Ergod. Th. & Dynam. Sys., (2021), **41**, 2591–2658 © The Author(s), 2020. Published by Cambridge University Press.
doi:10.1017/etds.2020.80

SURVEY

Symbolic dynamics for non-uniformly hyperbolic systems

YURI LIMA

Yuri Lima, Departamento de Matemática, Universidade Federal do Ceará (UFC), Campus do Pici, Bloco 914, CEP 60440-900. Fortaleza – CE, Brazil
(e-mail: yurilima@gmail.com)

(*Received* 4 *June* 2020 *and accepted in revised form* 4 *June* 2020)

Abstract. This survey describes the recent advances in the construction of Markov partitions for non-uniformly hyperbolic systems. One important feature of this development comes from a finer theory of non-uniformly hyperbolic systems, which we also describe. The Markov partition defines a symbolic extension that is finite-to-one and onto a non-uniformly hyperbolic locus, and this provides dynamical and statistical consequences such as estimates on the number of closed orbits and properties of equilibrium measures. The class of systems includes diffeomorphisms, flows, and maps with singularities.

Key words: Markov partition, Pesin theory, symbolic dynamics
2020 Mathematics Subject Classification: 37-02, 37B10, 37C05, 37D25 (Primary); 37C35, 37C83, 37D35 (Secondary)

Contents

1. *Introduction*

Markov partitions are a powerful tool in the modern theories of dynamical systems and ergodic theory. They were introduced to these fields at the end of the 1960s (see the foundational works by Adler and Weiss and by Sinaĭ [**AW67**, **AW70**, **Sin68a**, **Sin68b**] and references therein) and have played a crucial role ever since. Roughly speaking, a Markov partition is a partitioning of the phase space of a system into pieces, which allows trajectories to be represented by paths on a graph. The dynamics of paths on a graph is much simpler to understand, and many of its statistical properties can therefore be pushed to the original dynamical system. This approach was extensively developed in the late 1960s and early 1970s for uniformly hyperbolic systems, and its consequences include many breakthroughs in smooth ergodic theory. The method developed by Bowen [**Bow08**] will be of particular importance to us: locally representing the dynamics as a small perturbation of a hyperbolic matrix, he employed the theory of pseudo-orbits used by Anosov [**Ano70**] and by himself [**Bow70b**, **Bow71**, **Bow72b**] to obtain a Markov cover and then refine it into a Markov partition. Due to uniform hyperbolicity, the Markov partitions are finite.

Following Bowen's philosophy, Katok showed that a hyperbolic ergodic measure that is invariant under a $C^{1+\beta}$ diffeomorphism has horseshoes approximating its entropy. A measure is hyperbolic if its Lyapunov exponents are non-zero, and this introduces the concept of *non-uniform hyperbolicity*: the hyperbolicity is not necessarily observed at every iteration but only on average. In the late 1970s, Pesin developed a global theory to treat $C^{1+\beta}$ non-uniformly hyperbolic systems [**Pes76**, **Pes77a**, **Pes77b**], nowadays known as *Pesin theory*; see the book [**BP07**]. Pesin's idea was to construct local charts, nowadays called *Pesin charts*, to represent the dynamics of a non-uniformly hyperbolic diffeomorphism again as a small perturbation of a hyperbolic matrix. The difference from the uniformly hyperbolic situation is that the domain of the Pesin chart is no longer uniform in size and depends on the quality of hyperbolicity at the point. In [**Kat80**], Katok combined Pesin theory with a fine theory of pseudo-orbits and, to avoid the possible degeneracy of Pesin charts, restricted the analysis to *Pesin blocks*, which are non-invariant subsets of the phase space where non-uniform hyperbolicity is essentially uniform. For details, see the supplementary chapter by Katok and Mendoza [**KH95**]. Since a horseshoe naturally carries a Markov partition, Katok's result can be seen as the construction of finite Markov partitions that approximate the topological entropy. The applications using what are now called *Katok horseshoes* are countless. Nevertheless, this approach is not genuinely non-uniformly hyperbolic, since it does not treat at once regions where the degeneracy of Pesin charts occurs. In other words, a single Pesin block does not encompass the whole dynamics (for instance, it usually does not have full topological entropy).

This difficulty stood unresolved for many years, until Sarig recently bypassed it, constructing countable Markov partitions with full topological entropy for $C^{1+\beta}$ surface diffeomorphisms [**Sar13**]. His methods are more suitable for adaptations and

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

generalizations, and are now being further refined in settings outside the scope of the previous theory, such as billiard maps. Here are some of the developments:

- Lima and Sarig for three-dimensional flows without fixed points [**LS19**];
- Lima and Matheus for surface maps with discontinuities [**LM18**];
- Ben Ovadia for diffeomorphisms in any dimension [**BO18**].

Now the Markov partitions are countable. This is unavoidable to treat the regions where the Pesin charts degenerate. Not only are these latter results stronger than the previous ones in the literature, they also cover much broader classes of examples such as geodesic flows on surfaces with non-positive curvature and Bunimovich stadia. Using them, many dynamical and statistical properties have been established: counting on the number of closed orbits [**BD20, BO18, Buz20, LM18, LS19, Sar13**], counting on the number of measures of maximal entropy [**BCS18, BO18, LS19, Sar13**], ergodic properties of equilibrium measures [**LLS16, Sar11**], the almost Borel structure of surface diffeomorphisms [**BB17**], and the generic simplicity of the Lyapunov spectrum of non-uniformly hyperbolic diffeomorphisms [**BPVL20**].

Generally speaking, countable Markov partitions are indeed necessary to code non-uniformly hyperbolic systems: while the set of topological entropies associated to finite Markov partitions is countable, the set of topological entropies of non-uniformly hyperbolic systems is $[0, \infty)$. This occurs, for example, among C^∞ diffeomorphisms in surfaces, where the topological entropy is continuous [**New89**, Theorem 6]. For instance, consider the two-dimensional disc: the identity map has zero entropy, and Smale's horseshoe has topological entropy equal to $\log 2$ (see §2.1.1 for details on this latter example). Since these maps are homotopic, the set of values for the topological entropy contains $[0, \log 2]$ and thus, taking powers, it is equal to $[0, \infty)$. The same occurs for C^∞ diffeomorphisms in the two-dimensional torus: in the one-parameter family of standard maps $f_k(x, y) = (-y + 2x + k \sin(2\pi x), x)$, the topological entropy reaches arbitrarily large values [**Dua94**].

As already mentioned, Markov partitions provide many dynamical and statistical consequences because the dynamics of paths on a graph is simple to understand. In general, any partition generates a symbolic representation of the system, given by the shift map acting on a subset of the symbolic space of paths on a graph. For Markov partitions, such symbolic representation is defined not only on a subset but on the whole space of paths on the graph. This is already a big advantage, but for effectiveness of applications it is important to understand the *coding map*, which relates real trajectories to paths on the graph. If, for instance, the coding map is finite-to-one (that is, every point has finitely many pre-images) then measures on the original system are related to measures on the symbolic space, and the relation preserves entropy (by the Abramov–Rokhlin formula). This happens for uniformly hyperbolic systems almost automatically, but constitutes a major difficulty for non-uniformly hyperbolic ones. Indeed, all previous attempts before Sarig failed exactly at this point. Sarig did not prove that the coding map is finite-to-one, but that it is *morally* finite-to-one: after passing to recurrent subsets (defined by some recurrence assumptions), the coding map is finite-to-one. This was the motivation to use Pesin theory in a much finer way, which has a central importance in the recent constructions of Markov partitions.

Having this in mind, this survey has two main goals:

- to discuss the theory of non-uniformly hyperbolic systems;
- to use this theory to construct countable Markov partitions that generate finite-to-one coding maps.

Since the main reason to construct Markov partitions and finite-to-one codings is to understand dynamical and statistical properties of smooth dynamical systems, we also provide applications in this context.

Two words of caution. Firstly, we do not provide a historical account of the development of Markov partitions. Secondly, we do not discuss symbolic dynamics to any great extent, but only finite-to-one codings for systems with uniform and non-uniform hyperbolicity. Away from these contexts, there are various tools in symbolic dynamics that are important on their own and provide far-reaching conclusions, such as Milnor and Thurston's theory of kneading sequences [**MT88**], Hofbauer towers [**Hof78, Hof79, Hof81, Tak73**], symbolic extensions [**BD04, Bur11, Dow11**], Yoccoz puzzles [**Yoc15**], and Young towers [**You98**].

We divide this survey into three parts. In Part one, §2, we discuss the theory of invariant manifolds for uniformly and non-uniformly hyperbolic systems, including the construction of local charts and graph transforms. For simplicity of exposition, most of the arguments will be discussed in dimension two, both for diffeomorphisms and maps with discontinuities, but we also sketch how to make the constructions in higher dimension. In Part two, §3, we extend this theory to pseudo-orbits, and explain how to use them to construct Markov partitions and finite-to-one coding maps. In Part three, §4, we provide applications.

2. *Part one: Charts, graph transforms, and invariant manifolds*

We introduce tools that allow us to pass from the infinitesimal information given by the assumption on the derivative of the system to a representation of its local dynamics. The main goal is to introduce three tools:

- local charts, which locally represent the dynamics as a small perturbation of a hyperbolic matrix;
- graph transforms, which explore the hyperbolic feature of the local representation to identify points that remain close to trajectories;
- invariant manifolds, which provide dynamical coordinates and allow the future and past behavior of the system to be separated.

For methodological reasons, the discussion of this part of the paper is divided into sections, each of them treating a different class of systems. In §2.1 we deal with uniformly hyperbolic diffeomorphisms. In §2.2 we consider non-uniformly hyperbolic diffeomorphisms. In the last two sections, we discuss non-uniformly hyperbolic surface maps with discontinuities: in §2.3 we assume bounded derivative (for example, Poincaré return maps of flows without fixed points), and in §2.4 we allow the derivative to grow polynomially fast to infinity (for example, billiard maps). The discussion in each new section emphasizes the new input that is necessary to make the construction work, so we recommend the reader to follow the text as presented here.

2.1. *Uniformly hyperbolic systems.*

Uniformly hyperbolic systems are at the heart of the great developments that tailored the beginning of the modern theories of dynamical systems and ergodic theory, and constitute one of the nicest situations in which a system shows chaos in almost any sense of the word: exponential divergence of the trajectories, denseness of periodic orbits, among others. The study of uniformly hyperbolic systems has a long history, stretching back to the 19th century with the study of geodesic flows on surfaces of constant negative curvature by Hadamard [**Had98**]. This topic was extensively developed between 1920 and 1940; here we mention the work of Morse [**Mor34**], Hedlund [**Hed39**], and Hopf [**Hop39, Hop40**]. About 1940, it became clear that geodesic flows were a particular case of the real setup of interest, and Anosov realized that the theory goes through under a more general condition, which he called the (U)-*condition*. In his own words, a system satisfies the (U)-condition if it has 'exponential dichotomy of solutions' [**Ano69**, p. 22]. Anosov made fundamental contributions to the study of (U)-systems, including their ergodicity [**Ano69**, Theorem 4]. Nowadays, (U)-systems are called *Anosov systems*, and the assumption on exponential dichotomy of solutions is called *hyperbolicity*.

While the Russian school focused on the probabilistic aspects of dynamical systems, the American school led by Smale focused on the topological aspects. Smale discovered the *horseshoe*, which is the first example of a system shown to have infinitely many periodic points while being structurally stable. The history of the discovery is explained in [**Sma98**], where Smale claims that 'the horseshoe is a natural consequence of a geometrical way of looking at the equations of Cartwright–Littlewood and Levinson'. A horseshoe has similar properties to Anosov systems, because the recurrent (but not all) trajectories are hyperbolic. For the purpose of dynamics, this is satisfactory because a non-recurrent trajectory is uninteresting for dynamical purposes. Having this in mind, Smale introduced the notion of *Axiom A systems*, where hyperbolicity is required to hold only on the non-wandering set. For transitive Anosov systems, the notions of Anosov and Smale coincide, but there are Axiom A systems that are not Anosov. What we call *uniformly hyperbolic* are Anosov and Axiom A systems. Nowadays there are great textbooks describing such systems; see [**BS02, KH95, Shu87**].

The main result of this section is the existence of local invariant manifolds. It holds for C^1 uniformly hyperbolic systems; see [**Shu87**]. However, to maintain an analogy with the non-uniformly hyperbolic context to be discussed in §2.2, we will assume most of the time that the system is $C^{1+\beta}$; see definition in §2.1.2.

2.1.1. *Definitions and examples.* Let M be a closed (compact without boundary) connected smooth Riemannian manifold, and let $f : M \to M$ be a C^1 diffeomorphism.

Anosov diffeomorphism. We call f an *Anosov diffeomorphism* if there exist a continuous splitting $TM = E^s \oplus E^u$ and constants $C > 0$, $\kappa < 1$ such that:

(1) *Invariance.* $df(E^{s/u}_x) = E^{s/u}_{f(x)}$ for all $x \in M$.

(2) *Contraction.*
- Vectors in E^s contract in the future: $\|df^n v\| \leq C\kappa^n \|v\|$ for all $v \in E^s$, $n \geq 0$.
- Vectors in E^u contract in the past: $\|df^{-n} v\| \leq C\kappa^n \|v\|$ for all $v \in E^u$, $n \geq 0$.

A closed f-invariant set Λ satisfying the above properties is called *uniformly hyperbolic* or simply *hyperbolic*, hence a diffeomorphism is Anosov if the whole phase space M is

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

hyperbolic. The continuity condition of the splitting in the definition indeed follows from the other assumptions; see [BS02, Proposition 5.2.1]. As a matter of fact, the splitting is Hölder continuous, as proved by Anosov [Ano67]; see also the appendix of [Bal95] for a simpler proof. Condition (2) is the exponential dichotomy of solutions mentioned by Anosov. Usually, the above assumptions are rather restrictive because they require the properties on all of M, and sometimes parts of M are not dynamically relevant. The set where interesting dynamics can occur is called the non-wandering set.

Non-wandering set $\Omega(f)$. The *non-wandering set* of f, denoted by $\Omega(f)$, is the set of all $x \in M$ such that for every neighborhood $U \ni x$ there exists $n \neq 0$ such that $f^n(U) \cap U \neq \emptyset$.

In other words, a point is non-wandering if, no matter how small we choose a neighborhood, it does self-intersect in the future or in the past. In particular, every periodic point is non-wandering. Let $\mathrm{Per}(f)$ denote the set of periodic points of f.

Axiom A diffeomorphism. We call f an *Axiom A diffeomorphism* if:

(1) *Denseness of periodic orbits.* $\overline{\mathrm{Per}(f)} = \Omega(f)$.

(2) *Hyperbolicity.* $\Omega(f)$ is hyperbolic, that is, there exist a continuous splitting $T_{\Omega(f)}M = E^s \oplus E^u$ and constants $C > 0$, $\kappa < 1$ such that:
   - $df(E^{s/u}_x) = E^{s/u}_{f(x)}$ for all $x \in \Omega(f)$;
   - $\|df^n v\| \le C\kappa^n \|v\|$ for all $v \in E^s$, $n \ge 0$;
   - $\|df^{-n} v\| \le C\kappa^n \|v\|$ for all $v \in E^u$, $n \ge 0$.

Every Anosov diffeomorphism is Axiom A, but the converse is false. Now let $\varphi : M \to M$ be a flow generated by a vector field X of class C^1. The definitions of uniformly hyperbolic flows are similar to the ones above, bearing in mind that in the flow direction there is no contraction or expansion. Below, $\langle X \rangle$ represents the subbundle generated by X, whose vector space at x is the line generated by X_x.

Anosov flow. We call φ an *Anosov flow* if $X \neq 0$ everywhere and if there is a continuous splitting $TM = E^s \oplus \langle X \rangle \oplus E^u$ and constants $C > 0$, $\kappa < 1$ such that:

(1) *Invariance.* $d\varphi^t(E^{s/u}_x) = E^{s/u}_{\varphi^t(x)}$ for all $x \in M$, $t \in \mathbb{R}$.

(2) *Contraction.*
   - Vectors in E^s contract in the future: $\|d\varphi^t v\| \le C\kappa^t \|v\|$ for all $v \in E^s$, $t \ge 0$.
   - Vectors in E^u contract in the past: $\|d\varphi^{-t} v\| \le C\kappa^t \|v\|$ for all $v \in E^u$, $t \ge 0$.

Similarly, a closed f-invariant set Λ satisfying the above properties is called *uniformly hyperbolic* or simply *hyperbolic*.

