ f_x$ and see it as a perturbation of f_x. Since $x \in \mathrm{NUH}_\chi \mapsto e_x^s, e_x^u$ are not necessarily continuous, it might happen that $\mathrm{dist}(f(x), y) \ll 1$ and yet $\Psi_y^{-1} \circ \Psi_{f(x)}$ has a large norm. To bypass this, Sarig defined the notion of *ε–overlap.*

PARALLELIZATION OF M: Let $\mathscr{D}$ be a finite cover of M s.t. for each $D \in \mathscr{D}$ there is a map Θ_D s.t. $\Theta_D : T_xM \to \mathbb{R}^2$ is a linear isometry for all $x \in D$.

Actually, Θ_D needs to have some additional regularity properties, but we skip it to ease the presentation, see [Sar13, §3.1] for the rigorous definition. From now on we write $\frac{a}{b} = e^{\pm c}$ when $e^{-c} \le \frac{a}{b} \le e^c$.

ε–OVERLAP [Sar13]: Two Pesin charts $\Psi_{x_1}^{\eta_1}, \Psi_{x_2}^{\eta_2}$ are said to *ε–overlap* if $\frac{\eta_1}{\eta_2} = e^{\pm\varepsilon}$ and if there is $D \in \mathscr{D}$ s.t. $x_1, x_2 \in D$ and

$$\mathrm{dist}(x_1, x_2) + \|\Theta_D \circ C_\chi(x_1) - \Theta_D \circ C_\chi(x_2)\| < (\eta_1\eta_2)^4. \tag{5.1}$$

When this happens, write $\Psi_{x_1}^{\eta_1} \overset{\varepsilon}{\approx} \Psi_{x_2}^{\eta_2}$.

Theorem 5.5 ([Sar13])**.** *The following holds for all* $\varepsilon > 0$ *small enough. If* $\Psi_{f(x)}^\eta \overset{\varepsilon}{\approx} \Psi_y^{\eta'}$ *then* f_{xy} *is well-defined on* $[-10Q_\varepsilon(x), 10Q_\varepsilon(x)]^2$ *and can be put in the form*

$$f_{xy}(u,v) = (Au + h_1(u,v), Bv + h_2(u,v))$$

where:

(a) $|A| < e^{-\chi}$ *and* $|B| > e^\chi$.

(b) $h_1(0,0) = h_2(0,0) = 0$ *and* $\nabla h_1(0,0) = \nabla h_2(0,0) = 0$.

(c) $\|h_i\|_{C^{1+\beta/3}} < \varepsilon$*,* $i = 1,2$*, where the norm is taken in* $[-10Q_\varepsilon(x), 10Q_\varepsilon(x)]^2$.

If $\Psi_{f^{-1}(y)}^{\eta'} \overset{\varepsilon}{\approx} \Psi_x^\eta$*, then a similar statement holds for* $f_{xy}^{-1} = \Psi_x^{-1} \circ f^{-1} \circ \Psi_y$.

Note that we relaxed the norm in (c) from $C^{1+\beta/2}$ to $C^{1+\beta/3}$. We will use Theorem 5.5 to apply the graph transform method.

The parameters $q_\varepsilon, q_\varepsilon^s, q_\varepsilon^u$**.** To apply the graph transform from Ψ_x^η to $\Psi_{f(x)}^{\eta'}$ we require that $\frac{\eta}{\eta'} = e^{\pm\varepsilon}$, otherwise the images of admissible manifolds might not be admissible. Unfortunately Q_ε, our initial candidate for the size of Pesin charts, does not satisfy this[3]. Fortunately, there is a way of defining a smaller value $q_\varepsilon(x)$ satisfying this. Take $\delta_\varepsilon \in I_\varepsilon$ largest as possible s.t. $\delta_\varepsilon < \varepsilon$.

THE PARAMETER $q_\varepsilon(x)$: For $x \in \mathrm{NUH}_\chi^*$, let

$$q_\varepsilon(x) := \delta_\varepsilon \min\{e^{\varepsilon|n|}Q_\varepsilon(f^n(x)) : n \in \mathbb{Z}\}.$$

Note that $q_\varepsilon(x)$ depends on the whole orbit of x. Since non-uniformly hyperbolic systems might have different future and past behavior, we separate this dependence as follows.

PARAMETERS $q_\varepsilon^s(x), q_\varepsilon^u(x)$: For $x \in \mathrm{NUH}_\chi^*$, let

$$q_\varepsilon^s(x) := \delta_\varepsilon \min\{e^{\varepsilon n}Q_\varepsilon(f^n(x)) : n \geq 0\}$$
$$q_\varepsilon^u(x) := \delta_\varepsilon \min\{e^{\varepsilon n}Q_\varepsilon(f^{-n}(x)) : n \geq 0\}.$$

Lemma 5.6. *For all* $x \in \mathrm{NUH}_\chi^*$, *the following holds:*

(1) GOOD DEFINITION: $0 < q_\varepsilon(x), q_\varepsilon^s(x), q_\varepsilon^u(x) < \varepsilon Q_\varepsilon(x)$ *and* $q_\varepsilon^s(x) \wedge q_\varepsilon^u(x) = q_\varepsilon(x)$.
(2) TEMPEREDNESS: $\frac{q_\varepsilon(f(x))}{q_\varepsilon(x)} = e^{\pm\varepsilon}$.
(3) GREEDY ALGORITHM: *for all* $n \in \mathbb{Z}$ *it holds*

$$q_\varepsilon^s(f^n(x)) = \min\{e^\varepsilon q_\varepsilon^s(f^{n+1}(x)), \delta_\varepsilon Q_\varepsilon(f^n(x))\}$$
$$q_\varepsilon^u(f^n(x)) = \min\{e^\varepsilon q_\varepsilon^u(f^{n-1}(x)), \delta_\varepsilon Q_\varepsilon(f^n(x))\}.$$

Clearly $q_\varepsilon(x) \in I_\varepsilon$ and $q_\varepsilon(x) < \varepsilon Q_\varepsilon(x)$. By Lemma 5.3, $q_\varepsilon(x) > 0$. We leave (2)–(3) as exercises to the reader.

ε**–double Pesin charts.** We now define ε–double charts. These will be the vertices of the TMS we will construct.

ε–DOUBLE CHART: An ε*–double chart* is a pair of Pesin charts $\Psi_x^{p^s,p^u} = (\Psi_x^{p^s}, \Psi_x^{p^u})$ where $p^s, p^u \in I_\varepsilon$ with $0 < p^s, p^u < \varepsilon Q_\varepsilon(x)$.

The parameters p^s, p^u control the local future/past hyperbolicity of x.

EDGE $v \overset{\varepsilon}{\to} w$: Given ε–double charts $v = \Psi_x^{p^s,p^u}$ and $w = \Psi_y^{q^s,q^u}$, we draw an edge from v to w if the following conditions are satisfied:

(GPO1) $\Psi_{f(x)}^{q^s\wedge q^u} \overset{\varepsilon}{\approx} \Psi_y^{q^s\wedge q^u}$ and $\Psi_{f^{-1}(y)}^{p^s\wedge p^u} \overset{\varepsilon}{\approx} \Psi_x^{p^s\wedge p^u}$.
(GPO2) $p^s = \min\{e^\varepsilon q^s, \delta_\varepsilon Q_\varepsilon(x)\}$ and $q^u = \min\{e^\varepsilon p^u, \delta_\varepsilon Q_\varepsilon(y)\}$.

(GPO1) allows to pass from an ε–double chart at x to an ε–double chart at y, and vice-versa. (GPO2) is a greedy algorithm that chooses the local hyperbolicity parameters the largest as possible. Its motivation comes from Lemma 5.6(3). (GPO2) is crucial to prove the inverse theorem (Theorem 6.1).

[3]The ratio $\frac{Q_\varepsilon(f(x))}{Q_\varepsilon(x)}$ is uniformly bounded away from zero and infinity, but it is not necessarily equal to $e^{\pm\varepsilon}$.

ε–GENERALIZED PSEUDO-ORBIT (ε–GPO): An *ε–generalized pseudo-orbit* is a sequence $\underline{v} = \{\Psi_{x_n}^{p_n^s,p_n^u}\}_{n\in\mathbb{Z}}$ of ε–double charts s.t. $\Psi_{x_n}^{p_n^s,p_n^u} \overset{\varepsilon}{\to} \Psi_{x_{n+1}}^{p_{n+1}^s,p_{n+1}^u}$ for all $n \in \mathbb{Z}$.

Graph transform and shadowing. We can similarly apply the graph transform method for edges $v \overset{\varepsilon}{\to} w$, but we need a stronger definition for admissible manifolds. Here it is important that f is $C^{1+\beta}$. Let $v = \Psi_x^{p^s,p^u}$ be an ε–double chart.