Non-wandering set $\Omega(\varphi)$. The *non-wandering set* of φ, denoted by $\Omega(\varphi)$, is the set of all $x \in M$ such that for every neighborhood $U \ni x$ and for every $t > 0$ there exists $T \in \mathbb{R}$ with $|T| > t$ such that $\varphi^T(U) \cap U \neq \emptyset$.

The above definition is natural, since $\varphi^t(U) \cap U \neq \emptyset$ for any t sufficiently small. Let $\mathrm{Per}(\varphi)$ denote the set of periodic points of φ.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

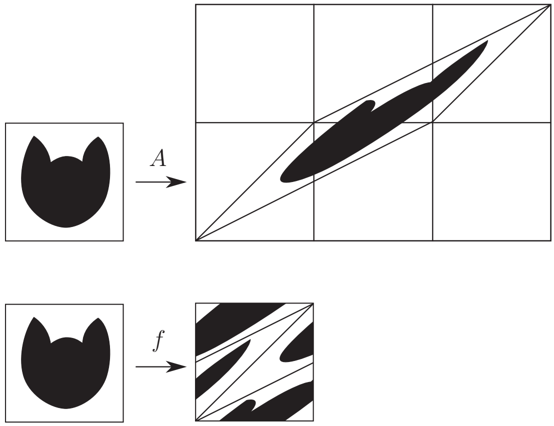

FIGURE 1. Arnold's cat map.

Axiom A flow. We call φ an *Axiom A flow* if $X \neq 0$ on $\Omega(\varphi)$ and:

(1) *Denseness of periodic orbits.* $\overline{\mathrm{Per}(\varphi)} = \Omega(\varphi)$.

(2) *Hyperbolicity.* $\Omega(\varphi)$ is hyperbolic, that is, there exist a continuous splitting $T_{\Omega(f)}M = E^s \oplus \langle X \rangle \oplus E^u$ and constants $C > 0$, $\kappa < 1$ such that:
- $d\varphi^t(E_x^{s/u}) = E_{\varphi^t(x)}^{s/u}$ for all $x \in \Omega(f)$, $t \in \mathbb{R}$;
- $\|d\varphi^t v\| \leq C\kappa^t \|v\|$ for all $v \in E^s$, $t \geq 0$;
- $\|d\varphi^{-t} v\| \leq C\kappa^t \|v\|$ for all $v \in E^u$, $t \geq 0$.

We call a system *uniformly hyperbolic* if it is either Anosov or Axiom A. Here are three classical examples.

(1) Every hyperbolic matrix (a matrix is hyperbolic if none of its eigenvalues lie on the unit circle) $A \in \mathrm{SL}(n, \mathbb{R})$ induces an Anosov diffeomorphism $f = f_A : \mathbb{T}^n \to \mathbb{T}^n$ on the n-dimensional torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. For $A = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right]$, the Anosov diffeomorphism f is known as *Arnold's cat map* or simply the *cat map*. Although the dynamics of A is simple, the dynamics of f as seen on the canonical fundamental domain $[0, 1]^2$ of $\mathbb{T}^2$ is rather complicated; see Figure 1. See also [**BS02**, §1.7].

(2) Smale's horseshoe, generated by the geometrical configuration in Figure 2. See [**BS02, Shu87**].

(3) The geodesic flow on a closed manifold with negative sectional curvature is Anosov; see Figure 3. Its hyperbolicity is more complicated to describe, since it is defined on the (unit) tangent bundle of the manifold and its derivative in the tangent bundle of this (unit) tangent bundle. We refer the reader to [**BP13**, Ch. 1] for a discussion on the two-dimensional case with constant curvature, to [**Ebe01**] for a more general discussion, and to [**Kni02**, §1.3] for a proof of hyperbolicity.

2.1.2. *Preliminaries on the geometry of M.* It is easy to define Hölder continuity for maps on Euclidean spaces. For instance, $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is β-Hölder if there is $\mathfrak{K} > 0$ such that $\|f(x) - f(x)\| \leq \mathfrak{K}\|x - y\|^\beta$ for all $x, y \in U$. Similarly, f is $C^{1+\beta}$ if it is C^1

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

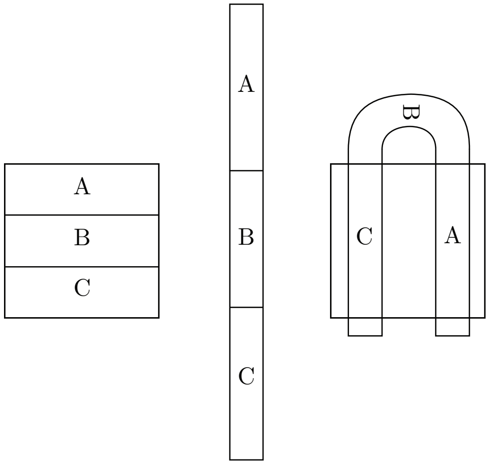

FIGURE 2. The geometrical mechanism in the creation of a horseshoe.

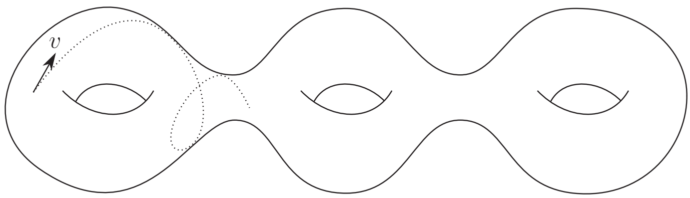

FIGURE 3. A surface with negative curvature, whose geodesic flow is Anosov. Note that the figure is somewhat misleading because it induces one to think that the curvature on the extreme left- and right-hand sides is positive.

and there is $\mathfrak{K} > 0$ such that $\|df_x^{\pm 1} - df_y^{\pm 1}\| \leq \mathfrak{K}\|x - y\|^{\beta}$ for all $x, y \in U$. For general manifolds, this is slightly more complicated because the derivatives are defined in different tangent spaces and so we need to compare the geometry of nearby tangent spaces. For that we use the local charts provided by exponential maps, which are intrinsic to the geometry of M. For the inexperienced reader, we suggest doing the calculations in the Euclidean situation, where all exponential maps are identity.

We are assuming M is a closed connected smooth Riemannian manifold. We denote open balls in M by $B(x, r)$. Given $r > 0$, let $B_x[r] \subset T_x M$ be the open ball with center 0 and radius r. For each $x \in M$, let $\exp_x : T_x M \to M$ be the *exponential map* at x, that is, $\exp_x(v) = \gamma(1)$ where γ is the unique geodesic such that $\gamma'(0) = v$. Given $x \in M$, let $\mathrm{inj}(x)$ be the *radius of injectivity* at x, that is, $\mathrm{inj}(x)$ is the largest $r > 0$ such that the restriction of $\exp_x$ to $B_x[r]$ is a diffeomorphism onto its image. Choose $\mathfrak{r}_0 > 0$ such that for $D_x := B(x, 2\mathfrak{r}_0)$ the following statements hold.

- $\exp_x : B_x[2\mathfrak{r}_0] \to M$ is a 2-bi-Lipschitz diffeomorphism onto its image.
- If $y \in D_x$ then $\mathrm{inj}(y) \geq 2\mathfrak{r}_0$ and $\exp_y^{-1} : D_x \to T_y M$ is a 2-bi-Lipschitz diffeomorphism onto its image.

Such $\mathfrak{r}_0 > 0$ exists because $d(\exp_x)_0 = \mathrm{Id}$ and M is compact.

For $x, x' \in M$, let $\mathscr{L}_{x,x'} := \{A : T_x M \to T_{x'} M : A \text{ is linear}\}$ and $\mathscr{L}_x := \mathscr{L}_{x,x}$. If $d(x, y) < \mathrm{inj}(x)$, then there is a unique radial geodesic γ joining x to y, and the parallel

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

transport $P_{x,y}$ along this geodesic is in $\mathscr{L}_{x,y}$. Let $x, x' \in M$ and $y, z \in M$ such that $d(x, y) < \text{inj}(x)$ and $d(x', z) < \text{inj}(x')$. Given $A \in \mathscr{L}_{y,z}$, 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 base points define a map that differs from $\widetilde{A}$ by pre- and postcomposition 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'. With this notation, we say that f is a $C^{1+\beta}$ diffeomorphism if $f \in C^1$ and there exists $\mathfrak{K} > 0$ such that the following statements hold:

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

2.1.3. *Lyapunov inner product.* By the definitions in §2.1.1, hyperbolicity implies that the restriction of df to E^s is eventually a contraction, exactly when n is large enough so that $C\kappa^n < 1$, and the same occurs to the restriction of df^{-1} to E^u. It turns out that we can define a new inner product, equivalent to the original, for which $df\restriction_{E^s}$ and $df^{-1}\restriction_{E^u}$ are contractions already since the first iterate. This inner product is known as an *adapted metric* or *Lyapunov inner product*. For consistency with the non-uniformly hyperbolic situation, we will use the later notation. The idea of changing an eventual contraction into a contraction is a popular trick in dynamics, and it appears in various contexts, from Picard's theorem on existence and uniqueness of solutions of ordinary differential equations to the construction of invariant manifolds, as we will see here. There are many different ways of defining such an inner product; see, for example, [**Shu87**, Proposition 4.2]. Here, we follow an approach similar to [**BS02**, Proposition 5.2.2].

We assume that $f : M \to M$ is a uniformly hyperbolic diffeomorphism, and we let $\langle \cdot, \cdot \rangle$ be the Riemannian metric on M. For simplicity of notation, we assume that f is Anosov, with invariant splitting $TM = E^s \oplus E^u$ (for Axiom A, the definitions are made inside $\Omega(f)$). Fix $\kappa < \lambda < 1$.

Lyapunov inner product. We define an inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ on M, called the *Lyapunov inner product*, by the following identities:

- for $v_1^s, v_2^s \in E^s$,
$$\langle\!\langle v_1^s, v_2^s \rangle\!\rangle = 2 \sum_{n \ge 0} \lambda^{-2n} \langle df^n v_1^s, df^n v_2^s \rangle;$$
- for $v_1^u, v_2^u \in E^u$,
$$\langle\!\langle v_1^u, v_2^u \rangle\!\rangle = 2 \sum_{n \ge 0} \lambda^{-2n} \langle df^{-n} v_1^u, df^{-n} v_2^u \rangle;$$
- for $v^s \in E^s$ and $v^u \in E^u$,
$$\langle\!\langle v^s, v^u \rangle\!\rangle = 0.$$

We can show, using the uniform hyperbolicity, that $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ is equivalent to and as smooth as $\langle \cdot, \cdot \rangle$. That is why it is also called an *adapted metric*. Letting $|||\cdot|||$ denote the norm induced

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

by $\langle\!\langle\cdot,\cdot\rangle\!\rangle$, if $v^s \in E^s\backslash\{0\}$ then

$$|||dfv^s|||^2 = 2\sum_{n\geq 0}\lambda^{-2n}\|df^n(dfv^s)\|^2 = \lambda^2[|||v^s|||^2 - 2] < \lambda^2|||v^s|||^2,$$

hence $|||dfv^s||| < \lambda|||v^s|||$. Similarly, if $v^u \in E^u\backslash\{0\}$ then $|||df^{-1}v^u||| < \lambda|||v^u|||$.

When M is a surface, the information of the Lyapunov inner product at each $x \in M$ can be recorded by three parameters $s(x), u(x), \alpha(x)$, which we now introduce. The bundles E^s, E^u are one-dimensional, so there are vectors $e^s_x \in E^s_x$ and $e^u_x \in E^u_x$, unitary in the metric $\langle\cdot,\cdot\rangle$. In the Lyapunov inner product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$, we have that $|||e^s_x|||, |||e^u_x||| \in [\sqrt{2}, C\lambda\sqrt{2/(\lambda^2-\kappa^2)}]$ are uniformly bounded away from zero and infinity. (Indeed, $2 < |||e^s_x|||^2 = 2\sum_{n\geq 0}\lambda^{-2n}\|df^n e^s_x\|^2 \leq 2C^2\sum_{n\geq 0}(\kappa/\lambda)^{2n} = 2C^2\lambda^2/(\lambda^2-\kappa^2)$.)

Parameters $s(x), u(x), \alpha(x)$.

$$s(x) = |||e^s_x||| = \sqrt{2}\bigg(\sum_{n\geq 0}\lambda^{-2n}\|df^n e^s_x\|^2\bigg)^{1/2},$$

$$u(x) = |||e^u_x||| = \sqrt{2}\bigg(\sum_{n\geq 0}\lambda^{-2n}\|df^{-n} e^u_x\|^2\bigg)^{1/2},$$

$$\alpha(x) = \angle(E^s_x, E^u_x).$$

As observed, $s(x), u(x)$ are uniformly bounded away from zero and infinity. Since the splitting $E^s \oplus E^u$ is continuous, $\alpha(x)$ is also uniformly bounded away from zero and π.

2.1.4. *Diagonalization of derivative.* For ease of exposition, we continue to assume that M is a surface. We now use the Lyapunov inner product (or, more specifically, the parameters $s(x), u(x), \alpha(x)$) to represent df_x as a hyperbolic matrix. Let $e_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $e_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ be the canonical basis of $\mathbb{R}^2$.

Linear map $C(x)$. For $x \in M$, let $C(x) : \mathbb{R}^2 \to T_xM$ be the linear map such that

$$C(x) : e_1 \mapsto \frac{e^s_x}{s(x)}, \quad C(x) : e_2 \mapsto \frac{e^u_x}{u(x)}.$$

The linear transformation $C(x)$ sends the canonical inner product on $\mathbb{R}^2$ to the Lyapunov inner product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ on T_xM. For a geometer, this may be a simple description of $\langle\!\langle\cdot,\cdot\rangle\!\rangle$, but for practical reasons we do not explore such description. Instead, we study the relation of $C(x)$ with the parameters $s(x), u(x), \alpha(x)$. Given a linear transformation, let $\|\cdot\|$ denote its sup norm and $\|\cdot\|_{\rm Frob}$ its Frobenius norm (the Frobenius norm of a 2×2 matrix $A = \left[\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right]$ is $\|A\|_{\rm Frob} = \sqrt{a^2+b^2+c^2+d^2}$). These two norms are equivalent, with $\|\cdot\| \leq \|\cdot\|_{\rm Frob} \leq \sqrt{2}\|\cdot\|$. The next lemma proves that C diagonalizes df.

LEMMA 2.1. *There is* $\mathscr{L} > 1$ *such that the following statements hold for all* $x \in M$.

(1) $\|C(x)\|_{\rm Frob} \leq 1$ *and* $\|C(x)^{-1}\|_{\rm Frob} = \sqrt{s(x)^2+u(x)^2}/|\sin\alpha(x)|$, *with* $\|C(x)\|$, $\|C(x)^{-1}\| \leq \mathscr{L}$.

(2) $C(f(x))^{-1}\circ df_x \circ C(x)$ *is a diagonal matrix with diagonal entries* $A, B \in \mathbb{R}$ *such that* $|A|, |B^{-1}| < \lambda$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Proof. (1) In the basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ and the basis $\{e^s_x, (e^s_x)^\perp\}$ of T_xM, $C(x)$ takes the form

$$\begin{bmatrix} \frac{1}{s(x)} & \frac{\cos\alpha(x)}{u(x)} \\ 0 & \frac{\sin\alpha(x)}{u(x)} \end{bmatrix},$$

hence

$$\|C(x)\|^2_{\mathrm{Frob}} = \frac{1}{s(x)^2} + \frac{1}{u(x)^2} \le 1.$$

The inverse of $C(x)$ is

$$\begin{bmatrix} s(x) & -\frac{s(x)\cos\alpha(x)}{\sin\alpha(x)} \\ 0 & \frac{u(x)}{\sin\alpha(x)} \end{bmatrix},$$

therefore

$$\|C(x)^{-1}\|_{\mathrm{Frob}} = \frac{\sqrt{s(x)^2 + u(x)^2}}{|\sin\alpha(x)|}.$$

Finally, since $s(x), u(x), \alpha(x)$ are uniformly bounded away from zero and infinity, there is $\mathscr{L} > 1$ such that $\|C(x)\|, \|C(x)^{-1}\| \le \mathscr{L}$ for all $x \in M$.

(2) It is clear that e_1, e_2 are eigenvectors of $C(f(x))^{-1} \circ df_x \circ C(x)$. We start by calculating the eigenvalue of e_1 . Since $df e^s_x = \pm\|df e^s_x\| e^s_{f(x)}$,

$$[df_x \circ C(x)](e_1) = df_x\left[\frac{e^s_x}{s(x)}\right] = \pm\frac{\|df e^s_x\|}{s(x)} e^s_{f(x)},$$

hence

$$[C(f(x))^{-1} \circ df_x \circ C(x)](e_1) = \pm\|df e^s_x\| \tfrac{s(f(x))}{s(x)} e_1.$$

Thus $A := \pm\|df e^s_x\| s(f(x))/s(x)$ is the eigenvalue of e_1. Since

$$s(f(x))^2 = \frac{2\lambda^2}{\|df e^s_x\|^2} \sum_{n\ge1} \lambda^{-2n} \|df^n e^s_x\|^2 = \frac{\lambda^2}{\|df e^s_x\|^2}(s(x)^2 - 2) < \frac{\lambda^2 s(x)^2}{\|df e^s_x\|^2},$$

we have $|A| < \lambda$. Similarly, $B := \pm\|df e^u_x\| u(f(x))/u(x)$ is the eigenvalue of e_2. Observing that $\|df^{-1} e^u_{f(x)}\| \cdot \|df e^u_x\| = 1$, we have

$$u(f(x))^2 = 2 + \sum_{n\ge1} \lambda^{-2n} \|df^{-n} e^u_{f(x)}\|^2 = 2 + \frac{u(x)^2}{\lambda^2\|df e^u_x\|^2} = 2 + \frac{u(f(x))^2}{B^2\lambda^2} > \frac{u(f(x))^2}{B^2\lambda^2},$$

and so $|B| > \lambda^{-1}$. □

Although s, u, α, C depend on the choice of λ, we will not emphasize this dependence because all calculations will be done for some a priori fixed λ.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

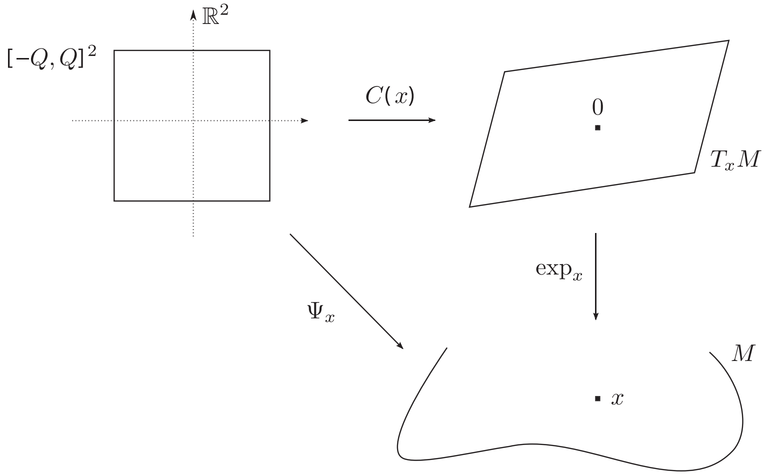

FIGURE 4. The Lyapunov chart Ψ_x at x.

2.1.5. *Lyapunov charts, change of coordinates.* From now on, we assume that f is $C^{1+\beta}$. The next step is to compose the linear transformation $C(x)$ with the exponential map to obtain a local chart of M in which f itself becomes a small perturbation of a hyperbolic matrix. Since this is a natural consequence of the use of the Lyapunov inner product, we will call these charts *Lyapunov charts*, as in [**BP13**, §6.4.2]. Fix a small number $\varepsilon \in (0, \mathfrak{r}_0)$ (how small depends on a finite number of inequalities that ε has to satisfy). Let $Q = \varepsilon^{3/\beta}$.

Lyapunov chart. The *Lyapunov chart* at x is the map $\Psi_x : [-Q, Q]^2 \to M$ defined by $\Psi_x := \exp_x \circ C(x)$; see Figure 4.

Since $Q < \varepsilon < \mathfrak{r}_0$ and $C(x)$ is a contraction, we have $C(x)([-Q, Q]^2) \subset B_x[2\mathfrak{r}_0]$ and so Ψ_x is a diffeomorphism onto its image. By Lemma 2.1(1), Ψ_x is 2-Lipschitz and its inverse is $2\mathscr{L}$-Lipschitz. In Lyapunov charts, f takes the form $f_x := \Psi_{f(x)}^{-1} \circ f \circ \Psi_x$. The next theorem shows that f_x is a small perturbation of a hyperbolic matrix.

THEOREM 2.2. *The following statements hold for all $\varepsilon > 0$ small enough.*

(1) $d(f_x)_0 = C(f(x))^{-1} \circ df_x \circ C(x) = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$ *with* $|A|, |B^{-1}| < \lambda$*; cf. Lemma 2.1.*

(2) $f_x(v_1, v_2) = (Av_1 + h_1(v_1, v_2), Bv_2 + h_2(v_1, v_2))$ *for* $(v_1, v_2) \in [-Q, Q]^2$ *where:*

  (a) $h_1(0, 0) = h_2(0, 0) = 0$ *and* $\nabla h_1(0, 0) = \nabla h_2(0, 0) = 0$*;*

  (b) $\|h_1\|_{1+\beta/2} < \varepsilon$ *and* $\|h_2\|_{1+\beta/2} < \varepsilon$*, where the norms are taken in* $[-Q, Q]^2$.

Similar statements hold for $f_x^{-1} := \Psi_x^{-1} \circ f^{-1} \circ \Psi_{f(x)}$.

Proof. Property (1) is clear since $d(\Psi_x)_0 = C(x)$ and $d(\Psi_{f(x)})_0 = C(f(x))$. By Lemma 2.1, $d(f_x)_0 = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$ with $|A|, |B^{-1}| < \lambda$. Define $h_1, h_2 : [-Q, Q]^2 \to \mathbb{R}$ by $f_x(v_1, v_2) = (Av_1 + h_1(v_1, v_2), Bv_2 + h_2(v_1, v_2))$. Then (a) is automatically satisfied. It remains to prove (b), which will follow from estimating $\|d(f_x)_{w_1} - d(f_x)_{w_2}\|$. For the inexperienced reader, we suggest doing the calculation in the Euclidean situation (hence all exponential maps are identity). Below we work through the general case. For $i = 1, 2$,

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

define

$$A_i = d(\widetilde{\exp_{f(x)}^{-1})_{(f\circ\exp_x)(w_i)}}, \quad B_i = d\widetilde{f_{\exp_x(w_i)}}, \quad C_i = d\widetilde{(\exp_x)_{w_i}}.$$

We first estimate $\|A_1B_1C_1 - A_2B_2C_2\|$. Note that:

- A_1, A_2 are derivatives of the map $\exp_{f(x)}^{-1}$ at nearby points, and so $\|A_1 - A_2\| \leq \mathscr{H}\|w_1 - w_2\|$, where $\mathscr{H} > 0$ is a constant that only depends on the regularity of exponential maps and their inverses;
- B_1, B_2 are derivatives of f at nearby points, so $\|B_1 - B_2\| \leq 2\mathfrak{K}\|w_1 - w_2\|^\beta$;
- C_1, C_2 are derivatives of exponential maps at nearby points, so $\|C_1 - C_2\| \leq \mathscr{H}\|w_1 - w_2\|$.

Applying some triangle inequalities, we obtain that

$$\|A_1B_1C_1 - A_2B_2C_2\| \leq 24\mathfrak{K}\mathscr{H}\|w_1 - w_2\|^\beta.$$

Now we estimate $\|d(f_x)_{w_1} - d(f_x)_{w_2}\|$:

$$\begin{aligned}\|d(f_x)_{w_1} - d(f_x)_{w_2}\| &\leq \|C(f(x))^{-1}\|\|A_1B_1C_1 - A_2B_2C_2\|\|C(x)\| \\ &\leq 24\mathfrak{K}\mathscr{H}\mathscr{L}\|w_1 - w_2\|^\beta.\end{aligned}$$

Since $\|w_1 - w_2\| < 4Q$, if $\varepsilon > 0$ is small enough then $24\mathfrak{K}\mathscr{H}\mathscr{L}\|w_1 - w_2\|^{\beta/2} \leq 96\mathfrak{K}\mathscr{H}\mathscr{L}\varepsilon^{3/2} < \varepsilon$, hence $\|d(f_x)_{w_1} - d(f_x)_{w_2}\| \leq \varepsilon\|w_1 - w_2\|^{\beta/2}$. □

2.1.6. *Graph transforms: construction of invariant manifolds.* A consequence of the hyperbolic behavior of f_x is that it sends curves that are almost parallel to the vertical axis to curves with the same property; similarly, the inverse map f_x^{-1} sends curves that are almost parallel to the horizontal axis to curves with the same property. This geometrical feature allows to construct local stable and unstable manifolds. According to Anosov [**Ano69**, p. 23], this construction was more or less known to Darboux, Poincaré and Lyapunov, but their proofs required additional assumptions on the system. Hadamard and Perron were the ones to observe that hyperbolicity is a sufficient condition. Below, we explain the method of Hadamard. The idea is to find the local invariant manifolds among graphs of functions, which we call *admissible manifolds*. The maps $f_x^{\pm1}$ define operators on the spaces of admissible manifolds, called *graph transforms*, and the local invariant manifolds are limit points of compositions of such operators. As already mentioned, usually f is only assumed to be C^1, but we take $f \in C^{1+\beta}$ to maintain the analogy with the remainder of the text, and make use of the Lyapunov charts constructed in §2.1.5.

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

We call F, G the *representing functions* of V^s, V^u, respectively. We prefer not to specify the quantifier for the condition $\|F'\|_{C^0}, \|G'\|_{C^0} \approx 0$. Instead, think of an s/u-admissible manifold as an almost horizontal/vertical curve that is tangent to the

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

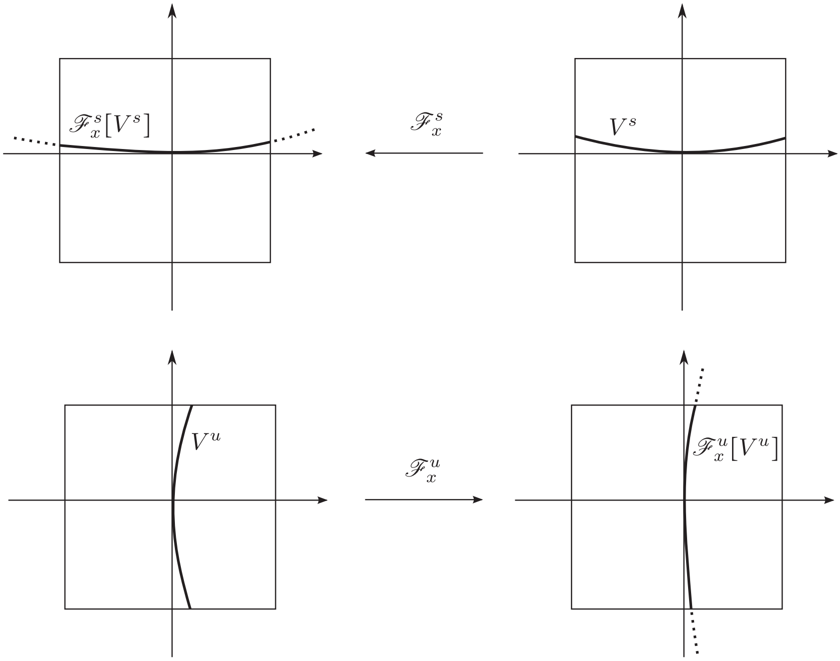

FIGURE 5. Graph transforms: the stable graph transform $\mathscr{F}_x^s$ sends an s-admissible manifold at $\Psi_{f(x)}$ to an s-admissible manifold at Ψ_x, while the unstable graph transform $\mathscr{F}_x^u$ sends a u-admissible manifold at Ψ_x to a u-admissible manifold at $\Psi_{f(x)}$.

horizontal/vertical axis at the origin. Let $\mathscr{M}_x^s, \mathscr{M}_x^u$ be the space of all s, u-admissible manifolds at Ψ_x, respectively. Introduce a metric on $\mathscr{M}_x^s$ as follows: for $V_1, V_2 \in \mathscr{M}_x^s$ 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 $\mathscr{M}_x^u$.

Graph transforms $\mathscr{F}_x^s, \mathscr{F}_x^u$. The *stable graph transform* $\mathscr{F}_x^s : \mathscr{M}_{f(x)}^s \to \mathscr{M}_x^s$ is the map that sends $V^s \in \mathscr{M}_{f(x)}^s$ to the unique $\mathscr{F}_x^s[V^s] \in \mathscr{M}_x^s$ with representing function F such that $\Psi_x\{(t, F(t)) : |t| \leq Q\} \subset f^{-1}(V^s)$. Similarly, the *unstable graph transform* $\mathscr{F}_x^u$: $\mathscr{M}_x^u \to \mathscr{M}_{f(x)}^u$ is the map that sends $V^u \in \mathscr{M}_x^u$ to the unique $\mathscr{F}_x^u[V^u] \in \mathscr{M}_{f(x)}^u$ with representing function G such that $\Psi_{f(x)}\{(G(t), t) : |t| \leq Q\} \subset f(V^u)$.

In other words, $\mathscr{F}_x^s$ sends an s-admissible manifold at $\Psi_{f(x)}$ with representing function F to an s-admissible manifold at Ψ_x whose graph of the representing function is contained in the graph of $f_x^{-1} \circ F$, and $\mathscr{F}_x^u$ sends a u-admissible manifold at Ψ_x with representing function G to a u-admissible manifold at $\Psi_{f(x)}$ whose graph of the representing function is contained in the graph of $f_x \circ G$. See Figure 5.

THEOREM 2.3. *$\mathscr{F}_x^s$ and $\mathscr{F}_x^u$ are well-defined contractions.*

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

The proof of the theorem follows from the hyperbolicity of f_x. Using it, we can construct local stable and unstable manifolds.

Stable/unstable manifolds. The *stable manifold* of $x \in M$ is the s-admissible manifold $V^s[x] \in \mathscr{M}_x^s$ defined by

$$V^s[x] := \lim_{n\to\infty} (\mathscr{F}_x^s \circ \cdots \circ \mathscr{F}_{f^{n-1}(x)}^s)[V_n]$$

for some (any) sequence $\{V_n\}_{n\geq 0}$ with $V_n \in \mathscr{M}_{f^n(x)}^s$. The *unstable manifold* of $x \in M$ is the u-admissible manifold $V^u[x] \in \mathscr{M}_x^u$ defined by

$$V^u[x] := \lim_{n\to -\infty} (\mathscr{F}_{f^{-1}(x)}^u \circ \cdots \circ \mathscr{F}_{f^n(x)}^u)[V_n]$$

for some (any) sequence $\{V_n\}_{n\leq 0}$ with $V_n \in \mathscr{M}_{f^n(x)}^u$.

The sets $V^s[x]$ and $V^u[x]$ are well defined because the graph transforms are contractions (Theorem 2.3 above), and they are indeed admissible curves. Note that $V^s[x]$ only depends on the future $\{f^n(x)\}_{n\geq 0}$, while $V^u[x]$ only depends on the past $\{f^n(x)\}_{n\leq 0}$. Also, since Ψ_x and its inverse have uniformly bounded norms (Lemma 2.1(a)), the stable and unstable manifolds have uniform sizes.

2.1.7. *Higher dimensions.* We sketch how to make the construction in higher dimensions. Our definition of Lyapunov inner product works in any dimension, but not the definition of $C(x)$. Since this matrix is used to send the canonical inner product of $\mathbb{R}^2$ to the Lyapunov inner product on T_xM, in arbitrary dimension we can similarly define $C(x) : \mathbb{R}^n \to T_xM$ to be a linear transformation such that $\langle v, w\rangle_{\mathbb{R}^n} = \langle\!\langle C(x)v, C(x)w\rangle\!\rangle$ for all $v, w \in \mathbb{R}^n$, that is, $C(x)$ is an isometry between $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_{\mathbb{R}^n})$ and $(T_xM, \langle\!\langle\cdot,\cdot\rangle\!\rangle)$. Let $d_s, d_u \in \mathbb{N}$ be the dimensions of E^s, E^u. The map $C(x)$ is not uniquely defined, and we can assume that it sends $\mathbb{R}^{d_s} \times \{0\}$ to E_x^s and $\{0\} \times \mathbb{R}^{d_u}$ to E_x^u. Doing this for all $x \in M$, we obtain a family $\{C(x)\}_{x\in M}$ of linear transformations. Although the splitting $TM = E^s \oplus E^u$ is continuous, we cannot always take $x \in M \mapsto C(x)$ continuously, because E^s and E^u may have non-real exponents, causing rotations inside them. But for our purpose, what matters is the behavior of the sequence $\{C(f^n(x))\}_{n\in\mathbb{Z}}$ for each $x \in M$. For what it is worth, $x \in M \mapsto C(x)$ can be chosen measurably; see, for example, [**BO18** footnote, at pp. 48].