ADMISSIBLE MANIFOLDS: We define an *s–admissible manifold at v* as a set of the form $\Psi_x\{(t, F(t)) : |t| \le p^s\}$ where $F : [-p^s, p^s] \to \mathbb{R}$ is a $C^{1+\beta/3}$ function s.t.:

(AM1) $|F(0)| \le 10^{-3}(p^s \wedge p^u)$.
(AM2) $|F'(0)| \le \frac{1}{2}(p^s \wedge p^u)^{\beta/3}$.
(AM3) $\|F'\|_{C^0} + \mathrm{Hol}_{\beta/3}(F') \le \frac{1}{2}$ where the norms are taken in $[-p^s, p^s]$.

Similarly, a *u–admissible manifold at v* is a set of the form $\Psi_x\{(G(t), t) : |t| \le p^u\}$ where $G : [-p^u, p^u] \to \mathbb{R}$ is a $C^{1+\beta/3}$ function satisfying (AM1)–(AM3), where the norms are taken in $[-p^u, p^u]$.

Note that $p^{s/u}$ control the domains of the representing functions. An analogue of Theorem 4.2 holds. Therefore we can similarly define stable and unstable manifolds $V^s[\underline{v}], V^u[\underline{v}]$ for every ε–gpo $\underline{v}$, and hence also the notion of shadowing.

SHADOWING: We say that $\underline{v} = \{\Psi_{x_n}^{p_n^s,p_n^u}\}_{n\in\mathbb{Z}}$ *shadows* x if $f^n(x) \in \Psi_{x_n}([-p_n^s \wedge p_n^u, p_n^s \wedge p_n^u]^2)$ for all $n \in \mathbb{Z}$. Necessarily $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$.

Coarse graining. The next result constructs a countable set of ε–double charts whose ε–gpo's shadow all of NUH_χ^*.

Theorem 5.7. *For all $\varepsilon > 0$ sufficiently small, there exists a countable set $\mathscr{A}$ of ε–double charts with the following properties:*

1. DISCRETENESS*: For all $t > 0$, the set $\{\Psi_x^{p^s,p^u} \in \mathscr{A} : p^s, p^u > t\}$ is finite.*
2. SUFFICIENCY: *If $x \in \mathrm{NUH}_\chi^*$ then there is a sequence $\underline{v} \in \mathscr{A}^{\mathbb{Z}}$ that shadows x.*
3. RELEVANCE: *For all $v \in \mathscr{A}$ there is an ε–gpo $\underline{v} \in \mathscr{A}^{\mathbb{Z}}$ with $v_0 = v$ that shadows a point in* NUH_χ.

Remark 5.8. In the original statement [Sar13, Thm 4.16], sufficiency is only proved on a subset $\mathrm{NUH}_\chi^\#$ of NUH_χ^*. The reason of the improvement is that our definition of $q_\varepsilon(x)$ differs from that of [Sar13], and the introduction of the parameters $q_\varepsilon^s(x), q_\varepsilon^u(x)$.

In the sequel we fix one such $\mathscr{A}$ and let (Σ, σ) be the TMS generated by it. As in the uniformly hyperbolic case, let $\pi : \Sigma \to M$, $\pi(\underline{v}) := V^s[\underline{v}] \cap V^u[\underline{v}]$.

6. INVERSE THEOREM

Introduction. The next step is to use π to construct a family $\mathscr{Z}$ of rectangles with a symbolic Markov property and then apply the Bowen-Sinai refinement. This resulting partition, being the refinement of the possibly infinite family $\mathscr{Z}$, could be uncountable (for example, the dyadic intervals in $[0,1]$ form a countable family whose refinement is the point partition). To avoid this, we require $\mathscr{Z}$ to be *locally finite*: every $Z \in \mathscr{Z}$ intersects finitely many others $Z' \in \mathscr{Z}$. The local finiteness is consequence of the *inverse theorem*: if $\underline{v}, \underline{w} \in \Sigma^\#$ and $\pi(\underline{v}) = \pi(\underline{w})$, then the parameters of $\underline{v}$ and $\underline{w}$ are comparable.

Theorem 6.1 (Inverse theorem [Sar13]). *Let* $\underline{v} = \{\Psi_{x_n}^{p_n^s,p_n^u}\}_{n\in\mathbb{Z}}, \underline{w} = \{\Psi_{y_n}^{q_n^s,q_n^u}\}_{n\in\mathbb{Z}}$ *in* $\Sigma^\#$. *If* $\pi(\underline{v}) = \pi(\underline{w})$, *then the following hold for all* $n \in \mathbb{Z}$:

(1) CONTROL OF α: $\frac{\sin\alpha(x_n)}{\sin\alpha(y_n)} = e^{\pm\text{const}}$.
(2) CONTROL OF s, u: $\frac{s(x_n)}{s(y_n)} = \frac{u(x_n)}{u(y_n)} = e^{\pm\text{const}}$.
(3) CONTROL OF p^s, p^u: $\frac{p_n^s}{q_n^s} = \frac{p_n^u}{q_n^u} = e^{\pm\text{const}}$.

Let $x = \pi(\underline{v}) = \pi(\underline{w})$. Below we sketch the proof Theorem 6.1 for $n = 0$.

Control of α. Since $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$, and since $V^{s/u}[\underline{v}]$ is almost parallel to $E_{x_0}^{s/u}$, we have $\alpha(x_0) \approx \alpha(x)$. Similarly $\alpha(y_0) \approx \alpha(x)$, hence $\alpha(x_0) \approx \alpha(y_0)$.

Control of s, u. This is the more delicate part. We show that $s(x), u(x)$ are finite, then we compare $s(x_0), u(x_0)$ with $s(x), u(x)$, as follows:

- DEFINITION OF $s(V^s[\underline{v}])$: the relevance of each vertex and the recurrence of $\underline{v}$ imply that $s(z) < \infty$ for every $z \in V^s[\underline{v}]$. By hyperbolicity, $\frac{s(z)}{s(z')} = e^{\pm\text{const}}$ for all $z, z' \in V^s[\underline{v}]$. Define $s(V^s[\underline{v}]) := s(\Psi_{x_0}(0, F(0)))$, where F is the representing function of $V^s[\underline{v}]$.
- RATIO IMPROVEMENT: if $v \to w$ and $V_1 \in \mathscr{M}_w^s$ s.t. $s(V_1) < \infty$ then $s(\mathscr{F}_{v,w}^s[V_1]) < \infty$. Moreover, if $V_2 \in \mathscr{M}_w^s$ s.t. $s(V_2) < \infty$ with $\frac{s(V_1)}{s(V_2)} = e^{\pm\rho}$ for $\rho \gg 1$ then $\frac{s(\mathscr{F}_{v,w}^s[V_1])}{s(\mathscr{F}_{v,w}^s[V_2])} = e^{\pm(\rho - Q_\varepsilon(x)^{\beta/4})}$, i.e. the ratio improves. Using the recurrence of $\underline{v}$ we conclude that $\frac{s(x_0)}{s(V^s[\underline{v}])} = e^{\pm\text{const}}$.

Similarly we have $\frac{s(y_0)}{s(V^s[\underline{w}])} = e^{\pm\text{const}}$, hence $\frac{s(x_0)}{s(y_0)} = e^{\pm\text{const}}$. An analogous argument for the unstable graph transforms gives that $\frac{u(x_0)}{u(y_0)} = e^{\pm\text{const}}$.

Note that parts (1)–(2) imply that $\frac{Q_\varepsilon(x_0)}{Q_\varepsilon(y_0)} = e^{\pm\text{const}}$.

Control of p^s, p^u. We show that are p_n^s, p_n^u are maximal infinitely often.

Lemma 6.2. *If* $\underline{v} = \{\Psi_{x_n}^{p_n^s,p_n^u}\}_{n\in\mathbb{Z}} \in \Sigma^\#$, *then* $p_n^s = \delta_\varepsilon Q_\varepsilon(x_n)$ *for infinitely many* $n > 0$, *and* $p_n^u = \delta_\varepsilon Q_\varepsilon(x_n)$ *for infinitely many* $n < 0$.

Proof. This follows from (GPO2) and the recurrence of $\underline{v}$. We prove the first part. If it is not true, then $\exists n > 0$ s.t. $p_m^s = e^\varepsilon p_{m+1}^s$ for all $m \geq n$, hence $p_n^s = e^{\varepsilon(m-n)} p_m^s$ for all $m \geq n$. This is a contradiction, since $\limsup_{m\to\infty} p_m^s > 0$. □

We check that $p_n^s \geq e^{-\text{const}} q_n^s$ for indices n satisfying the previous lemma, then we propagate this inequality to smaller indices. If $p_n^s = \delta_\varepsilon Q_\varepsilon(x_n)$, then $p_n^s = \delta\varepsilon Q_\varepsilon(x_n) \geq e^{-\text{const}} \delta_\varepsilon Q_\varepsilon(y_n) \geq e^{-\text{const}} q_n^s$. Now, by (GPO2) we have

$$p_{n-1}^s = \min\{e^\varepsilon p_n^s, \delta_\varepsilon Q_\varepsilon(x_n)\} \geq e^{-\text{const}} \min\{e^\varepsilon q_n^s, \delta_\varepsilon Q_\varepsilon(y_n)\} = e^{-\text{const}} q_{n-1}^s.$$

By induction, $p_m^s \geq e^{-\text{const}} q_m^s$ for all $m \leq n$. Since n can be taken arbitrarily large, we conclude that $p_m^s \geq e^{-\text{const}} q_m^s$ for all $m \in \mathbb{Z}$.

A locally finite Markov cover [Sar13]**.** We now use the map $\pi : \Sigma \to M$ defined in the previous section and Theorem 6.1 to construct a locally finite family of rectangles in M. This family is $\mathscr{Z} := \{Z_v : v \in \mathscr{A}\}$, where

$$Z_v := \{\pi(\underline{v}) : \underline{v} \in \Sigma^\# \text{ and } v_0 = v\}.$$

Each Z_v is a rectangle with respect to the Smale bracket. Local finiteness follows from Theorem 5.7(1) and Theorem 6.1(3): if $v = \Psi_x^{p^s,p^u}$ and $w = \Psi_y^{q^s,q^u}$ satisfy $Z_v \cap Z_w \neq \varnothing$ then $\frac{p^s}{q^s} = \frac{p^u}{q^u} = e^{\pm\text{const}}$, hence

$$\#\{Z' \in \mathscr{Z} : Z \cap Z' \neq \varnothing\} \leq \#\{w \in \mathscr{A} : q^s, q^u \geq e^{-\text{const}}(p^s \wedge p^u)\} < \infty.$$

The family $\mathscr{Z}$ satisfies a symbolic Markov property, inherited from Σ: if $\underline{v} \in \Sigma^\#$ and $\pi(\underline{v}) = x$, then

$$f(W^s(x, Z_{v_0})) \subset W^s(f(x), Z_{v_1}) \text{ and } f^{-1}(W^u(f(x), Z_{v_1})) \subset W^s(x, Z_{v_0}).$$

A refinement procedure. Since rectangles Z_v are usually neither closed nor empty, we apply a refinement method more abstract than in the uniformly hyperbolic case, not paying attention to relative interiors/closures. The price we pay is that we have less information on the resulting partition.

For $Z, Z' \in \mathscr{Z}$ s.t. $Z \cap Z' \neq \varnothing$, let $\mathscr{E}_{ZZ'}$ = cover of Z by rectangles:

$$\begin{aligned}
E^{su} &= \{x \in Z : W^s(x,Z) \cap Z' \neq \varnothing, W^u(x,Z) \cap Z' \neq \varnothing\}\\
E^{s\varnothing} &= \{x \in Z : W^s(x,Z) \cap Z' \neq \varnothing, W^u(x,Z) \cap Z' = \varnothing\}\\
E^{\varnothing u} &= \{x \in Z : W^s(x,Z) \cap Z' = \varnothing, W^u(x,Z) \cap Z' \neq \varnothing\}\\
E^{\varnothing\varnothing} &= \{x \in Z : W^s(x,Z) \cap Z' = \varnothing, W^u(x,Z) \cap Z' = \varnothing\}.
\end{aligned}$$

Let $\mathscr{R}$ be the partition that refines all of $\mathscr{E}_{ZZ'}$. Then $\mathscr{R}$ is a *Markov partition*:

(1) Product structure: If $x, y \in R \in \mathscr{R}$ then $[x,y] \in R$.
(2) Markov property: if $R, S \in \mathscr{R}$ and if $x \in R, f(x) \in S$ then:

$$f(W^s(x,R)) \subset W^s(f(x),S) \text{ and } f^{-1}(W^u(f(x),S)) \subset W^u(x,R).$$

This notion of Markov partition is weaker than the one considered for uniformly hyperbolic diffeomorphisms, nevertheless it generates a coding as before: $\widehat{\Sigma} = \Sigma(\widehat{V}, \widehat{E})$ where $\widehat{V} = \mathscr{R}$ and $R \to S$ if $f(R) \cap S \neq \varnothing$. We can therefore define $\widehat{\pi} : \widehat{\Sigma} \to M$ by

$$\widehat{\pi}(\underline{R}) := \bigcap_{n\geq 0} \overline{f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n)}.$$

The map $\widehat{\pi}$ is finite-to-one on the set $\widehat{\pi}[\widehat{\Sigma}^\#]$, therefore $(\widehat{\Sigma}, \widehat{\sigma}, \widehat{\pi})$ is a symbolic model for f. This completes the proof of Theorem 5.1.

Applications. Assume $h = h_{\rm top}(f) > 0$. Here are applications of Theorem 5.1:

- If f has a measure of maximal entropy, then $\limsup e^{-nh}\text{Per}_n(f) > 0$ [Sar13]. This improves an estimate obtained by Katok [Kat80].
- f has at most countably many ergodic measures of maximal entropy [Sar13].
- If μ is an ergodic equilibrium measure of a Hölder potential with $h_\mu(f) > 0$, then (f,μ) is either a Bernoulli automorphism or a Bernoulli automorphism times a finite rotation [Sar11].
- Buzzi-Crovisier-Sarig 2016: if f is transitive and C^∞ then it has a unique measure of maximal entropy.

Uniform hyperbolicity versus non-uniform hyperbolicity. Below is a summary of the main differences in the construction of symbolic dynamics for uniformly hyperbolic systems and non-uniformly hyperbolic ones.

	UNIFORMLY HYPERBOLIC	NON-UNIFORMLY HYPERBOLIC
Regularity	$f \in C^1$	$f \in C^{1+\beta}$
Coding	All points	μ–almost every point
Chart	Lyapunov chart: uniform size	Pesin chart: size Q_ε with $\frac{1}{n}\log Q_\varepsilon(f^n(x)) \to 0$
Vertex set	Finite number of Lyapunov charts	Countable number of ε–double charts
Edges	$\Psi_x \to \Psi_y$: $f(x) \approx y$ and $f^{-1}(y) \approx x$	$\Psi_x^{p^s,p^u} \overset{\varepsilon}{\to} \Psi_y^{q^s,q^u}$: (GPO1) and (GPO2)
Repr. function	$F \in C^1$ s.t. $\lvert F(0)\rvert < \varepsilon$ and $\lVert F'\rVert_{C^0} \le 1$	$F \in C^{1+\beta/3}$ s.t. (AM1)–(AM3)
$\pi : \Sigma \to M$	$\{\pi(\underline{v})\} = V^s[\underline{v}] \cap V^u[\underline{v}]$	Same
Cover $\mathscr{Z}$	$Z_v = \{\pi(\underline{v}) : v_0 = v\}$ are closed	$Z_v = \{\pi(\underline{v}) : \underline{v} \in \Sigma^\#, v_0 = v\}$ not open nor closed
Refinement	Relative interiors and closures	Set-theoretical refinement
Partition $\mathscr{R}$	Markov in the sense of Bowen	Markov in the sense of Sinaĭ
Graph $(\widehat{V}, \widehat{E})$	$\widehat{V} = \mathscr{R}$, $\widehat{E} = \{R \to S$: $f(R^*) \cap S^* \neq \varnothing\}$	$\widehat{V} = \mathscr{R}$, $\widehat{E} = \{R \to S$: $f(R) \cap S \neq \varnothing\}$
$\widehat{\pi} : \widehat{\Sigma} \to M$	$\{\widehat{\pi}(\underline{R})\} = \bigcap_{n\ge 0} f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n)$	$\{\widehat{\pi}(\underline{R})\} = \bigcap_{n\ge 0} \overline{f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n)}$
Finite-to-one	$\lvert\widehat{\pi}^{-1}(x)\rvert \le \lvert\mathscr{R}\rvert^2$, $\forall x \in M$	$\lvert\widehat{\pi}^{-1}(x) \cap \widehat{\Sigma}^\#\rvert < \infty$, $\forall x \in \widehat{\pi}(\widehat{\Sigma}^\#)$

7. Surface maps with discontinuities

Theorem 5.1 does not cover the case when at least one of the following conditions hold: M is not closed, f has discontinuities, df is not uniformly bounded. The main examples where (all) these conditions hold are *billiard maps.* Given a compact domain $T \subset \mathbb{R}^2$ with piecewise smooth boundary, consider the straight line motion of a particle inside T, with specular reflections in ∂T. Let $f : M \to M$ be the *billiard map*, where $M = \partial T \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with the convention that $(r,\theta) \in M$ represents r = collision position at ∂T and θ = angle of collision. The map f has a natural invariant Liouville measure $d\mu = \cos\theta dr d\theta$. Sinaĭ proved that for dispersing billiards μ is ergodic [Sin70]. Bunimovich constructed examples of ergodic nowhere dispersing billiards [Bun74a, Bun74b, Bun79]. These billiards, known as *Bunimovich billiards*, are non-uniformly hyperbolic: μ–almost every point has one positive Lyapunov exponent and one negative Lyapunov exponent, see [CM06, Chapter 8].