The composition $C(f(x))^{-1} \circ df_x \circ C(x)$ takes the block form

$$C(f(x))^{-1} \circ df_x \circ C(x) = \begin{bmatrix} D_s & 0 \\ 0 & D_u \end{bmatrix},$$

where D_s is a $d_s \times d_s$ matrix such that $\|D_s v\| \leq \lambda\|v\|$ for all $v \in \mathbb{R}^{d_s}$, and D_u is a $d_u \times d_u$ matrix such that $\|D_u^{-1} w\| \leq \lambda\|w\|$ for all $w \in \mathbb{R}^{d_u}$. This is the higher-dimensional counterpart of Lemma 2.1(2). Define the Lyapunov chart Ψ_x as in §2.1.5, which satisfies a higher-dimensional version of Theorem 2.2 with respect to the above block form. Defining an s-admissible manifold at Ψ_x as a set of the form $V^s = \Psi_x\{(t, F(t)) : t \in [-Q, Q]^{d_s}\}$, where $F : [-Q, Q]^{d_s} \to \mathbb{R}^{d_u}$ is a C^1 function such that $F(0) = 0$, $F'(0) = 0$ and $\|F'\|_{C^0} \approx 0$, and similarly u-admissible manifolds at Ψ_x, Theorem 2.3 holds. Hence, every $x \in M$ has local stable and unstable manifolds.

2.2. *Non-uniformly hyperbolic systems.* As discussed in §2.1, uniformly hyperbolic systems had a big impact on the development of dynamical systems and ergodic theory. Unfortunately, uniform hyperbolicity is a condition that is not usually satisfied. For instance, if a three-dimensional manifold admits an Anosov flow then its fundamental group has exponential growth [**PT72**]. During the 1970s, new notions of hyperbolicity were proposed. These notions substitute the uniform assumption with weaker ones. One of them, called *non-uniform hyperbolicity*, was introduced by Pesin [**Pes76, Pes77a, Pes77b**]. In contrast to uniform hyperbolicity, which requires hyperbolicity to hold every time (for all $n \in \mathbb{Z}$ or $t \in \mathbb{R}$) and everywhere (for all $x \in M$), the notion of non-uniform hyperbolicity assumes an asymptotic hyperbolicity (on average) not necessarily in the whole phase space (almost everywhere), that is, hyperbolicity occurs but in a non-uniform way.

Here is a simplistic way of comparing uniform and non-uniform hyperbolicity. To prepare the dough for bread, a baker needs repeatedly to contract and stretch the dough. An ideal baker would perform such an operation with every movement, all over the dough. This is uniform hyperbolicity. On the other hand, a real-life baker practices non-uniform hyperbolicity: he does not perform the operation with every movement (he might get tired from time to time) and he can forget some tiny parts of the dough. As it turns out, it is the notion of non-uniform hyperbolicity that allows for applications outside of mathematics.

Non-uniform hyperbolicity is simultaneously weak enough to include many new examples and applications, and strong enough to recover many of the properties of uniformly hyperbolic systems, such as stable manifolds and graph transforms. This is one of the reasons for the success of the theory of non-uniformly hyperbolic systems, known as *Pesin theory*. Since its beginnings, it has been an important tool for the understanding of ergodic and statistical properties of smooth dynamical systems. Nowadays, Pesin theory is classical and there are great textbooks on the topic; see [**BP07, FHY83, KM95**]. For the applications to symbolic dynamics in Part two, §3, we follow the modern approach recently developed by Sarig [**Sar13**], which has been slightly improved in the past two years or so [**BO18, LM18, LS19**].

We now make *essential* use of the $C^{1+\beta}$ regularity. Indeed, the theory is just not true under C^1 regularity; see, for example, Pugh's example in [**BP13**, Ch. 15].

2.2.1. *Definitions and examples.* Let M be a closed (compact without boundary) connected smooth Riemannian manifold, and let $f : M \to M$ be a C^1 diffeomorphism. The objects that identify the asymptotic hyperbolicity are the *Lyapunov exponents*.

Lyapunov exponent. For a non-zero vector $v \in TM$, the *Lyapunov exponent* of f at v is defined by

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

when the limit exists.

The mere existence of the limit should not be taken for granted, even for uniformly hyperbolic systems. It comes from the *Oseledets theorem*, which is a measure-theoretic

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

statement that we now explain. Assume that μ is a probability measure on M, invariant under f. For simplicity, assume that μ is ergodic. The Oseledets theorem proves that, for μ-almost every (a.e.) $x \in M$, the Lyapunov exponents of every non-zero $v \in T_x M$ exist [**Ose68**]. Furthermore, they exist and are equal for future or past iterations. In our context, we state this as follows.

THEOREM 2.4. (Oseledets) *Let (f, μ) be as above. Then there exist an f-invariant subset $\widetilde{M} \subset M$ with $\mu[\widetilde{M}] = 1$, real numbers $\chi_1 < \chi_2 < \cdots < \chi_k$, and a splitting $T\widetilde{M} = E^1 \oplus E^2 \oplus \cdots \oplus E^k$ satisfying the following properties.*

(1) *Invariance. $df(E^i_x) = E^i_{f(x)}$ for all $x \in \widetilde{M}$ and $i = 1, \ldots, k$.*

(2) *Lyapunov exponents. For all $x \in \widetilde{M}$ and all non-zero $v \in E^i_x$,*

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

When μ is not ergodic, we can apply a standard argument of ergodic decomposition to conclude that $\chi(v)$ exists μ-a.e., but now its value depends on x. Theorem 2.4 above follows from the general version of the Oseledets theorem on cocycles satisfying an *integrability condition*; see, for example, the recent survey of Filip [**Fil19**]. In our setting, the integrability condition is that $\log \|df^{\pm 1}\| \in L^1(\mu)$. Since f is a diffeomorphism on a closed manifold, this condition is automatically satisfied.

The notion of *non-uniform hyperbolicity* also depends on a measure. Let (f, μ) be as above, where μ is not necessarily ergodic.

Non-uniformly hyperbolic diffeomorphism. The pair (f, μ) is called *non-uniformly hyperbolic* if, for μ-a.e. $x \in M$, we have $\chi(v) \neq 0$ for all non-zero $v \in T_x M$. In this case, μ is called a *hyperbolic* measure.

If f is uniformly hyperbolic (see the notation of §2.1.1), then $\chi(v) \leq \log \kappa < 0$ for non-zero $v \in E^s$, and $\chi(v) \geq -\log \kappa > 0$ for non-zero $v \in E^u$, wherever the Lyapunov exponents exist. Thus a uniformly hyperbolic diffeomorphism is non-uniformly hyperbolic, for any invariant probability measure. Here is another example: if f is a diffeomorphism and $p \in M$ is a hyperbolic periodic point with period n, then $\mu = 1/n \sum_{k=0}^{n-1} \delta_{f^k(p)}$ is a hyperbolic measure. We will usually assume that the Lyapunov exponents are bounded away from zero.

χ-hyperbolic measure. Given (f, μ) non-uniformly hyperbolic and $\chi > 0$, μ is called *χ-hyperbolic* if, for μ-a.e. $x \in M$, we have $|\chi(v)| > \chi$ for all non-zero $v \in T_x M$.

In the latter example, μ is χ-hyperbolic for all χ smaller than the multiplier of p. Now let $\varphi : M \to M$ be a flow generated by a vector field X of class C^1, and let μ be a φ-invariant probability measure on M. Since $d\varphi^t \circ X = X \circ \varphi^t$, we have $\chi(X_x) = 0$ for all $x \in M$, hence the assumption of non-zero exponents is required in the remaining directions.

Non-uniformly hyperbolic flow. The pair (φ, μ) is called non-uniformly hyperbolic if, for μ-a.e. $x \in M$, we have $\chi(v) \neq 0$ for all $v \in T_x M$ not proportional to X_x. When this happens, the measure μ is called hyperbolic.

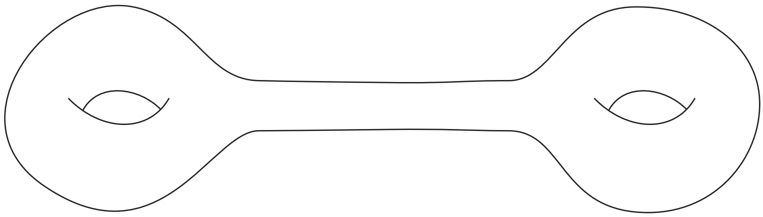

FIGURE 6. Example of a surface with non-positive curvature.

Again, it is easy to see that a uniformly hyperbolic flow is non-uniformly hyperbolic for any invariant probability measure, and that the Dirac measure of a hyperbolic closed orbit is hyperbolic. The notion of χ-hyperbolic measure is defined accordingly.

χ-hyperbolic measure. Given (φ, μ) non-uniformly hyperbolic and $\chi > 0$, μ is called *χ-hyperbolic* if, for μ-a.e. $x \in M$, we have $|\chi(v)| > \chi$ for all non-zero $v \in T_x M$ transverse to X_x.

Let us mention some classical examples.

(1) The slowdown of $f_A : \mathbb{T}^2 \to \mathbb{T}^2$; see [**Kat79**] and [**BP13**, §1.3].

(2) Let f be a C^1 surface diffeomorphism, and let μ be an ergodic f-invariant probability measure. Let $h = h_\mu(f)$ be the Kolmogorov–Sinaĭ entropy, and assume that $h > 0$. Then (f, μ) is non-uniformly hyperbolic, as consequence of the Ruelle inequality applied to f and to f^{-1}:

- f has a positive Lyapunov exponent $\chi^+ \geq h > 0$;
- we have $h_\mu(f^{-1}) = h$, and the Lyapunov spectrum of (f^{-1}, μ) is minus the Lyapunov spectrum of (f, μ), hence f has a Lyapunov exponent χ^- such that $-\chi^- \geq h$, that is, $\chi^- \leq -h < 0$. If in addition $h > \chi$, then μ is χ-hyperbolic.

(3) Let N be a closed manifold with non-positive sectional curvature, for example the surface in Figure 6 containing a flat cylinder between two regions of negative curvature. The geodesic flow on N, which is defined on $M = T_1 N$, has a natural invariant volume measure μ. Pesin showed that if the trajectory of a vector $x \in M$ spends a positive fraction of time in regions of negative sectional curvature, then $\chi(v) \neq 0$ for all $v \in T_x M$ transverse to X_x; see [**Pes77a**, Theorem 10.5]. The underlying philosophy (although not entirely correct), is that in regions of negative sectional curvature the derivative behaves as in a uniformly hyperbolic flow, and in regions of zero sectional curvature it only varies linearly, so the overall exponential behavior beats the linear. Therefore, if μ is ergodic then it is hyperbolic. Unfortunately, the ergodicity of μ is still an open problem (even when N is a surface).

2.2.2. *The non-uniformly hyperbolic locus* NUH_χ. As we have cast it above, the notion of non-uniform hyperbolicity is an almost-everywhere statement that depends on a measure. Due to the Oseledets theorem, we can still get almost-everywhere statements if we only consider *Lyapunov regular* points, which are points that satisfy Theorem 2.4 and a non-degeneracy assumption on the angles $\angle(E^i, E^j)$ between the invariant subbundles. For some applications, this restriction is cumbersome. For example, if x, y are Lyapunov

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

regular, then most likely points in $W^s(x) \cap W^u(y)$ are *not* Lyapunov regular (this happens, for example, when x, y have different Lyapunov exponents).

In what follows, we employ a different approach. We fix some $\chi > 0$ and consider the set of points satisfying a weaker notion of non-uniform hyperbolicity that still allows us to construct local invariant manifolds. This perspective appeared in an essential way in the work carried out by the author with Buzzi and Crovisier [**BCL**]. Independently and simultaneously, Ben Ovadia recently obtained a similar characterization in higher dimensions [**BO19**].

Let $f : M \to M$ be a C^1 diffeomorphism. As in §3, we start by assuming that M is a closed surface. Let $\chi > 0$.

The non-uniformly hyperbolic locus NUH_χ. This is the set of points $x \in M$ for which there are transverse unitary vectors $e^s_x, e^u_x \in T_xM$ such that the following conditions hold.

(NUH1) e^s_x contracts in the future at least $-\chi$ and expands in the past:

$$\limsup_{n\to+\infty} \frac{1}{n} \log \|df^n e^s_x\| \le -\chi \quad \text{and} \quad \liminf_{n\to+\infty} \frac{1}{n} \log \|df^{-n} e^s_x\| > 0.$$

(NUH2) e^u_x contracts in the past at least $-\chi$ and expands in the future:

$$\limsup_{n\to+\infty} \frac{1}{n} \log \|df^{-n} e^u_x\| \le -\chi \quad \text{and} \quad \liminf_{n\to+\infty} \frac{1}{n} \log \|df^{n} e^u_x\| > 0.$$

(NUH3) The parameters $s(x), u(x)$ below are finite:

$$s(x) = \sqrt{2}\bigg(\sum_{n\ge 0} e^{2n\chi} \|df^n e^s_x\|^2\bigg)^{1/2} \in [\sqrt{2}, \infty),$$

$$u(x) = \sqrt{2}\bigg(\sum_{n\ge 0} e^{2n\chi} \|df^{-n} e^u_x\|^2\bigg)^{1/2} \in [\sqrt{2}, \infty).$$

Clearly, NUH_χ is invariant by f. Observe that the definitions of $s(x), u(x)$ are the same as those given in §2.1.3, where we change λ to $e^{-\chi}$. Conditions (NUH1) and (NUH2) guarantee that e^s_x, e^u_x are defined up to a sign, and condition (NUH3) guarantees asymptotic contractions of rates at least $-\chi$. These conditions are weaker than Lyapunov regularity, hence NUH_χ contains all Lyapunov regular points with exponents greater than χ in absolute value. In particular, NUH_χ carries all χ-hyperbolic measures. But NUH_χ might contain points with some Lyapunov exponents equal to $\pm\chi$, and even non-regular points, where the contraction rates oscillate infinitely often. Usually, NUH_χ is a non-compact subset of M. Observe that if (NUH3) holds, then the first conditions of (NUH1)–(NUH2) hold as well. In practice, this is how we will show that $x \in \mathrm{NUH}_\chi$.

The quality of hyperbolicity can be measured from the parameters $s(x), u(x)$ and from the angle $\alpha(x) = \angle(e^s_x, e^u_x)$. More specifically, $x \in \mathrm{NUH}_\chi$ has bad hyperbolicity when at least one of the following situations occurs.

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

None of these situations happen for uniformly hyperbolic systems: as we have seen in §2.1.3, for uniformly hyperbolic systems the parameters s, u, α are uniformly bounded away from zero and infinity. For non-uniformly hyperbolic systems, the behavior is more complicated. Another reason for complication is that, contrary to uniformly hyperbolic systems, the maps $x \in \mathrm{NUH}_\chi \mapsto e^s_x, e^u_x$ are usually no more than just measurable.

2.2.3. *Diagonalization of derivative.* As in §2.1.4, we define linear maps $C(x)$ that diagonalize df, the difference being that we only take $x \in \mathrm{NUH}_\chi$.

Linear map $C(x)$. For $x \in \mathrm{NUH}_\chi$, let $C(x) : \mathbb{R}^2 \to T_xM$ be the linear map such that

$$C(x) : e_1 \mapsto \frac{e^s_x}{s(x)}, \quad C(x) : e_2 \mapsto \frac{e^u_x}{u(x)}.$$

Above, $\{e_1, e_2\}$ is the canonical basis for $\mathbb{R}^2$. If, for each $x \in \mathrm{NUH}_\chi$, we define a Lyapunov inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ on T_xM, then $C(x)$ sends the canonical metric on $\mathbb{R}^2$ to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Lemma 2.1 remains valid, except for the uniform bound on $\|C(x)^{-1}\|$. This result is known as *Oseledets–Pesin reduction*; see, for example, [**BP13**, Theorem 6.10].