It is possible to adapt the methods of [Sar13] to the context of billiards [LM]. Let M be a smooth Riemannian surface with finite diameter, possibly with boundary. We assume that the diameter of M is smaller than one. Let $\mathscr{D}^+, \mathscr{D}^-$ be closed subsets of M. Fix $f : M\backslash\mathscr{D}^+ \to M$ a diffeomorphism onto its image, s.t. f has an inverse $f^{-1} : M\backslash\mathscr{D}^- \to M$ that is a diffeomorphism onto its image. Let $\mathscr{D} := \mathscr{D}^+ \cup \mathscr{D}^-$ be the *set of discontinuities of* f. We require some regularity conditions on M, f. The first four assumptions are on the geometry of M. Given $x \in M\backslash\mathscr{D}$, let $\mathrm{inj}(x)$ be the injectivity radius of M at x, and let $\exp_x$ be the *exponential map* at x. Given $r > 0$, let $B_x[r] \subset T_xM$ be the ball with center 0 and radius r. The Riemannian metric on M induces a Riemannian metric on TM, called the *Sasaki metric*, see e.g. [BMW12, §2]. Denote the Sasaki metric by $d_{\rm Sas}(\cdot,\cdot)$. Similarly, we denote the Sasaki metric on $TB_x[r]$ by the same notation, and the context will be clear in which space we are. For nearby small vectors, the Sasaki metric is almost a product metric in the following sense. Given a geodesic γ joining y to x, let $P_\gamma : T_yM \to T_xM$ be the parallel transport along γ. If $v \in T_xM$, $w \in T_yM$ then $d_{\rm Sas}(v,w) \asymp d(x,y) + \|v - P_\gamma w\|$ as $d_{\rm Sas}(v,w) \to 0$, see e.g. [BMW12, Appendix A]. The rate of convergence depends on the curvature tensor of the metric on M. Here are the first assumptions on M.

Regularity of $\exp_x$: $\exists a > 1$ s.t. for all $x \in M\backslash\mathscr{D}$ there is $d(x,\mathscr{D})^a < \mathfrak{r}(x) < 1$ s.t. for $D_x := B(x, 2\mathfrak{r}(x))$ the following holds:

(A1) If $y \in D_x$ then $\mathrm{inj}(y) \geq 2\mathfrak{r}(x)$, $\exp_y^{-1} : D_x \to T_yM$ is a diffeomorphism onto its image, and $\frac{1}{2}(d(x,y) + \|v - P_{y,x}w\|) \leq d_{\rm Sas}(v,w) \leq 2(d(x,y) + \|v - P_{y,x}w\|)$ for all $y \in D_x$ and $v \in T_xM, w \in T_yM$ s.t. $\|v\|, \|w\| \leq 2\mathfrak{r}(x)$, where $P_{y,x} := P_\gamma$ is the radial geodesic γ joining y to x.

(A2) If $y_1, y_2 \in D_x$ then $d(\exp_{y_1} v_1, \exp_{y_2} v_2) \leq 2 d_{\rm Sas}(v_1, v_2)$ for $\|v_1\|$, $\|v_2\| \leq 2\mathfrak{r}(x)$, and $d_{\rm Sas}(\exp_{y_1}^{-1} z_1, \exp_{y_2}^{-1} z_2) \leq 2[d(y_1,y_2) + d(z_1,z_2)]$ for $z_1, z_2 \in D_x$ wherever the expression makes sense. In particular $\|d(\exp_x)_v\| \leq 2$ for $\|v\| \leq 2\mathfrak{r}(x)$, and $\|d(\exp_x^{-1})_y\| \leq 2$ for $y \in D_x$.

The next two assumptions are on the regularity of $d\exp_x$. For $x, x' \in M\backslash\mathscr{D}$, let $\mathscr{L}_{x,x'} := \{A : T_xM \to T_{x'}M : A \text{ is linear}\}$ and $\mathscr{L}_x := \mathscr{L}_{x,x}$. Then the parallel transport $P_{y,x}$ considered in (A1) is in $\mathscr{L}_{y,x}$. Given $y \in D_x, z \in D_{x'}$ and $A \in \mathscr{L}_{y,z}$, let $\widetilde{A} \in \mathscr{L}_{x,x'}$, $\widetilde{A} := P_{z,x'} \circ A \circ P_{x,y}$. By definition, $\widetilde{A}$ depends on x, x' but different basepoints define a map that differs from $\widetilde{A}$ by pre and post composition with isometries. In particular, $\|\widetilde{A}\|$ does not depend on the choice of x, x'. Similarly, if $A_i \in \mathscr{L}_{y_i,z_i}$ then $\|\widetilde{A_1} - \widetilde{A_2}\|$ does not depend on the choice of x, x'. Define the map $\tau = \tau_x : D_x \times D_x \to \mathscr{L}_x$ by $\tau(y,z) = d\widetilde{(\exp_y^{-1})}_z$, where we use the identification $T_v(T_yM) \cong T_yM$ for all $v \in T_yM$.

Regularity of $d\exp_x$:

(A3) If $y_1, y_2 \in D_x$ then $\|d\widetilde{(\exp_{y_1})}_{v_1} - d\widetilde{(\exp_{y_2})}_{v_2}\| \leq d(x,\mathscr{D})^{-a} d_{\rm Sas}(v_1,v_2)$ for all $\|v_1\|, \|v_2\| \leq 2\mathfrak{r}(x)$, and $\|\tau(y_1,z_1) - \tau(y_2,z_2)\| \leq d(x,\mathscr{D})^{-a}[d(y_1,y_2) + d(z_1,z_2)]$ for all $z_1, z_2 \in D_x$.

(A4) If $y_1, y_2 \in D_x$ then the map $\tau(y_1,\cdot) - \tau(y_2,\cdot) : D_x \to \mathscr{L}_x$ has Lipschitz constant $\leq d(x,\mathscr{D})^{-a} d(y_1,y_2)$.

Conditions (A1)–(A2) guarantee that the exponential maps and their inverses are well-defined and have uniformly bounded Lipschitz constants in balls of radii $d(x,\mathscr{D})^a$. Condition (A3) controls the Lipschitz constants of the derivatives of these maps, and condition (A4) controls the Lipschitz constants of their second derivatives. Here are some case when (A1)–(A4) are satisfied, in increasing order of generality:

- The curvature tensor R of M is globally bounded, e.g. when M is the phase space of a billiard map.
- $R, \nabla R, \nabla^2 R, \nabla^3 R$ grow at most polynomially fast with respect to the distance to $\mathscr{D}$, e.g. when M is a moduli space of curves equipped with the Weil-Petersson metric [BMW12].

Regularity of f: There are constants $0 < \beta < 1 < b$ s.t. for all $x \in M \backslash \mathscr{D}$:

(A5) If $y \in D_x$ then $\|df_y^{\pm 1}\| \le d(x, \mathscr{D})^{-b}$.

(A6) If $y_1, y_2 \in D_x$ and $f(y_1), f(y_2) \in D_{x'}$ then $\|\widetilde{df_{y_1}} - \widetilde{df_{y_2}}\| \le \mathfrak{K} d(y_1, y_2)^\beta$, and if $y_1, y_2 \in D_x$ and $f^{-1}(y_1), f^{-1}(y_2) \in D_{x''}$ then $\|\widetilde{df^{-1}_{y_1}} - \widetilde{df^{-1}_{y_2}}\| \le \mathfrak{K} d(y_1, y_2)^\beta$.

Although technical, conditions (A5)–(A6) hold in most cases of interest, e.g. if $\|df^{\pm 1}\|, \|d^2 f^{\pm 1}\|$ grow at most polynomially fast with respect to the distance to $\mathscr{D}$. We finally define the measures we code. Fix $\chi > 0$.

χ–hyperbolic measure: An f–invariant probability measure on M is called *χ–hyperbolic* if μ–a.e. $x \in M$ has one Lyapunov exponent $> \chi$ and another $< -\chi$.

f–adapted measure: An f–invariant measure on M is called *f–adapted* if

$$\int_M \log d(x, \mathscr{D}) d\mu(x) > -\infty.$$

A fortiori $\mu(\mathscr{D}) = 0$.

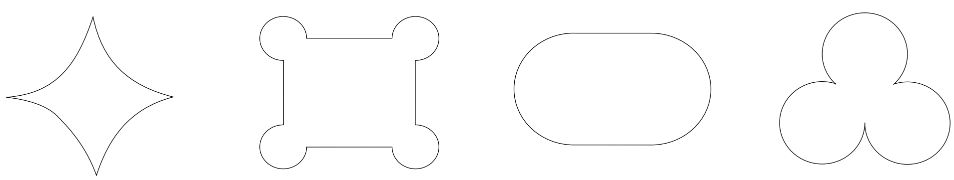
Theorem 7.1. *Let M, f satisfy* (A1)–(A6)*. For all $\chi > 0$, there exists a TMS (Σ, σ) and a Hölder continuous map $\pi : \Sigma \to M$ s.t.:*

(1) $\pi \circ \sigma = f \circ \pi$.

(2) *$\pi[\Sigma^\#]$ has full μ–measure for every f–adapted χ–hyperbolic measure μ.*

(3) *For all $x \in \pi[\Sigma^\#]$, $\#\{\underline{v} \in \Sigma^\# : \pi(\underline{v}) = x\} < \infty$.*

Theorem 7.1 applies to Sinai billiards and to Bunimovich billiards, see figure 5.



(a) (b) (c) (d)

Figure 5. (a) is a Sinai billiard table. The others are Bunimovich billiard tables: (b) is the pool table with pockets, (c) is the stadium, (d) is the flower.

The proof of Theorem 7.1 requires a finer definition of Pesin charts: their images cannot intersect $\mathscr{D}$, and the distortions caused by f have to be controlled. Since the derivative grows at most polynomially fast with the distance to $\mathscr{D}$, we can define Q_ε with the temperedness property: $Q_\varepsilon(x)$ is largest element of I_ε that is

$$\le \min\left\{\|C_\chi(x)^{-1}\|_{\text{Frob}}^{-24/\beta}, \|C_\chi(f(x))^{-1}\|_{\text{Frob}}^{-12/\beta} \rho(x)^{72a/\beta}\right\}.$$

Above, $\rho(x) = d(\{f^{-1}(x), x, f(x)\}, \mathscr{D})$. Note that this definition is stronger than the one of [Sar13].

8. Three dimensional flows with positive speed

The next result we want to discuss is a version of Theorem 5.1 for flows. Let M be a three dimensional closed smooth Riemannian manifold, and $\varphi : M \to M$ a flow s.t. its generating vector field $X = d\varphi$ is $C^{1+\beta}$ and $X \neq 0$ everywhere. Since $d\varphi^t(X) = X \circ \varphi^t$, the Lyapunov exponent of φ in the direction of X is zero. For $\chi > 0$, we can similarly define χ–hyperbolic measures.

χ–HYPERBOLIC MEASURE: A *χ–hyperbolic measure* is a φ–invariant probability measure s.t. μ–a.e. $x \in M$ has one Lyapunov exponent $> \chi$ and another $< -\chi$.

By the same reason for diffeomorphisms, every ergodic invariant measure with $h_\mu(\varphi) > \chi$ is χ–hyperbolic. Contrary to Theorem 5.1, we code each χ–hyperbolic measure at a time.

Theorem 8.1 ([LS])**.** *Let $\varphi : M \to M$ as above. For each χ–hyperbolic measure μ, there is a TMF (Σ_r, σ_r) and $\pi_r : \Sigma_r \to M$ Hölder continuous s.t.:*
(1) *$\pi_r \circ \sigma_r^t = \varphi^t \circ \pi_r$ for all $t \in \mathbb{R}$.*
(2) *$\pi_r[\Sigma_r^\#]$ has full μ–measure.*
(3) *π_r is finite-to-one on $\pi_r[\Sigma_r^\#]$.*

Above, $\Sigma_r^\# = \{(\underline{v}, t) \in \Sigma_r : \underline{v} \in \Sigma^\#\}$. By the Oseledets theorem, every χ–hyperbolic measure is carried by the set NUH_χ of points $x \in M$ for which there are vectors $\{e^s_{\varphi^t(x)}\}_{t\in\mathbb{R}}$, $\{e^u_{\varphi^t(x)}\}_{t\in\mathbb{R}}$ s.t. for every $y = \varphi^t(x)$ the following hold:

(1) $e_y^{s/u} \in T_yM$, $\|e_y^{s/u}\| = 1$.
(2) $\mathrm{span}(d\varphi^{t'} e_y^{s/u}) = \mathrm{span}(e^{s/u}_{\varphi^{t'}(y)})$ for all $t' \in \mathbb{R}$.
(3) $\lim_{t\to\pm\infty} \frac{1}{t}\log\|d\varphi^t e^s_y\| < -\chi$ and $\lim_{t\to\pm\infty} \frac{1}{t}\log\|d\varphi^t e^u_y\| > \chi$.
(4) $\lim_{t\to\pm\infty} \frac{1}{t}\log|\sin\alpha(\varphi^t(y))| = 0$, where $\alpha(\varphi^t(y)) = \angle(e^s_{\varphi^t(y)}, e^u_{\varphi^t(y)})$.

Flow boxes, standard sections. For $x \in M$ and $r > 0$, consider the disc

$$S_r(x) := \exp_x\{v \in \mathbb{R}^3 : v \perp X(x), \|v\| \leq r\}.$$

There is a constant $\mathfrak{r} = \mathfrak{r}(\varphi)$ s.t. if $r < \mathfrak{r}$ then $S_r(x)$ is an embedded surface on M, transversal to X. In the sequel we always assume that $r < \mathfrak{r}$.

FLOW BOX $\mathrm{FB}(x,r)$: We define the *flow box* $\mathrm{FB}(x,r)$ by

$$\mathrm{FB}(x,r) := \bigcup_{|t|\leq r} \varphi^t[S_r(x)].$$

$\mathrm{FB}(x,r)$ has a simple system of coordinates $(y,t) \in S_r(x) \times [-r,r] \mapsto \varphi^t(y)$. In these coordinates, φ is a unit speed vertical flow.

STANDARD SECTION: A *standard section* is a finite disjoint union $\Lambda = \bigsqcup_{i=1}^N S_r(x_i)$ of discs s.t. $M = \bigcup_{i=1}^N \mathrm{FB}(x_i, r)$.

Standard sections exist [LS, Lemma 2.7]. The idea of proof can be summarized as follows:

- By compactness, M can be covered by finitely many flow boxes $\mathrm{FB}(x_i, R)$.
- If the $S_R(x_i)$'s are disjoint, then we are done. If not, change each $S_R(x_i)$ by $S_R(\varphi^{\varepsilon_i}(x))$ for small parameters ε_i. If there are ε_i's s.t. these discs are disjoint, then we are done.

∘ If not, replace each $S_R(x_i)$ by $\frac{R^2}{r^2}$ discs $S_r(\varphi^{\varepsilon_{ij}}(y_j^{(i)}))$ for $r \ll R$, $y_j^{(i)} \in S_R(x_i)$, and ε_{ij} small. This procedure adds one degree of freedom to the choice of ε_{ij}, and now they can be chosen to make the discs disjoint.

Poincaré return map. Fix a standard section Λ.

Poincaré return map $f : \Lambda \to \Lambda$: It is the map $x \in \Lambda \mapsto \varphi^t(x)$, where t is the smallest positive number s.t. $\varphi^t(x) \in \Lambda$.

The measure μ induces an f–invariant measure μ_Λ on Λ, sometimes called the *flux measure*. μ_Λ relates to μ explicity: if $R : \Lambda \to (0, r]$ is the *return time function* of Λ, $R(x) = \min\{t > 0 : \varphi^t(x) \in \Lambda\}$, then

$$\mu = \frac{1}{\int_\Lambda R d\mu_\Lambda} \int_\Lambda \left[\int_0^{R(x)} \delta_{\varphi^t(x)} dt \right] d\mu_\Lambda(x)$$

where $\delta_{\varphi^t(x)}$ is the Dirac measure at $\varphi^t(x)$.

For $x \in \Lambda \cap \mathrm{NUH}_\chi$, let $v_x^s, v_x^u \in T_x\Lambda$ be unit vectors s.t.

$$v_x^s = \gamma^s(x) e_x^s + n^s(x) X(x)$$
$$v_x^u = \gamma^u(x) e_x^u + n^u(x) X(x)$$

for scalars $\gamma^{s/u}(x), n^{s/u}(x)$. These vectors exist because Λ is tranversal to X. We can explicitly calculate the Lyapunov exponents of f in the directions v^s and v^u and conclude the following.

Lemma 8.2 ([LS]). *μ_Λ is χ'–hyperbolic for f, where $\chi' = \chi \inf R$.*

At this level, we would like to apply the method of [Sar13] to f, but unfortunately f is not $C^{1+\beta}$. Actually f is not even continuous, because Λ has boundaries.

Regular and singular sets of f: The *regular set* Λ' of f is the set of $x \in \Lambda$ that posses a neighborhood $V \subset \Lambda \backslash \partial\Lambda$ s.t. the restrictions $f \restriction_V : V \to f(V)$ and $f^{-1} \restriction_V : V \to f^{-1}(V)$ are diffeomorphisms onto their images. The *singular set* is $\mathfrak{S}(\Lambda) = \Lambda \backslash \Lambda'$.

The good news is that f is well-behaved inside Λ'.

Lemma 8.3 ([LS]). *R, f, f^{-1} are differentiable on Λ', and $\exists \mathfrak{C}$ only depending on M and φ s.t. $\sup_{x\in\Lambda'} \|dR_x\| < \mathfrak{C}$, $\sup_{x\in\Lambda'} \|df_x\| < \mathfrak{C}$, $\sup_{x\in\Lambda'} \|(df_x)^{-1}\| < \mathfrak{C}$, $\|f|_U\|_{C^{1+\beta}} < \mathfrak{C}$ and $\|f^{-1}|_U\|_{C^{1+\beta}} < \mathfrak{C}$ for all open and connected $U \subset \Lambda'$.*

Adapted sections. Another potential problem is when $\mu_\Lambda[\mathfrak{S}(\Lambda)] > 0$ (actually, in Lemma 8.2 we needed to require that $\mu_\Lambda[\mathfrak{S}(\Lambda)] = 0$). The requirement $\mu_\Lambda[\mathfrak{S}(\Lambda)] = 0$ is not enough: we also want almost all trajectories of f *not* to converge exponentially fast to $\mathfrak{S}(\Lambda)$. If so, then we can define $Q_\varepsilon(x)$ satisfying an analogue of Lemma 5.3.