LEMMA 2.5. (Oseledets–Pesin reduction) *The following statements hold for all $x \in \mathrm{NUH}_\chi$.*

(1) $\|C(x)\|_{\mathrm{Frob}} \le 1$ *and* $\|C(x)^{-1}\|_{\mathrm{Frob}} = \sqrt{s(x)^2 + u(x)^2}/|\sin \alpha(x)|$.

(2) $C(f(x))^{-1} \circ df_x \circ C(x)$ *is a diagonal matrix with diagonal entries $A, B \in \mathbb{R}$ such that $|A|, |B^{-1}| < e^{-\chi}$.*

The proof is the same as for Lemma 2.1.

2.2.4. *Pesin charts, the parameter $Q(x)$ and change of coordinates.* From now on, we assume that f is $C^{1+\beta}$. Remember we are also assuming that M is a closed surface. Fix $\varepsilon \in (0, \mathfrak{r}_0)$ small.

Pesin chart. The *Pesin chart* at $x \in \mathrm{NUH}_\chi$ is the map $\Psi_x : [-\varepsilon^{3/\beta}, \varepsilon^{3/\beta}]^2 \to M$ defined by $\Psi_x := \exp_x \circ C(x)$.

This is exactly the same as the definition of the Lyapunov chart given in §2.1.5, but we call it the Pesin chart for historical reasons. The map Ψ_x is well defined for each $x \in \mathrm{NUH}_\chi$. In Pesin charts, f takes the form $f_x := \Psi^{-1}_{f(x)} \circ f \circ \Psi_x$. Unfortunately, we might not be able to see hyperbolicity for f_x, but only for a restriction: while in the uniformly hyperbolic situation $C(x), C(x)^{-1}$ are uniformly bounded, now the parameters s, u, α can degenerate and so $\|C(f(x))^{-1}\|$ can be arbitrarily large, causing a big distortion. To decrease the domain of definition of f_x, we multiply its current size by a large negative power of $\|C(f(x))^{-1}\|$.

Parameter $Q(x)$. For $x \in \mathrm{NUH}_\chi$, define $Q(x) = \varepsilon^{3/\beta}\|C(f(x))^{-1}\|^{-12/\beta}_{\mathrm{Frob}}$.

The choice of the powers $3/\beta$ and $12/\beta$ is not canonical but just an artifact of the proof, and any choice of powers bigger than these would also make the proof work. This more complicated definition of $Q(x)$ is the price we pay for detecting hyperbolicity among non-uniformly hyperbolic systems, as stated in the following theorem.

THEOREM 2.6. (Pesin) *The following statements hold for all $\varepsilon > 0$ small. If $x \in \mathrm{NUH}_\chi$ then:*

(1) $d(f_x)_0 = C(f(x))^{-1} \circ df_x \circ C(x) = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right]$ *with* $|A|, |B^{-1}| < e^{-\chi}$*; cf. Lemma 2.5.*

(2) $f_x(v_1, v_2) = (Av_1 + h_1(v_1, v_2), Bv_2 + h_2(v_1, v_2))$*,* $(v_1, v_2) \in [-10Q(x), 10Q(x)]^2$*, where:*

(a) $h_1(0, 0) = h_2(0, 0) = 0$ *and* $\nabla h_1(0, 0) = \nabla h_2(0, 0) = 0$*;*

(b) $\|h_1\|_{1+\beta/2} < \varepsilon$ *and* $\|h_2\|_{1+\beta/2} < \varepsilon$*, with norms taken in* $[-10Q(x), 10Q(x)]^2$.

Similar statements hold for $f_x^{-1} = \Psi_x^{-1} \circ f^{-1} \circ \Psi_{f(x)}$.

The above theorem and its proof below are similar to [**Sar13**, Theorem 2.7]; see also [**BP07**, Theorem 5.6.1].

Proof. We proceed as in the proof of Theorem 2.2. The main difficulty resides in part (2)(b). We still have the estimate

$$\|A_1B_1C_1 - A_2B_2C_2\| \le 24\mathfrak{K}\mathscr{H}\|w_1 - w_2\|^\beta,$$

but now

$$\begin{aligned}\|d(f_x)_{w_1} - d(f_x)_{w_2}\| &\le \|C(f(x))^{-1}\|\|A_1B_1C_1 - A_2B_2C_2\|\|C(x)\| \\ &\le 24\mathfrak{K}\mathscr{H}\|C(f(x))^{-1}\|\|w_1 - w_2\|^\beta.\end{aligned}$$

If $w_1, w_2 \in [-10Q(x), 10Q(x)]^2$ then $\|w_1 - w_2\| < 40Q(x)$, hence for $\varepsilon > 0$ small,

$$24\mathfrak{K}\mathscr{H}\|C(f(x))^{-1}\|\|w_1 - w_2\|^{\beta/2} \le 200\mathfrak{K}\mathscr{H}\varepsilon^{3/2}\|C(f(x))^{-1}\|^{-5} \le 200\mathfrak{K}\mathscr{H}\varepsilon^{3/2} < \varepsilon.$$

This completes the proof. □

Therefore, at a smaller scale that depends on the quality of hyperbolicity at x, the map f_x is again the perturbation of a hyperbolic matrix.

2.2.5. *Temperedness and the parameter $q(x)$.* After successfully detecting hyperbolicity for f_x, the next step is to define graph transforms. As seen in §2.1, for uniformly hyperbolic systems the domains of all Lyapunov charts have the same size. Since forward images of u-admissible manifolds and backward images of s-admissible manifolds grow essentially as λ^{-1}, their images do cross the successive domains from one side to the other; see Figure 5. By Theorem 2.6, for non-uniformly hyperbolic systems the forward images of u-admissible manifolds and backward images of s-admissible manifolds grow essentially as e^χ. Therefore we face a problem when the ratio $Q(f(x))/Q(x)$ is far from 1.

- If $Q(f(x)) \gg Q(x)$, then the image of a u-admissible manifold at x does not cross the domain of $\Psi_{f(x)}$ from top to bottom.
- If $Q(f(x)) \ll Q(x)$, then the image of an s-admissible manifold at $f(x)$ does not cross the domain of Ψ_x from left to right.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

The parameters $s(x), u(x), \alpha(x)$ and $s(f(x)), u(f(x)), \alpha(f(x))$ differ roughly by the action of df_x, so there is a constant $\mathscr{C} = \mathscr{C}(f) > 1$ such that $\mathscr{C}^{-1} \leq Q(f(x))/Q(x) \leq \mathscr{C}$ for all $x \in \mathrm{NUH}_\chi$. Nevertheless, this control is yet not enough to rule out the above problems, since we can still have $\mathscr{C} \gg e^\chi$. To solve this issue, we need to further reduce the domains of Pesin charts, introducing a parameter that varies regularly.

Parameter $q(x)$. For $x \in \mathrm{NUH}_\chi$, define $q(x) = \inf\{e^{\varepsilon|n|}Q(f^n(x)) : n \in \mathbb{Z}\}$.

While $Q(x)$ (essentially) does not depend on ε, the parameter $q(x)$ does and, if positive, it does behave nicely along orbits:

$$e^{-\varepsilon} \leq \frac{q(f(x))}{q(x)} \leq e^{\varepsilon}.$$

The above definition is motivated by the proof of [**Pes76**, Lemma 1.1.1], and provides the optimal value for $q(x) \leq Q(x)$ satisfying the above inequalities. This is known as the *tempering kernel lemma*; see, for example, [**BP13**, Lemma 6.11]. We remark that there are other proofs of the tempering kernel lemma, but they do not provide optimal $q(x)$; see, for example, [**BP07**, Lemma 3.5.7].

Since $q(x) \leq Q(x)$, the restriction of f_x to the smaller domain $[-q(x), q(x)]^2$ is a small perturbation of a hyperbolic matrix. Now we are safe: restricting Ψ_x to $[-q(x), q(x)]^2$, if $\varepsilon > 0$ is small enough then the growth of u/s-admissible manifolds beats the possible increase/decrease of domains. Motivated by this, we consider the subset of NUH_χ where q is positive.

The non-uniformly hyperbolic locus NUH_χ^*.

$$\mathrm{NUH}_\chi^* = \{x \in \mathrm{NUH}_\chi : q(x) > 0\}.$$

By the next lemma, NUH_χ^* carries the same finite invariant measures as NUH_χ.

LEMMA 2.7. *If μ is an f-invariant probability measure supported on* NUH_χ*, then μ is supported on* NUH_χ^*.

Proof. By assumption, $\mu[\mathrm{NUH}_\chi] = 1$. Clearly, if $\lim_{n\to\pm\infty}(1/n)\log Q(f^n(x)) = 0$ then $q(x) > 0$. We will prove that $\lim_{n\to\pm\infty}(1/n)\log Q(f^n(x)) = 0$ for μ-a.e. $x \in \mathrm{NUH}_\chi$. Define the function $\varphi : \mathrm{NUH}_\chi \to \mathbb{R}$ by

$$\varphi(x) := \log\left[\frac{Q(f(x))}{Q(x)}\right] = \log Q(f(x)) - \log Q(x).$$

Since $\mathscr{C}^{-1} \leq Q(f(x))/Q(x) \leq \mathscr{C}$ for $x \in \mathrm{NUH}_\chi$, we have $\varphi \in L^1(\mu)$. Let $\varphi_n = \log(Q \circ f^n) - \log Q$ be the nth Birkhoff sum of φ. By the Birkhoff ergodic theorem, $\lim_{n\to+\infty}(\varphi_n(x)/n)$ exists μ-a.e. Since by the Poincaré recurrence theorem we have $\liminf_{n\to+\infty}|\varphi_n(x)| = \liminf_{n\to+\infty}|\log Q(f^n(x)) - \log Q(x)| < \infty$ μ-a.e., it follows that $\lim_{n\to+\infty}(\varphi_n(x)/n) = 0$ for μ-a.e. $x \in \mathrm{NUH}_\chi$. Proceeding in the same way for $n \to -\infty$, we conclude that $\lim_{n\to\pm\infty}(1/n)\log Q(f^n(x)) = 0$ for μ-a.e. $x \in \mathrm{NUH}_\chi$. $\square$

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

2.2.6. *Sizes of invariant manifolds: the parameters $q^s(x), q^u(x)$.* Using what we have done so far, we can proceed as in §2.1 to construct invariant manifolds: define s/u-admissible manifolds as graphs of functions $F : [-q(x), q(x)] \to \mathbb{R}$ satisfying some regularity assumptions (which we will explain later), and define graph transforms $\mathscr{F}_x^s, \mathscr{F}_x^u$. Hence Theorem 2.3 holds, so we can construct (local) stable and unstable manifolds for every $x \in \mathrm{NUH}_\chi^*$. This is essentially what is done in Pesin theory; see, for example, [**BP07**, Ch. 7].

In general, $q(x)$ is not the optimal size for the local invariant manifolds, and in some applications we need bigger sizes for them. This is the case for the construction of countable Markov partitions that we will discuss in Part two, §3. Observe that $q(x)$ might be small for two different reasons:

- there is $n > 0$ for which $e^{\varepsilon n} Q(f^n(x))$ is small;
- there is $n > 0$ for which $e^{\varepsilon n} Q(f^{-n}(x))$ is small.

In the first case, the forward behavior of $Q(f^n(x))$ is bad, so we expect to construct a small stable manifold; but we are also constructing a small unstable manifold, that is, the bad forward behavior is influencing the size of the unstable manifold! Since the unstable manifold only depends on the past, its size should not be affected by the future. To deal with this, we introduce two new parameters $q^s(x)$ and $q^u(x)$, the first controlling the future behavior and the second controlling the past behavior. Then we use them to construct invariant manifolds with larger sizes.

Parameters $q^s(x)$ and $q^u(x)$. For $x \in \mathrm{NUH}_\chi^*$, define

$$q^s(x) = \inf\{e^{\varepsilon n} Q(f^n(x)) : n \geq 0\},$$
$$q^u(x) = \inf\{e^{\varepsilon n} Q(f^{-n}(x)) : n \geq 0\}.$$

In other words, $q^s(x), q^u(x)$ are the one-sided versions of $q(x)$. Just like q, the parameters q^s, q^u depend on ε. We will use $q^s(x)$ as the scale for considering the stable graph transform and $q^u(x)$ as the scale for considering the unstable graph transform.

LEMMA 2.8. *For all $x \in \mathrm{NUH}_\chi^*$, the following statements hold.*

(1) *Good definition.* $q^s(x), q^u(x) > 0$ *and* $q(x) = \min\{q^s(x), q^u(x)\}$.

(2) *Greedy algorithm.*

$$q^s(x) = \min\{e^{\varepsilon} q^s(f(x)), Q(x)\},$$
$$q^u(x) = \min\{e^{\varepsilon} q^u(f^{-1}(x)), Q(x)\}.$$

The proofs are direct; see also [**LM18**, Lemma 4.2].

2.2.7. *Graph transforms: construction of invariant manifolds.* There are dynamical explanations for Lemma 2.8(2). Let us discuss the first equality. Assume that s-admissible manifolds at x have representing functions defined in the interval $[-q^s(x), q^s(x)]$. If $\varepsilon > 0$ is small enough, then the stable graph transform $\mathscr{F}_x^s$ takes the graph of a representing function defined in $[-q^s(f(x)), q^s(f(x))]$ and expands it at least by a factor of e^{ε}, so the new representing function is well defined in $[-e^{\varepsilon} q^s(f(x)), e^{\varepsilon} q^s(f(x))]$. Since its domain

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

of definition should not go beyond $[-Q(x), Q(x)]$ (where we have a good control on f_x), the best we can do is to define it in $[-q^s(x), q^s(x)]$. In summary, q^s provides maximal scales for the definition of stable graph transforms. Similarly, q^u provides maximal scales for the definition of unstable graph transforms. With this in mind, we give a new definition of s/u-admissible manifolds.

Admissible manifolds. An *s-admissible manifold* at Ψ_x is a set of the form $V^s = \Psi_x\{(t, F(t)) : |t| \leq q^s(x)\}$, where $F : [-q^s(x), q^s(x)] \to \mathbb{R}$ is a $C^{1+\beta/3}$ function such that $F(0) = F'(0) = 0$ and $\|F'\|_0 + \mathrm{Hol}_{\beta/3}(F') \leq \frac{1}{2}$, where the norms are taken in $[-q^s(x), q^s(x)]$. Similarly, a *u-admissible manifold* at Ψ_x is a set of the form $V^u = \Psi_x\{(G(t), t) : |t| \leq q^u(x)\}$, where $G : [-q^u(x), q^u(x)] \to \mathbb{R}$ is a $C^{1+\beta/3}$ function such that $G(0) = G'(0) = 0$ and $\|G'\|_0 + \mathrm{Hol}_{\beta/3}(G') \leq \frac{1}{2}$, with norms taken in $[-q^u(x), q^u(x)]$.

As before, F, G are called the *representing functions* of V^s, V^u, respectively. Let $\mathscr{M}^s_x, \mathscr{M}^u_x$ be the space of all s, u-admissible manifolds at Ψ_x respectively, which are metric spaces with the C^0 distance. Let $x \in \mathrm{NUH}^*_\chi$.

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

The difference from the previous definition is that the stable and unstable graph transforms are defined at different scales; see Figure 7.

THEOREM 2.9. *$\mathscr{F}^s_x$ and $\mathscr{F}^u_x$ are well-defined contractions.*

For non-uniformly hyperbolic systems, this theorem was first proved by Pesin; see [**Pes76**, Theorem 2.3]. The proof is similar to the proof of Theorem 2.3. In its present form, with scales q^s and q^u, the above result is a special case of [**Sar13**, Proposition 4.12]. For $x \in \mathrm{NUH}^*_\chi$, let $V^s[x]$ and $V^u[x]$ be the stable and unstable manifolds of x, defined as in §2.1.6. Then $V^s[x]$ is the image under Ψ_x of the graph of a function defined in $[-q^s(x), q^s(x)]$, while $V^u[x]$ is the image under Ψ_x of the graph of a function defined in $[-q^u(x), q^u(x)]$.

2.2.8. *Higher dimensions.* Now consider diffeomorphisms in any dimension. The discussion follows [**BO18**] and in some sense [**BO19**]. We can no longer perform the construction using only the parameters $s(x), u(x), \alpha(x)$, because now the spaces E^s, E^u are higher-dimensional, and each vector defines its own parameter. More specifically, consider the following definition, for each fixed $\chi > 0$.

The non-uniformly hyperbolic locus NUH_χ. This is the set of points $x \in M$ for which there is a splitting $T_xM = E^s_x \oplus E^u_x$ such that the following conditions hold.