Adapted section: A standard section Λ is called *adapted* if:

(1) $\mu_\Lambda[\mathfrak{S}(\Lambda)] = 0$.
(2) $\lim_{n\to\pm\infty} \frac{1}{n} \log \mathrm{dist}(f^n(x), \mathfrak{S}(\Lambda)) = 0$ for μ_Λ–a.e. $x \in \Lambda$.

There is no reason for a single standard section to be adapted. To construct an adapted section, we consider a 1–parameter family of standard sections $\{\Lambda_r\}$ and show that $\exists r$ s.t. Λ_r is adapted.

Theorem 8.4 ([LS])**.** *If $\Lambda_r=\bigsqcup_{i=1}^N S_r(x_i)$ is a standard section for all $r\in[a,b]$, then Λ_r is adapted for Lebesgue–a.e. $r\in[a,b]$.*

Proof. Let $f_r=f_{\Lambda_r}$ and $\mathfrak{S}_r=\mathfrak{S}(\Lambda_r)$. It is enough to show that

$$\mu_{\Lambda_r}\left\{x\in\Lambda_r:\liminf_{|n|\to\infty}\frac{1}{|n|}\log\operatorname{dist}(f_r^n(x),\mathfrak{S}_r)<0\right\}=0\text{ for a.e. }r\in[a,b].\tag{8.1}$$

For $\alpha>0$, let

$$A_\alpha(r):=\{x\in\Lambda_b:\exists\text{ infinitely many }n\in\mathbb{Z}\text{ s.t. }\tfrac{1}{|n|}\log\operatorname{dist}(f_b^n(x),\mathfrak{S}_r)<-\alpha\}.$$

Hence (8.1) follows from the statement

$$\forall\ \alpha>0\text{ rational }\big(\mu_{\Lambda_b}[A_\alpha(r)]=0\text{ for a.e. }r\in[a,b]\big).\tag{8.2}$$

Let $I_\alpha(x):=\{a\le r\le b:x\in A_\alpha(r)\}$, then $1_{A_\alpha(r)}(x)=1_{I_\alpha(x)}(r)$, whence by Fubini's Theorem $\int_a^b\mu_{\Lambda_b}[A_\alpha(r)]dr=\int_{\Lambda_b}\text{Leb}[I_\alpha(x)]d\mu_{\Lambda_b}(x)$. So (8.2) follows from

$$\text{Leb}[I_\alpha(x)]=0\text{ for all }x\in\Lambda_b.$$

The set $I_\alpha(x)$ is contained in the lim sup of intervals $\{I_n\}_{n\in\mathbb{Z}}$ with $|I_n|\approx e^{-\alpha|n|}$. Since $\sum_{n\in\mathbb{Z}}e^{-\alpha|n|}<\infty$, the Borel-Cantelli lemma gives that $\text{Leb}[I_\alpha(x)]=0$. □

Fix an adapted section Λ.

THE PARAMETER $Q_\varepsilon(x)$: $Q_\varepsilon(x)$ = largest element of I_ε that is

$$\le\min\{\operatorname{dist}(x,\mathfrak{S}(\Lambda)),e^{3/\beta}\|C_\chi(f(x))^{-1}\|^{-12/\beta}\}.$$

The values Q_ε defined above satisfy Lemma 5.3. Now adapt the methods of [Sar13] to the triple (Λ,f,μ_Λ). Although some modifications are needed (for example in the coarse graining), the core ideas are the same. The final result is Theorem 8.1.

Applications. Assume that $h=h_{\rm top}(\varphi)>0$.

- If φ has a measure of maximal entropy, then $\liminf Te^{-hT}\#\{\text{closed orbits of length}\le T\}>0$ [LS].
- φ has at most countably many ergodic measures of maximal entropy [LS].
- If μ is an ergodic equilibrium measure of a Hölder potential with $h_\mu(\varphi)>0$, then (φ,μ) is either a Bernoulli flow or a Bernoulli flow times a rotational flow [LLS16].

The second application is proved as in the case of diffeomorphisms. The first uses a dichotomy obtained throughout the proof of the third application. In the sequel we discuss how to establish the Bernoulli property.

9. The Bernoulli property

The precise statement we prove is the following one.

Theorem 9.1 ([LLS16])**.** *If $\varphi:M\to M$ is as above, then every ergodic equilibrium measure of a Hölder continuous potential with positive entropy is either a Bernoulli flow or a Bernoulli flow times a rotational flow.*

Corollary 9.2 ([LLS16])**.** *Let S be a closed smooth orientable Riemannian surface, with nonpositive and non-identically zero curvature. Then the geodesic flow of S is Bernoulli with respect to its (unique) measure of maximal entropy.*

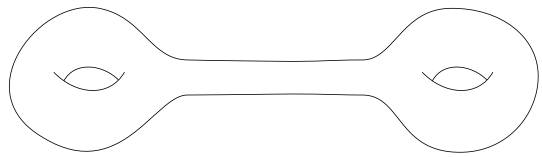


FIGURE 6. A surface with nonpositive and non-identically zero curvature.

Since the Bernoulli property is preserved under factors (this is part of Ornstein's theory), it is enough to prove the Bernoulli property for the TMF $(\Sigma_r, \sigma_r, \widehat{\mu})$ obtained in Theorem 8.1. Here, $\widehat{\mu}$ is the lift of μ to Σ_r. It is an equilibrium measure of a Hölder potential and it has positive entropy. The measure $\widehat{\mu}$ projects[4] to a measure ν on Σ, again an equilibrium measure of a Hölder potential with positive entropy. Such measures were studied in [BS03]. One of their properties is the local product structure.

Bernoulli flows. Let $(X, \mathscr{B}, \mu, \{T^t\})$ be a measurable flow.

BERNOULLI FLOW: $(X, \mathscr{B}, \mu, \{T^t\})$ is called a *Bernoulli flow* if $(X, \mathscr{B}, \mu, T^t)$ is a Bernoulli automorphism for every $t \neq 0$.

Here are some examples of Bernoulli flows:

- Totoki flow: suspension flow over $\{0,1\}^{\mathbb{Z}}$, with $r\restriction_{[0]} \equiv 1$ and $r\restriction_{[1]} \equiv \alpha \notin \mathbb{Q}$, see figure 7.
- Geodesic flows on closed manifolds with negative sectional curvature [OW73, Rat74].
- Sinai billiards and Bunimovich billiards, for their invariant volume measures.

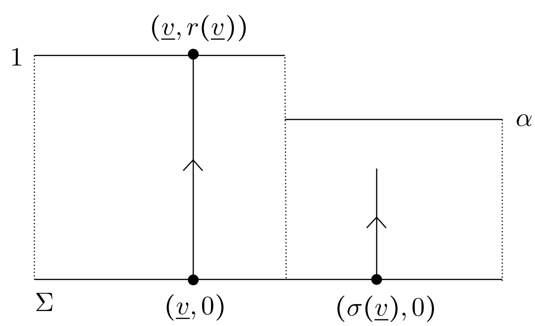


FIGURE 7. Totoki flow: a suspension over the shift on $\{0,1\}^{\mathbb{Z}}$ with uniform Bernoulli measure and piecewise constant roof function.

The Bernoulli property is one the strongest stochastic notions for measure-preserving systems. For instance, we have the following hierarchy:

Ergodic $\subset$ Weak mixing $\subset$ Mixing $\subset$ K property $\subset$ Bernoulli property

[4]Just like μ projects to μ_Λ, $\widehat{\mu}$ projects to a σ–invariant probability measure ν on Σ.

The K property. The intermediate K property is usually established before the Bernoulli property. Let $(X,\mathscr{B},\mu,T)$ be an automorphism.

PINSKER ALGEBRA: The *Pinsker algebra* of $(X,\mathscr{B},\mu,T)$ is

$$\mathscr{P}(T)=\{A\in\mathscr{B}:h_\mu(T,\{A,X\backslash A\})=0\}.$$

K-AUTOMORPHISM: $(X,\mathscr{B},\mu,T)$ is called a *K-automorphism* if $\mathscr{P}(T)=\{\varnothing,X\}$.

The equality of sigma-algebras is modulo μ. An automorphism is K iff it satisfies the Kolmogorov zero-one law, iff tail sigma-algebras of finite partitions are trivial. For systems with hyperbolicity (e.g. Axiom A diffeomorphisms and TMS's), there are two dynamically relevant tail sigma-algebras, associated to partitions ξ^s,ξ^u subordinated to the stable, unstable foliations respectively[5]. A consequence of Rokhlin-Sinai theory is that

$$\mathscr{P}(T)\leq \mathrm{Tail}(\xi^s)\wedge\mathrm{Tail}(\xi^u), \tag{9.1}$$

hence $\mathscr{P}(T)$–measurable functions are μ–a.e. constant along global stable and unstable manifolds.

Now let $(X,\mathscr{B},\mu,\{T^t\})$ be a flow. We have $\mathscr{P}(T^t)=\mathscr{P}(T^1)$ for all $t\neq 0$.

K-FLOW: $(X,\mathscr{B},\mu,\{T^t\})$ is called a *K-flow* if $\mathscr{P}(T^1)=\{\varnothing,X\}$.

The holonomy group. While analyzing mixing properties for flows, the main difficulty is to mix in the flow direction. One way of doing this is to mix along the stable and unstable directions altogether, and push this to the flow direction.

Remeber (Σ_r,σ_r), the TMF obtained in Theorem 8.1. For $\underline{v},\underline{w}\in\Sigma$, write $\underline{v}\sim\underline{w}$ if $\underline{v},\underline{w}$ belong to the same *weak* stable or unstable manifold of σ. Call a sequence $\gamma=\langle\underline{v}^0,\ldots,\underline{v}^n\rangle$ an *su–loop* if $\underline{v}^0=\underline{v}^n$ and $\underline{v}^i\sim\underline{v}^{i+1}$ for all $i=0,\ldots,n-1$. Each *su*–loop γ lifts to a path $\widehat{\gamma}:=\langle(\underline{v}^0,0),(\underline{v}^1,t_1),\ldots,(\underline{v}^n,t_n)\rangle$ on Σ_r s.t. $(\underline{v}^i,t_i),(\underline{v}^{i+1},t_{i+1})$ belong to the same *strong* stable or unstable manifold of σ_r. There is no need for $\widehat{\gamma}$ to be a closed path, see figure 8 below. Define the *weight* of γ as $P(\gamma):=t_n$. It is along these times that we can mix the flow direction.

HOLONOMY GROUP: The *holonomy group* of σ_r is $G:=\overline{\{P(\gamma):\gamma \text{ is a } su\text{–loop}\}}$.

Rigorously, to define G we fix some $\underline{v}\in\Sigma$ and consider only *su*–loops with $\underline{v}^0=\underline{v}$. This definition is the same for ν–a.e. $\underline{v}\in\Sigma$. See [LLS16, §4] for details.

Lemma 9.3 ([LLS16])**.** *G is a closed additive subgroup of $\mathbb{R}$.*

The proof uses the simple observation that the concatenation of *su*–paths γ_1,γ_2 with $\underline{v}^0=\underline{v}$ is an *su*–path γ with $\underline{v}^0=\underline{v}$, and that $P(\gamma)=P(\gamma_1)+P(\gamma_2)$.

A dichotomy theorem. By Lemma 9.3, either $G=\{0\}$ or $G=c\mathbb{Z}$ for some $c\neq 0$ or $G=\mathbb{R}$. The first case never occurs, hence we get the following dichotomy.

Theorem 9.4 ([LLS16])**.** *If $(\Sigma_r,\sigma_r,\widehat{\mu})$ is as above, then either* (a) *or* (b) *below holds:*

(a) *$(\Sigma_r,\sigma_r,\widehat{\mu})$ is measurably conjugate to a constant suspension over a TMS.*
(b) *$(\Sigma_r,\sigma_r,\widehat{\mu})$ is a K-flow.*

[5]Subordination of ξ^s to the stable foliation means that the atoms of ξ^s are local stable manifolds.

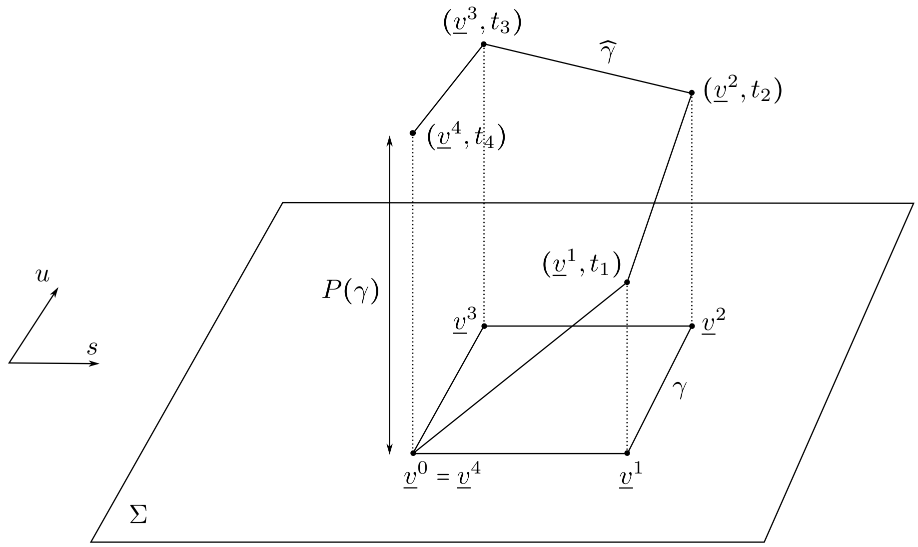


FIGURE 8. The holonomy group G: above γ is an su–loop forming a quadrilateral. The lift $\widehat{\gamma}$ is not necessarily a loop.

Idea of the proof. We analyze the three possibilities for G.

CASE 1: $G = \{0\}$.

Apply the methods of Livschitz theory for partially hyperbolic systems [KK96] to get that r is a measurable coboundary. This is a contradiction, since $\inf r > 0$.

CASE 2: $G = c\mathbb{Z}$ for some $c \neq 0$.

Again applying the ideas of [KK96], we get that r is measurably cohomologous to a function $\widetilde{r} : \Sigma \to c\mathbb{Z}$. When this happens, we can recode the TMS and conclude that (a) holds.

CASE 3: $G = \mathbb{R}$.

Let f be $\mathscr{P}(\sigma_r^1)$–measurable. By (9.1), f is $\mathrm{Tail}(\xi^s) \wedge \mathrm{Tail}(\xi^u)$–measurable. This implies that f is σ_r^t–invariant for every $t \in G$, hence σ_r–invariant. By ergodicity, f is constant. Therefore $\mathscr{P}(\sigma_r^1)$ is the trivial sigma-algebra, and so (b) holds. □

Proof of Theorem 9.4 and Corollary 9.2. We prove firstly Theorem 9.1. If Theorem 9.4(a) holds, then by Ornstein's theory $(\Sigma_r, \sigma_r, \widehat{\mu})$ is measurably conjugate to the product of a Bernoulli flow and a rotational flow. If Theorem 9.4(b) holds, then we adapt the methods of Ratner [Rat74] to construct a generating partition with the very weak Bernoulli property. By Ornstein's theory, this implies that $(\Sigma_r, \sigma_r, \widehat{\mu})$ is a Bernoulli flow.

Remark 9.5. To adapt the methods of [Rat74], we need two facts from the thermodynamical formalism of countable Markov shifts:

- The projection ν of $\widehat{\mu}$ to Σ has local product structure [BS03].
- The g–functions of ν have bounded total variation [Led74, BS03].

To prove Corollary 9.2, we use that geodesic flows are Reeb flows. These flows have the property that the distribution generated by the stable and unstable directions is not integrable. In particular, they do not posses quadrilaterals whose sides are small pieces of stable and unstable manifolds. This implies that $G \neq c\mathbb{Z}$ for all $c \in \mathbb{R}$, therefore $G = \mathbb{R}$ and so Theorem 9.4(b) holds.

Appendix

Uniformly hyperbolic systems. Let M be a closed (compact without boundary) connected orientable smooth Riemannian manifold, and let $f : M \to M$ be a C^1 diffeomporphism.

Anosov diffeomorphism: f is an *Anosov diffeomorphism* if there are $C > 0$, $\lambda < 1$ and a df–invariant decomposition $TM = E^s \oplus E^u$ s.t. $\|df^n v^s\|, \|df^{-n} v^u\| \leq C\lambda^n \|v^s\|$ for all $v^s \in E^s$, $v^u \in E^u$, $n \geq 0$.

Sometimes it is impossible to find a decomposition for every point, one of the reasons being that the relevant part of the dynamics is not all of M.

Non-wandering set $\Omega(f)$: The *non-wandering set* of f is the set $\Omega(f)$ of all $x \in M$ s.t. for every neighborhood $U \ni x$ there exists $n \in \mathbb{Z}$ s.t. $f^n(U) \cap U \neq \varnothing$.

In particular, every periodic point belongs to $\Omega(f)$.

Axiom A diffeomorphism: f is an *Axiom A diffeomorphism* if:

(1) The periodic points are dense in $\Omega(f)$.
(2) There are $C > 0$, $\lambda < 1$ and a df–invariant decomposition $T_{\Omega(f)}M = E^s \oplus E^u$ s.t. $\|df^n v^s\|, \|df^{-n} v^u\| \leq C\lambda^n \|v^s\|$ for all $v^s \in E^s$, $v^u \in E^u$, $n \geq 0$.