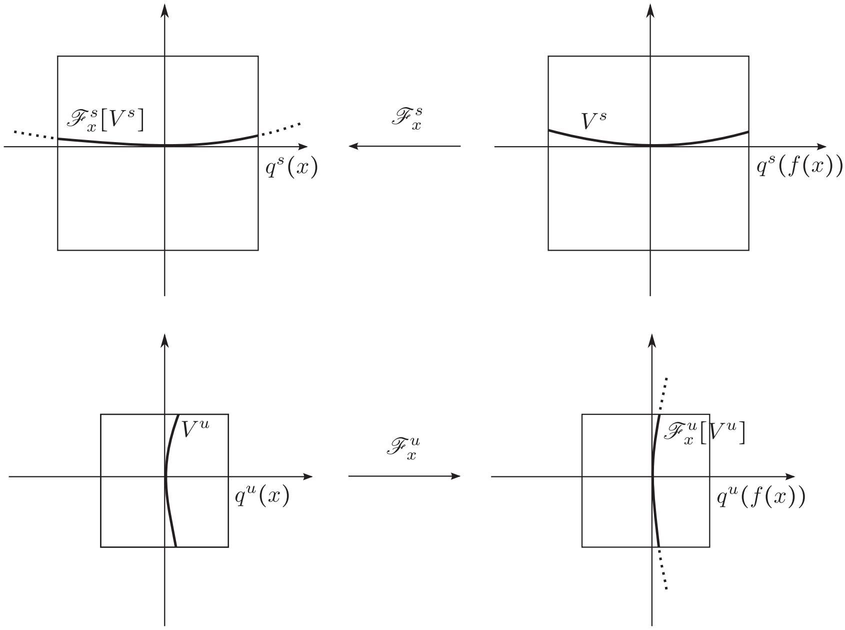

FIGURE 7. The stable graph transforms are defined at scales q^s, and the unstable graph transforms at scales q^u.

(NUH1) Every $v \in E^s_x$ contracts in the future at least $-\chi$ and expands in the past:

$$\limsup_{n\to+\infty} \frac{1}{n} \log \|df^n v\| \leq -\chi \quad \text{and} \quad \liminf_{n\to+\infty} \frac{1}{n} \log \|df^{-n} v\| > 0.$$

(NUH2) Every $v \in E^u_x$ contracts in the past at least $-\chi$ and expands in the future:

$$\limsup_{n\to+\infty} \frac{1}{n} \log \|df^{-n} v\| \leq -\chi \quad \text{and} \quad \liminf_{n\to+\infty} \frac{1}{n} \log \|df^{n} v\| > 0.$$

(NUH3) The parameters $s(x) = \sup_{\substack{v\in E^s_x \\ \|v\|=1}} S(x, v)$ and $u(x) = \sup_{\substack{w\in E^u_x \\ \|w\|=1}} U(x, w)$ are finite, where

$$S(x, v) = \sqrt{2}\bigg(\sum_{n\geq 0} e^{2n\chi} \|df^n v\|^2\bigg)^{1/2},$$

$$U(x, w) = \sqrt{2}\bigg(\sum_{n\geq 0} e^{2n\chi} \|df^{-n} w\|^2\bigg)^{1/2}.$$

In [**BO19**], this definition is similar to the definition of the set χ-summ. Again, NUH_χ is f-invariant, and for each $x \in \text{NUH}_\chi$ we can define a linear transformation $C(x) : \mathbb{R}^n \to T_x M$ that sends the canonical metric on $\mathbb{R}^n$ to the Lyapunov inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ on $T_x M$. We again have the block representation

$$C(f(x))^{-1} \circ df_x \circ C(x) = \begin{bmatrix} D_s & 0 \\ 0 & D_u \end{bmatrix},$$

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

where D_s is a $d_s \times d_s$ matrix such that $\|D_s v\| \le e^{-\chi}\|v\|$ for all $v \in \mathbb{R}^{d_s}$ and D_u is a $d_u \times d_u$ matrix such that $\|D_u^{-1} w\| \le e^{-\chi}\|w\|$ for all $w \in \mathbb{R}^{d_u}$. This is the higher-dimensional Oseledets–Pesin reduction; see Lemma 2.5(2). Define the Pesin chart Ψ_x as in §2.2.4, and the parameter $Q(x) = \mathscr{K}\|C(f(x))^{-1}\|^{-48/\beta}$, where $\mathscr{K} = \mathscr{K}(\beta, \varepsilon)$ is a constant that allows us to keep the estimates of order ε and to absorb multiplicative constants. Then a higher-dimensional version of Theorem 2.6 holds; see [**BO18**, Theorem 1.13]. From now on, we can repeat the two-dimensional construction, defining the parameters $q(x), q^s(x), q^u(x)$, the non-uniformly hyperbolic locus NUH_χ^*, an s-admissible manifold at Ψ_x as a set of the form $V^s = \Psi_x\{(t, F(t)) : t \in [-q^s(x), q^s(x)]^{d_s}\}$, where $F : [-q^s(x), q^s(x)]^{d_s} \to \mathbb{R}^{d_u}$ is a $C^{1+\beta/3}$ function such that $F(0) = F'(0) = 0$ and $\|F'\|_0 + \mathrm{Hol}_{\beta/3}(F') \le \frac{1}{2}$, where the norms are taken in $[-q^s(x), q^s(x)]^{d_s}$, and similarly u-admissible manifolds. Then Theorem 2.3 holds (see [**BO18**, Proposition 2.8]), and so every $x \in \mathrm{NUH}_\chi^*$ has local stable and unstable manifolds.

2.3. *Maps with discontinuities and bounded derivative.* In the previous section we considered diffeomorphisms defined on closed (compact without boundary) surfaces. There are natural examples that do not fit into this context, for example Poincaré return maps of flows and billiard maps. Their common feature is the presence of discontinuities, and the possible explosion of derivatives. In this section and the next we will discuss how to adapt the methods of §2.2 to cover these examples, focusing on the changes that are needed to make the arguments work. We start by dealing with surface maps with discontinuities and bounded derivative. The reference is [**LS19**].

2.3.1. *Definitions and examples.* Let M be a compact surface, possibly with boundary. To avoid multiplicative constants in the calculations, we assume that M has diameter smaller than one. Let $\mathscr{D}^+, \mathscr{D}^-$ be closed subsets of M, and consider a map $f : M\backslash\mathscr{D}^+ \to M$ with inverse $f^{-1} : M\backslash\mathscr{D}^- \to M$. Let $\mathscr{D} := \mathscr{D}^+ \cup \mathscr{D}^-$ be the *set of discontinuities of f*. We require f, f^{-1} to be local $C^{1+\beta}$ diffeomorphisms.

Regularity of f. There is a constant $\mathscr{L} > 0$ with the following property.
- For every $x \in M\backslash\mathscr{D}^+$ there is an open set $U \ni x$ such that $f\restriction_U$ is a diffeomorphism onto its image with $C^{1+\beta}$ norm at most $\mathscr{L}$.
- For every $x \in M\backslash\mathscr{D}^-$ there is an open set $V \ni x$ such that $f^{-1}\restriction_V$ is a diffeomorphism onto its image with $C^{1+\beta}$ norm at most $\mathscr{L}$.

In particular, $\|df^{\pm 1}\|$ is bounded away from zero and infinity, so the integrability condition in the Oseledets theorem holds for any f-invariant probability measure. The main difficulty when dealing with f as above is that, as x approaches $\mathscr{D}$, the open sets U, V become smaller, hence the domains of Pesin charts also need to be smaller. To avoid this issue, we only consider trajectories that do not approach $\mathscr{D}$ exponentially fast.

Here is the example to have in mind. Let N be a three-dimensional closed Riemannian manifold, let X be a $C^{1+\beta}$ vector field on N such that $X(p) \neq 0$ for all $p \in N$, and let

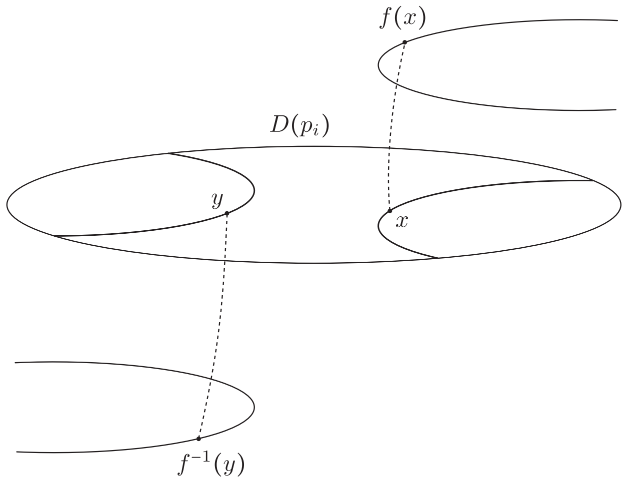

FIGURE 8. Discontinuities for f: in the picture, $x \in \mathscr{D}^+$ and $y \in \mathscr{D}^-$.

$\varphi = \{\varphi^t\}_{t\in\mathbb{R}}$ be the flow generated by X. We can reduce the dynamics of φ to the dynamics of a surface map by constructing a global Poincaré section M for φ as follows.

- Fix $\varepsilon > 0$ small enough.
- For each $p \in N$, consider a closed differentiable disc $D(p)$ centered at p with diameter smaller than ε such that $\angle(T_q D(p), X(q)) > \pi/2 - \varepsilon$ for all $q \in D(p)$.
- Let $\mathrm{FB}(p) := \bigcup_{|t|\leq\varepsilon} \varphi^t[D(p)]$ be the *flow box* defined by $D(p)$. Using that $X \neq 0$, we see that $\mathrm{FB}(p)$ contains an open ball centered at p.
- By compactness, N is covered by finitely many flow boxes $\mathrm{FB}(p_1), \ldots, \mathrm{FB}(p_\ell)$.

Therefore $M = D(p_1) \cup \cdots \cup D(p_\ell)$ is a global Poincaré section for φ. With some extra work, we can make the discs $D(p_1), \ldots, D(p_\ell)$ pairwise disjoint, hence the return time function $\mathfrak{t} : M \to (0, \infty)$ is bounded away from zero and infinity. See [**LS19**, §2] for details.

Let $f : M \to M$ be the Poincaré return map of M, that is, $f(x) = \varphi^{\mathfrak{t}(x)}(x)$. The map f has discontinuities, with $\mathscr{D}^\pm = \{x \in M : f^{\pm 1}(x) \in \partial M\}$ (observe that in this case f is defined on all of M, but $f^{\pm 1}$ is discontinuous on $\mathscr{D}^\pm$); see Figure 8.

Nevertheless, where $f^{\pm 1}$ is continuous, its $C^{1+\beta}$ norm is uniformly bounded. This occurs because, at continuity points, $f^{\pm 1}$ has the form φ^τ where τ has uniformly bounded $C^{1+\beta}$ norm; see [**LS19**, Lemma 2.5] for details.

2.3.2. *Non-uniform hyperbolicity.* To apply the methods of §2.2, we only consider trajectories that do not approach $\mathscr{D}$ exponentially fast. Let d be the distance in M.

The non-uniformly hyperbolic locus NUH_χ^*. This is the set of points $x \in M$ satisfying conditions (NUH1)–(NUH3) of page 2609 and the following additional condition.

(NUH4) Subexponential convergence to $\mathscr{D}$:

$$\lim_{n\to\pm\infty} \frac{1}{n} \log d(f^n(x), \mathscr{D}) = 0.$$

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

The idea of looking at trajectories satisfying condition (NUH4) is not new. It goes back to Sinaĭ in the context of billiards [**Sin70**], which we will discuss in §2.4. See also the section 'Overcoming influence of singularities' in [**KSLP86**]. At the level of invariant measures, (NUH4) is related to the following notion.

f-adapted measure. An f-invariant measure on M is called *f-adapted* if the function $\log d(x, \mathscr{D}) \in L^1(\mu)$. A fortiori, $\mu(\mathscr{D}) = 0$.

By the Birkhoff ergodic theorem, if μ is f-adapted then (NUH4) holds μ-a.e. If in addition μ is χ-hyperbolic, then (NUH1)–(NUH3) hold μ-a.e. and therefore μ is carried by NUH_χ, that is, $\mu[\mathrm{NUH}_\chi] = 1$.

For each $x \in \mathrm{NUH}_\chi$, the linear map $C(x)$ can be defined as before, and Lemma 2.5 remains valid with the same proof. To define the Pesin chart Ψ_x, we just need to adjust its domain of definition, according to the distance of x to $\mathscr{D}$. Let $\delta(x) = \varepsilon^{3/\beta} d(x, \mathscr{D})$.

Pesin chart. The *Pesin chart* at $x \in \mathrm{NUH}_\chi$ is the map $\Psi_x : [-\delta(x), \delta(x)]^2 \to M$ defined by $\Psi_x := \exp_x \circ C(x)$.

We also redefine $Q(x)$ accordingly. Let $\rho(x) := d(\{f^{-1}(x), x, f(x)\}, \mathscr{D})$.

Parameter $Q(x)$. For $x \in \mathrm{NUH}_\chi$, let $Q(x) = \varepsilon^{3/\beta} \min\{\|C(f(x))^{-1}\|_{\mathrm{Frob}}^{-12/\beta}, \varepsilon\rho(x)\}$.

With this definition, the representation of f in Pesin charts $f_x := \Psi_{f(x)}^{-1} \circ f \circ \Psi_x$ is well defined in $[-10Q(x), 10Q(x)]^2$. (We take the opportunity to observe that the definition of $Q(x)$ in [**LS19**] has a small error, since it does not depend on $d(f(x), \mathscr{D})$ and so we cannot guarantee that f_x is well defined; see [**LS19**, Theorem 3.2 and Corollary 3.6]. Nevertheless, this can be easily fixed with the definition we give here.) Indeed, $Q(x) \le \varepsilon\delta(f(x))$ and so

$$(f \circ \Psi_x)([-10Q(x), 10Q(x)]^2) \subset \Psi_{f(x)}([-\delta(f(x)), \delta(f(x))]^2).$$

Again, in the domain $[-10Q(x), 10Q(x)]^2$ the map f_x is a small perturbation of a hyperbolic matrix.

Now define the parameters q, q^s, q^u, the set NUH_χ^*, and the graph transforms $\mathscr{F}_x^{s/u}$ as in the previous section, then construct local invariant manifolds for each $x \in \mathrm{NUH}_\chi^*$. We finish this section by proving an analogue of Lemma 2.7.

LEMMA 2.10. *If μ is an f-invariant probability measure supported on* NUH_χ*, then μ is supported on* NUH_χ^*.

Proof. By assumption, $\lim_{n\to\pm\infty}(1/n)\log d(f^n(x), \mathscr{D}) = 0$ for μ-a.e. $x \in M$. Let $\widetilde{Q}(x) = \varepsilon^{3/\beta}\|C(f(x))^{-1}\|_{\mathrm{Frob}}^{-12/\beta}$ be the 'old' Q. Since df is uniformly bounded, we can proceed exactly as in Lemma 2.7 to conclude that $\lim_{n\to\pm\infty}(1/n)\log \widetilde{Q}(f^n(x)) = 0$ for μ-a.e. $x \in M$. But then $\lim_{n\to\pm\infty}(1/n)\log Q(f^n(x)) = 0$ for μ-a.e. $x \in M$. □

2.4. *Maps with discontinuities and unbounded derivative.* Next, we consider surface maps with discontinuities and unbounded derivative. Added to the difficulty that Pesin

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

charts are defined in smaller domains, now $\|df^{\pm 1}\|$ can approach zero and infinity, so even the integrability condition in the Oseledets theorem is no longer automatic. The first development of Pesin theory in this context was [**KSLP86**], where the interest was in applying it to billiard maps. Since its beginnings, the ergodic theory of billiard maps was mainly focused on a reference Liouville measure. This is the case in [**KSLP86**], where the authors construct invariant manifolds Lebesgue almost everywhere and use them to prove ergodic theoretic properties such as the Ruelle inequality. On the contrary, in the next sections we follow the same approach as §2.2, not focusing on a particular measure but rather on the set of points with some hyperbolicity. The reference for this section is [**LM18**].