Every Anosov diffeomorphism is Axiom A, but not necessarily the opposite (an example is the Smale horseshoe). In these notes Axiom A diffeomorphisms are also called *uniformly hyperbolic*.

If f is uniformly hyperbolic, then there is a way of defining a metric on M, equivalent to $\|\cdot\|$, that satisfies (1)–(2) with $C = 1$, see [Shu87]. A metric with this property is called *adapted*.

Now let $\varphi : M \to M$ be a C^1 flow, and $X = d\varphi$.

Anosov flow: φ is called an *Anosov flow* if $X \neq 0$ everywhere, and if there are $C > 0$, $\lambda < 1$ and a $d\varphi$–invariant decomposition $TM = E^s \oplus \langle X \rangle \oplus E^u$ s.t. $\|d\varphi^t v^s\|, \|d\varphi^{-t} v^u\| \leq C\lambda^t \|v^s\|$ for all $v^s \in E^s$, $v^u \in E^u$, $t \geq 0$.

Axiom A flow: φ is called an *Axiom A flow* if:

(1) The closed orbits are dense in $\Omega(\varphi)$.
(2) $X \neq 0$ in $\Omega(\varphi)$ and there are $C > 0$, $\lambda < 1$ and a df–invariant decomposition $T_{\Omega(\varphi)}M = E^s \oplus \langle X \rangle \oplus E^u$ s.t. $\|d\varphi^t v^s\|, \|d\varphi^{-t} v^u\| \leq C\lambda^t \|v^s\|$ for all $v^s \in E^s$, $v^u \in E^u$, $t \geq 0$.

We also call an Axiom A flow by *uniformly hyperbolic flow*.

Lyapunov exponents. Let $f : M \to M$ be C^1 diffeomorphism, and let μ a f–invariant measure probability measure.

LYAPUNOV EXPONENT: The *Lyapunov exponent* of f in the direction of $v \in T_xM$ is

$$\chi(v) := \lim_{n\to+\infty} \frac{1}{n} \log \|df^n v\|$$

when the limit exists.

The Oseledets theorem guarantees the μ–a.e. existence of the limit.

Theorem 9.6 (Oseledets)**.** *If $f : M \to M$ is as above, then there exists a set $M_0 \subset M$ of full μ–measure s.t., for all $x \in M_0$ and $v \in T_xM$:*

$$\chi(v) = \lim_{n\to\pm\infty} \frac{1}{n} \log \|df^n v\|.$$

NON-UNIFORMLY HYPERBOLIC DIFFEOMORPHISM: (f, μ) is called a *non-uniformly hyperbolic diffeomorphism* if for μ–a.e. $x \in M$ it holds $\chi(v) \neq 0$ for all $v \in T_xM$.

Similar definitions are valid for flows. In these notes we assume more than non-uniform hyperbolicity: we require that the measure is χ–hyperbolic.

Equilibrium measures. Let X be a set, $\mathscr{B}$ a sigma-algebra, $T : X \to X$ a bimeasurable map, and $\psi : X \to \mathbb{R}$ be a function. Given μ a probability measure preserving T, let $h_\mu(T)$ denote its *Kolmogorov-Sinai entropy.*

TOPOLOGICAL PRESSURE: The *topological pressure* of ψ is $h_{\rm top}(\psi) := \sup\{h_\mu(T) + \int \psi d\mu\}$, where the supremum ranges over all T–invariant probability measures.

EQUILIBRIUM MEASURE: An *equilibrium measure* for ψ is a T–invariant probability measure μ s.t. $h_{\rm top}(\psi) = h_\mu(T) + \int \psi d\mu$.

We interpret $-\int \psi d\mu$ as the potential energy of ψ with respect to the distribution μ. Hence an equilibrium measure is a distribution that minimizes the free energy. This is in accordance with the principle of Maupertius: nature minimizes the free energy.

A special case is when $\psi \equiv 0$. An equilibrium measure for this potential is called a *measure of maximal entropy*, for obvious reasons.

REFERENCES

[AW67] R. L. Adler and B. Weiss. Entropy, a complete metric invariant for automorphisms of the torus. *Proc. Nat. Acad. Sci. U.S.A.*, 57:1573–1576, 1967.

[BMW12] K. Burns, H. Masur, and A. Wilkinson. The Weil-Petersson geodesic flow is ergodic. *Ann. of Math. (2)*, 175(2):835–908, 2012.

[Bow70] Rufus Bowen. Markov partitions for Axiom A diffeomorphisms. *Amer. J. Math.*, 92:725–747, 1970.

[Bow73] Rufus Bowen. Symbolic dynamics for hyperbolic flows. *Amer. J. Math.*, 95:429–460, 1973.

[Bow78] Rufus Bowen. Markov partitions are not smooth. *Proc. Amer. Math. Soc.*, 71(1):130–132, 1978.

[Bow08] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.

[Bow75a] Rufus Bowen. Bernoulli equilibrium states for Axiom A diffeomorphisms. *Math. Systems Theory*, 8(4):289–294, 1974/75.

[Bow75b] Rufus Bowen. Some systems with unique equilibrium states. *Math. Systems Theory*, 8(3):193–202, 1974/75.

[BS03] Jérôme Buzzi and Omri Sarig. Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps. *Ergodic Theory Dynam. Systems*, 23(5):1383–1400, 2003.

[Bun74a] L. A. Bunimovič. Billiards that are close to scattering billiards. *Mat. Sb. (N.S.)*, 94(136):49–73, 159, 1974.

[Bun74b] L. A. Bunimovič. The ergodic properties of certain billiards. *Funkcional. Anal. i Priložen.*, 8(3):73–74, 1974.

[Bun79] L. A. Bunimovich. On the ergodic properties of nowhere dispersing billiards. *Comm. Math. Phys.*, 65(3):295–312, 1979.

[CM06] Nikolai Chernov and Roberto Markarian. *Chaotic billiards*, volume 127 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[Kat79] A. Katok. Bernoulli diffeomorphisms on surfaces. *Ann. of Math. (2)*, 110(3):529–547, 1979.

[Kat80] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, (51):137–173, 1980.

[KK96] A. Katok and A. Kononenko. Cocycles' stability for partially hyperbolic systems. *Math. Res. Lett.*, 3(2):191–210, 1996.

[KSLP86] Anatole Katok, Jean-Marie Strelcyn, F. Ledrappier, and F. Przytycki. *Invariant manifolds, entropy and billiards; smooth maps with singularities*, volume 1222 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986.

[Led74] F. Ledrappier. Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 30:185–202, 1974.

[LLS16] François Ledrappier, Yuri Lima, and Omri Sarig. Ergodic properties of equilibrium measures for smooth three dimensional flows. *Comment. Math. Helv.*, 91(1):65–106, 2016.

[LM] Yuri Lima and Carlos Matheus. Symbolic dynamics for non-uniformly hyperbolic surface maps with discontinuities. 35 pages, Preprint available at http://arxiv.org/abs/1606.05863.

[LS] Yuri Lima and Omri Sarig. Symbolic dynamics for three dimensional flows with positive entropy. 48 pages, Preprint available at http://arxiv.org/abs/1408.3427.

[OW73] Donald S. Ornstein and Benjamin Weiss. Geodesic flows are Bernoullian. *Israel J. Math.*, 14:184–198, 1973.

[PP83] William Parry and Mark Pollicott. An analogue of the prime number theorem for closed orbits of Axiom A flows. *Ann. of Math. (2)*, 118(3):573–591, 1983.

[Rat69] M. E. Ratner. Markov decomposition for an U-flow on a three-dimensional manifold. *Mat. Zametki*, 6:693–704, 1969.

[Rat73] M. Ratner. Markov partitions for Anosov flows on n-dimensional manifolds. *Israel J. Math.*, 15:92–114, 1973.

[Rat74] M. Ratner. Anosov flows with Gibbs measures are also Bernoullian. *Israel J. Math.*, 17:380–391, 1974.

[Sar11] Omri M. Sarig. Bernoulli equilibrium states for surface diffeomorphisms. *J. Mod. Dyn.*, 5(3):593–608, 2011.

[Sar13] Omri M. Sarig. Symbolic dynamics for surface diffeomorphisms with positive entropy. *J. Amer. Math. Soc.*, 26(2):341–426, 2013.

[Shu87] Michael Shub. *Global stability of dynamical systems*. Springer-Verlag, New York, 1987. With the collaboration of Albert Fathi and Rémi Langevin, Translated from the French by Joseph Christy.

[Sin68] Ja. G. Sinaĭ. Markov partitions and U-diffeomorphisms. *Funkcional. Anal. i Priložen*, 2(1):64–89, 1968.

[Sin70] Ja. G. Sinaĭ. Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat. Nauk*, 25(2 (152)):141–192, 1970.

[Sin72] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[Sma98] Steve Smale. Finding a horseshoe on the beaches of Rio. *Math. Intelligencer*, 20(1):39–44, 1998.