2.4.1. *Definitions and examples.* Let M be a compact surface, possibly with boundary. Again, we assume that M has diameter smaller than one. Let $\mathscr{D}^+$, $\mathscr{D}^-$ be closed subsets of M, and consider $f : M\backslash\mathscr{D}^+ \to M$ with inverse $f^{-1} : M\backslash\mathscr{D}^- \to M$. Let $\mathscr{D} := \mathscr{D}^+ \cup \mathscr{D}^-$ be the *set of discontinuities of f*. We require the following conditions on *f*.

Regularity of f. There are constants $0 < \beta < 1 < a$ and $\mathfrak{K} > 0$ such that for all $x \in M\backslash\mathscr{D}$ there is $d(x, \mathscr{D})^a < \mathfrak{r}(x) < d(x, \mathscr{D})$ such that if $D_x = B(x, \mathfrak{r}(x))$ then the following assumptions hold.

- If $y \in D_x$ then $\|df_y^{\pm 1}\| \leq d(x, \mathscr{D})^{-a}$.
- If $y_1, y_2 \in D_x$ and $f(y_1), f(y_2) \in D_{x'}$ then $\|\widetilde{df_{y_1}} - \widetilde{df_{y_2}}\| \leq \mathfrak{K} d(y_1, y_2)^\beta$, and if $y_1, y_2 \in D_x$ and $f^{-1}(y_1), f^{-1}(y_2) \in D_{x''}$ then $\|\widetilde{df^{-1}_{y_1}} - \widetilde{df^{-1}_{y_2}}\| \leq \mathfrak{K} d(y_1, y_2)^\beta$.

The first assumption says that $df^{\pm 1}$ blows up at most polynomially fast, and the second says that $df^{\pm 1}$ is locally β-Hölder. The examples to have in mind are billiard maps, as we now explain. Given a compact domain $T \subset \mathbb{R}^2$ or $T \subset \mathbb{T}^2$ with piecewise C^3 boundary, consider the straight-line motion of a particle inside T, with specular reflections in ∂T. The phase space of configurations is $M = \partial T \times [-\pi/2, \pi/2]$ with the convention that $(r, \theta) \in M$ represents r = collision position at ∂T and θ = angle of collision. Given $(r, \theta) \in M$, let (r^+, θ^+) be the next collision and (r^-, θ^-) be the previous collision. Let $\{r_1, \ldots, r_k\}$ be the break points of ∂T, and define:

$$\mathscr{D}^+ = \{r^+ = r_i \text{ for some } i\} \cup \left\{\theta^+ = \pm\frac{\pi}{2}\right\},$$

$$\mathscr{D}^- = \{r^- = r_i \text{ for some } i\} \cup \left\{\theta^- = \pm\frac{\pi}{2}\right\}.$$

The *billiard map* is $f : M\backslash\mathscr{D}^+ \to M$ defined by $f(r, \theta) = (r^+, \theta^+)$, with inverse $f : M\backslash\mathscr{D}^- \to M$ defined by $f(r, \theta) = (r^-, \theta^-)$. Since ∂T (usually) has two normal vectors at r_i, we cannot define $f^{\pm 1}(r, \theta)$ if $r^\pm = r_i$. When $\theta^\pm = \pm\pi/2$, the trajectory has a grazing collision, and $f^{\pm 1}$ is usually discontinuous on (r, θ). Furthermore, $df^{\pm 1}$ becomes arbitrarily large in a neighborhood of (r, θ). This justifies the choice of $\mathscr{D}^\pm$ above. See [**CM06**] for details.

Sinaĭ showed that f has a natural invariant Liouville measure $\mu_{\text{SRB}} = \cos\theta dr d\theta$, which is ergodic for dispersing billiards [**Sin70**]. Bunimovich constructed examples

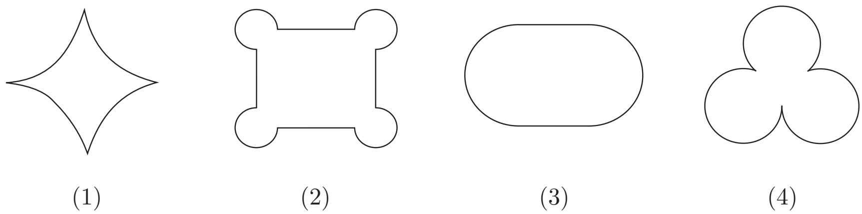

FIGURE 9. (1) is a Sinaĭ billiard table. The others are Bunimovich billiard tables: (2) is a pool table with pockets, (3) is a stadium, (4) is a flower.

of ergodic nowhere dispersing billiards [**Bun74a, Bun74b, Bun79**]. These billiards, known as *Bunimovich billiards*, are non-uniformly hyperbolic. See some examples in Figure 9. Recently, Baladi and Demers constructed measures of maximal entropy for some finite-horizon periodic Lorentz gases [**BD20**]; for more see §4.1.

2.4.2. *Non-uniform hyperbolicity.* We continue only considering trajectories that do not approach $\mathscr{D}$ exponentially fast, and define the non-uniformly hyperbolic locus NUH_χ as in §2.3.2. Similarly, an f-invariant measure μ is called *f-adapted* if $\log d(x, \mathscr{D}) \in L^1(\mu)$.

Since $\log \|df^{\pm 1}\|$ is usually unbounded, some measures might not satisfy the integrability condition in the Oseledets theorem. Due to the regularity of f, the functions $\log \|df^{\pm 1}\|$ and $\log d(x, \mathscr{D})$ are comparable, therefore $\log \|df^{\pm 1}\| \in L^1(\mu)$ when $\log d(x, \mathscr{D}) \in L^1(\mu)$. Hence the Oseledets theorem holds for f-adapted measures which shows that f-adaptability is a natural assumption. In particular, if μ is f-adapted and χ-hyperbolic, then (NUH1)–(NUH4) hold μ-a.e. and so μ is carried by NUH_χ. For each $x \in \mathrm{NUH}_\chi$, we define $C(x)$ as before, and Lemma 2.5 remains valid with the same proof.

Pesin chart. The *Pesin chart* at $x \in \mathrm{NUH}_\chi$ is $\Psi_x : [-d(x, \mathscr{D})^a, d(x, \mathscr{D})^a]^2 \to M$, defined by $\Psi_x := \exp_x \circ C(x)$.

The definition of $Q(x)$ is more complicated. Let $\rho(x) = d(\{f^{-1}(x), x, f(x)\}, \mathscr{D})$.

Parameter $Q(x)$. For $x \in \mathrm{NUH}_\chi$, define

$$Q(x) = \varepsilon^{3/\beta} \min\{\|C(x)^{-1}\|_{\mathrm{Frob}}^{-24/\beta}, \|C(f(x))^{-1}\|_{\mathrm{Frob}}^{-12/\beta} \rho(x)^{72a/\beta}\}.$$

As before, the choice of the powers is not canonical but just an artifact of the proof. The above definition depends on $f^{-1}(x), x, f(x)$, and is strong enough to construct local invariant manifolds, and to run the methods of Part two, §3. Firstly, in the domain $[-10Q(x), 10Q(x)]^2$ the representation of f in Pesin charts $f_x := \Psi_{f(x)}^{-1} \circ f \circ \Psi_x$ is a small perturbation of a hyperbolic matrix; see [**LM18**, Theorem 3.3]. Defining the parameters q, q^s, q^u, the set NUH_χ^*, and the graph transforms $\mathscr{F}_x^{s/u}$ as before, we construct local invariant manifolds for each $x \in \mathrm{NUH}_\chi^*$. Finally, we establish an analogue of Lemmas 2.7 and 2.10.

LEMMA 2.11. *If μ is an f-adapted probability measure supported on NUH_χ, then μ is supported on NUH_χ^*.*

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Proof. It is enough to show that

$$\lim_{n\to\pm\infty}\frac{1}{n}\log\|C(f^n(x))^{-1}\|_{\text{Frob}}=\lim_{n\to\pm\infty}\frac{1}{n}\log\rho(f^n(x))=0$$

for μ-a.e. $x\in M$. By (NUH4), the second equality holds. For the first equality, let $\widetilde{Q}(x)=\varepsilon^{3/\beta}\|C(f(x))^{-1}\|_{\text{Frob}}^{-12/\beta}$ be the 'old' Q, and observe that the regularity assumption on f implies that $\log[(\widetilde{Q}\circ f)/\widetilde{Q}]\in L^1(\mu)$ if and only if $\log d(x,\mathscr{D})\in L^1(\mu)$. Since μ is f-adapted, we get that $\log[(\widetilde{Q}\circ f)/\widetilde{Q}]\in L^1(\mu)$. Now proceed as in Lemma 2.7 to conclude that $\lim_{n\to\pm\infty}(1/n)\log\|C(f^n(x))^{-1}\|_{\text{Frob}}=0$ for μ-a.e. $x\in M$. □

3. *Part two: Symbolic dynamics*

Symbolic dynamics is an important tool for the understanding of ergodic and statistical properties of dynamical systems, both smooth and non-smooth. The field of symbolic dynamics is enormous and covers various contexts, from the study of symbolic spaces to the theory of complexity functions; see, for example, [**Fer99**, **Kit98**]. Here, we only discuss the use of symbolic dynamics to represent smooth dynamical systems. The main idea is simple and can be summarized in two steps: firstly, divide the phase space of a system into finitely or countably many pieces, which we call *rectangles*; secondly, instead of describing the trajectory of a point by the exact positions in the phase space, just record the sequence of rectangles that the trajectory visits. We call the second step above a *coding*. This procedure can be performed in a wide setting. For instance, any partition defines a coding in the usual way. Such flexibility allows its use in various contexts.

- Periodic points of continuous intervals maps: proof of the Sharkovsky theorem using *Markov graphs*; see, for example, [**BH11**].
- Milnor and Thurston's kneading theory of continuous intervals maps: description of the trajectory of the critical point with respect to monotonicity intervals [**MT88**].
- Geodesics on surfaces of constant negative curvature: Hadamard represented closed geodesics using sequences of symbols [**Had98**]; see also [**KU07**].

In this survey we focus on symbolic dynamics for smooth systems with some hyperbolicity, uniform and non-uniform. The final goal is to describe the invariant measures and ergodic theoretical properties of such systems, and for that a mere coding is not enough: it is important to recover codings, that is, to know which trajectories are coded in the same way. This reverse procedure is called *decoding*. Good codings are those for which we can satisfactorily decode. It has long been observed that uniform expansion provides a good decoding: two different trajectories eventually stay far apart and therefore cannot visit the same rectangles. This property, that different trajectories eventually stay far apart, is known as *expansivity*. It also occurs for (U)-systems, due to the exponential dichotomy of solutions mentioned in §2.1.

Another required property on the coding is that the space of sequences coding the trajectories should be as simple as possible and at the same time rich enough to reflect the structure of the original smooth system. This property is certainly satisfied if the rectangles have the *Markov property*, since in this case every path on the graph is naturally associated to a genuine orbit. Apart from some technical assumptions, when this occurs we call the partition a *Markov partition*.

Let us give a simple example of its usefulness. Let $f : K \to K$ be the horseshoe map described in Example 2 of §2.1.1 (Smale's horseshoe). Since K has a fractal structure, it seems rather complicated to understand its periodic points and invariant measures. But the system has a Markov partition that induces a continuous bijection $\pi : \Sigma \to K$ between the symbolic space $\Sigma = \{0, 1\}^{\mathbb{Z}}$ and K, that commutes the shift map on Σ and the map f. Hence we can analyze the dynamical properties of f by means of the dynamical properties of the shift map.

In the next sections, we explain how to construct Markov partitions for smooth systems with some hyperbolicity, both uniform and non-uniform. The conclusion is the existence of a *symbolic model*. Let us give the definitions. Let $\mathscr{G} = (V, E)$ be an oriented graph. We assume that V is finite or countable, and that for each $v, w \in V$ there is at most one edge $v \to w$.

Topological Markov shift. Let

$$\Sigma = \{\{v_n\}_{n\in\mathbb{Z}} \in V^{\mathbb{Z}} : v_n \to v_{n+1}, \forall n \in \mathbb{Z}\}$$

be the set of $\mathbb{Z}$-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* (TMS).

An element of Σ is denoted by $\underline{v} = \{v_n\}_{n\in\mathbb{Z}}$. We endow Σ with the distance $d(\underline{v}, \underline{w}) := \exp[-\min\{|n| : n \in \mathbb{Z} \text{ such that } v_n \neq w_n\}]$. Let $\Sigma^\#$ be the *recurrent set* of Σ, defined by

$$\Sigma^\# = \left\{\{v_n\}_{n\in\mathbb{Z}} \in \Sigma : \begin{array}{l} \exists v, w \in V \text{ such that } v_n = v \text{ for infinitely many } n > 0 \\ \text{and } v_n = w \text{ for infinitely many } n < 0 \end{array}\right\}.$$

When V is finite, $\Sigma^\# = \Sigma$. Let $f : M \to M$ be a diffeomorphism.

Symbolic model for diffeomorphism. A *symbolic model* for $f : M \to M$ is a triple (Σ, σ, π) where (Σ, σ) is a TMS and $\pi : \Sigma \to M$ is a Hölder continuous map such that $\pi \circ \sigma = f \circ \pi$ and the restriction $\pi \restriction_{\Sigma^\#}: \Sigma^\# \to \pi[\Sigma^\#]$ is finite-to-one.

Hence a symbolic model is a TMS together with a projection map π that commutes f and σ, and that is finite-to-one on $\pi[\Sigma^\#]$. A diffeomorphism can have many symbolic models. Some of them are bad, when $\pi[\Sigma^\#]$ is much smaller than the subsets where f displays an interesting dynamics. A good symbolic model is one for which $\pi[\Sigma^\#]$ contains the subset where f displays chaotic dynamics. For us, this occurs when $\pi[\Sigma^\#]$ carries χ-hyperbolic measures. To define a symbolic model for flows, we add the flow direction to the TMS.

Topological Markov flow. 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 pair (Σ_r, σ_r) by:

- $\Sigma_r = \{(\underline{v}, t) : \underline{v} \in \Sigma, 0 \leq t \leq r(\underline{v})\}$ with the identification $(\underline{v}, r(\underline{v})) \sim (\sigma(\underline{v}), 0)$;
- $\sigma_r = \{\sigma_r^t\}_{t\in\mathbb{R}} : \Sigma_r \to \Sigma_r$ the unit speed vertical flow on Σ_r, called the *suspension flow* (see Figure 10).

The pair (Σ_r, σ_r) is called a *topological Markov flow* (TMF).

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

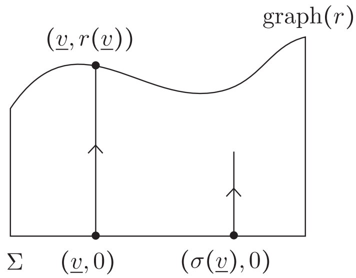

FIGURE 10. The suspension flow σ_r: starting at $(\underline{v}, 0)$, flow at unit speed until hitting the graph of r, then return to the basis via the identification $(\underline{v}, r(\underline{v})) \sim (\sigma(\underline{v}), 0)$ and continue flowing.

An element of Σ_r is denoted by $(\underline{v}, t)$. Let $\Sigma_r^\#$ be the *recurrent set* of Σ_r,

$$\Sigma_r^\# = \{(\underline{v}, t) \in \Sigma_r : \underline{v} \in \Sigma^\#\}.$$

See [**LS19**] for basic properties on (Σ_r, σ_r). Let $\varphi : M \to M$ be a flow.

Symbolic model for flow. A *symbolic model* for $\varphi : M \to M$ is a triple $(\Sigma_r, \sigma_r, \pi_r)$ where (Σ_r, σ_r) is a TMF and $\pi_r : \Sigma_r \to M$ is a Hölder continuous map such that $\pi_r \circ \sigma_r^t = \varphi^t \circ \pi_r$ for all $t \in \mathbb{R}$ for which the restriction $\pi_r \restriction_{\Sigma_r^\#} : \Sigma_r^\# \to \pi_r[\Sigma_r^\#]$ is finite-to-one.

3.1. *Symbolic dynamics for uniformly hyperbolic systems.* There are at least two general ways of constructing Markov partitions for uniformly hyperbolic systems. One of them, due to Sinaĭ, is called the *method of successive approximations* [**Sin68a, Sin68b**]. The second, due to Bowen, is called the *method of pseudo-orbits* [**Bow08**]. Their common feature is the use of the local invariant manifolds constructed in Part one, §2, as dynamically defined systems of coordinates. For completeness and ease of understanding, we also describe the construction of Adler and Weiss for hyperbolic toral automorphisms [**AW67**]. Since for non-uniformly hyperbolic systems we will make use of the method of pseudo-orbits, we will only sketch the other techniques, and the details can be found in the original papers. We start by defining Markov partitions (and their flow counterpart) in the context of uniformly hyperbolic systems, and explain how they generate symbolic models. It is important to mention that, for uniformly hyperbolic systems, the vertex set V of the oriented graph $\mathscr{G}$ is finite.

3.1.1. *Markov partitions/sections.* Let $f : M \to M$ be a diffeomorphism. As already mentioned, our goal is to construct a partition of M so that the dynamics of f can be represented by a TMS. Let $\mathscr{G} = (V, E)$ be the graph defining the TMS (Σ, σ). The vertex set V is the set of partition elements, and each edge in E will represent one possible transition by the iteration of f. If $v_0 \to v_1$ and $v_1 \to v_2$ are edges, then their concatenation is a path from v_0 to v_2. This property, translated to the dynamics of f, is the *Markov property*. More precisely, if R_0, R_1, R_2 are the partition elements associated to v_0, v_1, v_2, then there is a point $x \in R_0$ such that $f(x) \in R_1$ and there is a point $y \in R_1$ such that $f(y) \in R_2$. The Markov property ensures that there is a point $z \in R_0$ such that $f(z) \in R_1$ and $f^2(z) \in R_2$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Imagine, for a moment, that f is uniformly expanding. If we define edges $R_0 \to R_1$ when $f(R_0) \cap R_1 \neq \emptyset$, then the above property is not guaranteed. If instead we define $R_0 \to R_1$ when $f(R_0) \supset R_1$, then the above concatenation holds. When f is a uniformly hyperbolic diffeomorphism, the definition of edges requires two inclusions, one for each invariant direction. Let us give the definitions. The references for the discussion in this section are [**Bow08**] for diffeomorphisms and [**Bow73**] for flows. Let $f : M \to M$ be an Axiom A diffeomorphism, and fix $\varepsilon > 0$ small. By Part one, §2, each $x \in \Omega(f)$ has a local stable manifold $W^s_{\rm loc}(x) = V^s[x]$ and a local unstable manifold $W^u_{\rm loc}(x) = V^u[x]$. By definition, $W^{s/u}_{\rm loc}(x)$ is tangent to $E^{s/u}_x$ at x, hence $W^s_{\rm loc}(x)$, $W^u_{\rm loc}(x)$ are transversal at x. The maps $x \in \Omega(f) \mapsto W^{s/u}_{\rm loc}(x)$ are continuous; see, for example, [**Shu87**, Theorem 6.2(2)]. Hence, if $x, y \in \Omega(f)$ with $\mathrm{dist}(x, y) \ll 1$ then $W^s_{\rm loc}(x)$ and $W^u_{\rm loc}(y)$ intersect transversally at a single point. Fix $\delta \ll \varepsilon$, and consider the following definition.

Smale bracket. For $x, y \in \Omega(f)$ with $d(x, y) < \delta$, the *Smale bracket* of x and y is defined by $\{[x, y]\} := W^s_{\rm loc}(x) \cap W^u_{\rm loc}(y)$.

For $R \subset \Omega(f)$, let R^* denote the interior of R in the induced topology of $\Omega(f)$.

Rectangle. A subset $R \subset \Omega(f)$ is called a *rectangle* if it satisfies the following properties.

(1) *Regularity*. $R = \overline{R^*}$ and $\mathrm{diam}(R) < \delta$.

(2) *Product structure*. $x, y \in R \Rightarrow [x, y] \in R$.

The product structure means that R is a rectangle in the system of coordinates given by the local invariant manifolds. Let $W^{s/u}(x, R) := W^{s/u}_{\rm loc}(x) \cap R$. Regardless of whether $W^{s/u}_{\rm loc}(x)$ are smooth manifolds, since $\Omega(f)$ is usually a fractal set, $W^{s/u}(x, R)$ are also usually fractal.

It is easy to construct rectangles: given $\rho > 0$, let $W^{s/u}_\rho(x) = W^{s/u}_{\rm loc}(x) \cap B(x, \rho)$; if $x \in \Omega(f)$, then $[W^u_\rho(x) \cap \Omega(f), W^s_\rho(x) \cap \Omega(f)]$ is a rectangle for all $\rho > 0$ small enough. Let $\mathscr{R}$ be a finite cover of $\Omega(f)$ by rectangles.

Markov partition. $\mathscr{R}$ is called a *Markov partition* for f if it satisfies the following properties.

(1) *Disjointness*. The elements of $\mathscr{R}$ can only intersect at their boundaries (boundaries are considered with respect to the relative topology of $W^{s/u}(x, R)$; see [**Bow08**]).

(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*. The two latter inclusions represent two Markov properties, one for each invariant direction. Geometrically, they ensure that if two rectangles intersect, then the intersection occurs all the way from one side to the other, with respect to the system of coordinates of the local invariant manifolds; see Figure 11. We stress that, while here every rectangle has non-empty interior, in the non-uniformly hyperbolic situation we will not be able to guarantee this.

Now let $\varphi : M \to M$ be an Axiom A flow. Recall the definitions of §2.3.1. Given an interval $I \subset \mathbb{R}$ and $Y \subset M$, let $\varphi^I(Y) := \bigcup_{t \in I} \varphi^t(Y)$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

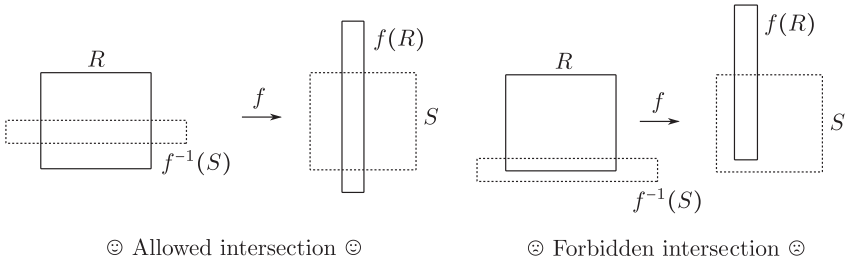

FIGURE 11. The Markov property: if $f(R)$ intersects S non-trivially, then $f(R)$ crosses S completely all the way from one side to the other.

Proper section. A finite family $\mathscr{M} = \{B_1, \ldots, B_n\}$ is a *proper section of size* α if there are closed differentiable discs $D_1, \ldots, D_n$ transverse to the flow direction such that the following properties hold.

(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*. $B_i \subset \text{int}(D_i)$ 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, $D_i \cap D_j = \emptyset$.

For simplicity, denote $B_1 \cup \cdots \cup B_n$ also by $\mathscr{M}$. Let $f : \mathscr{M} \to \mathscr{M}$ be the Poincaré return map of $\mathscr{M}$, and $\mathfrak{t} : \mathscr{M} \to (0, \infty)$ the return time function. By properties (2) and (4), $0 < \inf \mathfrak{t} \leq \sup \mathfrak{t} \leq \alpha$. By transversality, the stable/unstable directions of φ project to stable/unstable directions of the Poincaré map f. Also, local invariant manifolds of f are projections, in the flow direction, of local invariant manifolds of φ, and we can similarly define the Smale bracket $[\cdot, \cdot]$ for f.

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

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

Considering points in $\mathscr{M}'$ 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 $f(B_i) \cap B_j$ is a subset of ∂B_j.

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

Markov section. $\mathscr{M}$ is called a *Markov section* of size α for φ 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

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

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Above, $W^s(x, B_i) = \{[x, y] : y \in B_i\}$ is the intersection of the local stable manifold of f at x with B_i. The definition of $W^u(y, B_j)$ is similar.

3.1.2. *Markov partitions/sections generate symbolic models.* If $\mathscr{R}$ is a Markov partition for f, then we have a symbolic model for f.

- $\mathscr{G} = (V, E)$ with $V = \mathscr{R}$ and $E = \{R \to S : f(R^*) \cap S^* \neq \emptyset\}$.
- $\pi : \Sigma \to \Omega(f)$ is defined for $\underline{R} = \{R_n\}_{n\in\mathbb{Z}} \in \Sigma$ by

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

Alternatively, $\pi(\underline{R})$ is the unique $x \in \Omega(f)$ such that $f^n(x) \in R_n$, for all $n \in \mathbb{Z}$. The map π is well defined due to the Markov property and uniform hyperbolicity. Clearly $f \circ \pi = \pi \circ \sigma$. Additionally, π is a finite-to-one continuous surjection that is one-to-one on a residual subset of $\Omega(f)$; see [Bow08, Theorem 3.18] for details.

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

- $\mathscr{G} = (V, E)$ with $V = \mathscr{M}$ and $E = \{B_i \to B_j : \exists x \in \mathscr{M}'$ such that $x \in B_i^*, f(x) \in B_j^*\}$.
- $\pi : \Sigma \to \mathscr{M}$ is defined for $\underline{B} = \{B_n\}_{n\in\mathbb{Z}} \in \Sigma$ by

$$\{\pi(\underline{B})\} := \bigcap_{n\geq 0} f^n(B_{-n}) \cap \cdots \cap f^{-n}(B_n) = \bigcap_{n\geq 0} \overline{f^n(B_{-n}) \cap \cdots \cap f^{-n}(B_n)}.$$

- $r : \Sigma \to \mathbb{R}$ is defined by $r := \mathfrak{t} \circ \pi$.
- $\pi_r : \Sigma_r \to \Omega(\varphi)$ is defined by $\pi_r(\underline{B}, t) := \varphi^t[\pi(\underline{B})]$.

Again, π is well defined because of the Markov property and uniform hyperbolicity, and satisfies $f \circ \pi = \pi \circ \sigma$. Also, π is a finite-to-one continuous surjection that is one-to-one on $\mathscr{M}'$; see [Bow73] for details.

Therefore, to get symbolic models for uniformly hyperbolic symbolic systems, it is enough to construct Markov partitions/sections.

3.1.3. *Markov partitions for two-dimensional hyperbolic toral automorphisms.* This method, developed by Adler and Weiss [AW67], constructs finite Markov partitions for two-dimensional hyperbolic toral automorphisms. A particular case was constructed by Berg [Ber68]. Consider the cat map introduced in Example 1 of §2.1.1 , and let $\vec{0} = (0, 0) \in \mathbb{T}^2$. Clearly, $f(\vec{0}) = \vec{0}$. Since the matrix A is hyperbolic, $\vec{0}$ has two eigendirections; let us call W^s the contracting one and W^u the expanding one. By linearity, W^s and W^u are the (global) stable and unstable manifolds of $\vec{0}$.

To obtain a Markov partition, we construct a fundamental domain of $\mathbb{T}^2$ whose sides are pieces of W^s and W^u, and then subdivide this domain into finitely many rectangles satisfying the Markov property. In Figure 12, we draw one possibility for the tessellation of $\mathbb{R}^2$ by one such fundamental domain. For a general two-dimensional hyperbolic toral automorphism, the construction of the fundamental domain consists of two steps.

Step 1. Take a cover $\mathscr{R}$ of $\mathbb{T}^2$ by finitely many rectangles whose sides belong to W^s and W^u such that every non-trivial intersection $f(R^*) \cap S^*$ is connected, that is, $f(R^*)$ does not intersect S^* ‘twice’.

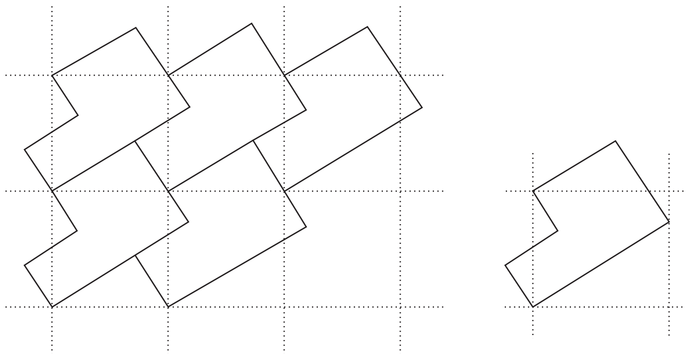

FIGURE 12. A tessellation of $\mathbb{R}^2$ by fundamental domains whose sides are parallel to W^s and W^u.

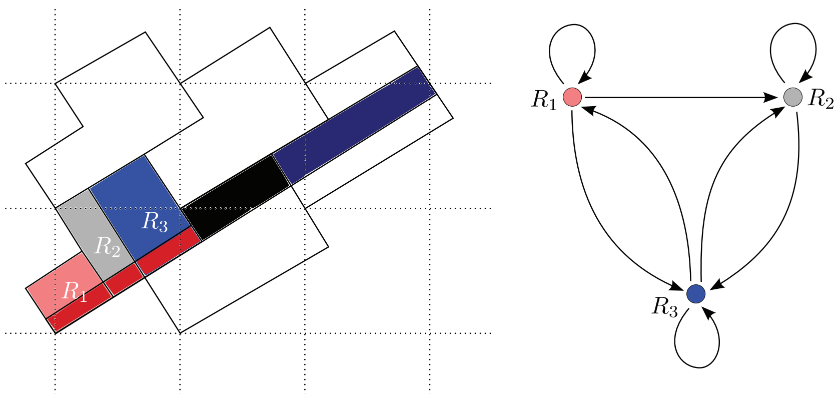

FIGURE 13. A Markov partition for the cat map by three rectangles R_1, R_2, R_3, and the graph defining the respective TMS.

Step 2. Since f contracts W^s, which is the stable manifold of the fixed point $\vec{0}$, 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 finite Markov partition; see [**AW70**] for details. The projection map $\pi : \Sigma \to \mathbb{T}^2$ is a finite-to-one continuous surjection that is one-to-one on the set $\{x \in \mathbb{T}^2 : f^n(x) \in \bigcup_{R\in\mathscr{R}} R^*, \forall n \in \mathbb{Z}\}$.

In our example, it is enough to divide the fundamental domain into three rectangles R_1, R_2, R_3 as in Figure 13. We leave it as an exercise for the reader to show that the images $f(R_1), f(R_2), f(R_3)$ are as depicted in Figure 13, so that $\{R_1, R_2, R_3\}$ is a Markov partition. The graph defining the TMS is also depicted in Figure 13.

For higher-dimensional hyperbolic toral automorphisms, a similar construction works, but there is an important difference from the two-dimensional case: the boundary of a Markov partition is not smooth [**Bow78a**].

3.1.4. *The method of successive approximations for diffeomorphisms.* This method, due to Sinaĭ [**Sin68a, Sin68b**], provides Markov partitions for Anosov diffeomorphisms. It

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

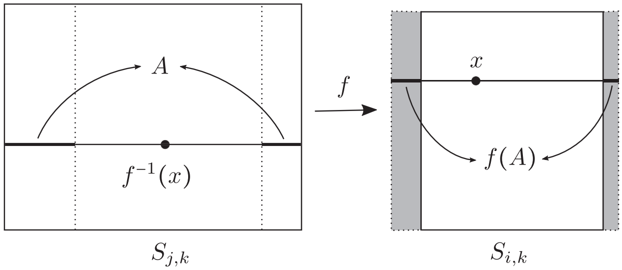

FIGURE 14. The horizontal component of $S_{i,k+1}\backslash S_{i,k}$ is the union of sets of the form $f(A)$, where A is a horizontal subset of $S_{j,k}\backslash S_{j,k-1}$. Above, A is composed of the two ticker segments in the left figure and the added set is in gray in the right figure.

was later modified by Bowen to also work for Axiom A diffeomorphisms [Bow70a]. The construction consists of three main steps. Below, we explain them for Anosov diffeomorphisms.

Step 1 (Coarse graining). Let $\mathscr{T} = \{T_i\}$ be a finite cover of M by rectangles (as we have argued in §3.1.1, 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 Markov cover.

Let us understand the above definitions. Representing the stable direction by the horizontal direction, we identify what are the horizontal components of $S_{i,k+1}\backslash S_{i,k}$; see Figure 14. Start by observing that the horizontal component of $S_{i,1}\backslash S_{i,0}$ is the union of sets of the form $f(A)$, where $A = W^s(f^{-1}(x), S_{j,0})\backslash f^{-1}(W^s(x, S_{i,0}))$. Each such A has diameter less than 1, hence $f(A)$ has diameter less than κ, thus $S_{i,1}$ equals the union of $S_{i,0}$ and a set of horizontal diameter less than κ. Similarly, the horizontal component of $S_{i,2}\backslash S_{i,1}$ is the union of sets of the form $f(A)$, where A is a horizontal subset of $S_{j,1}\backslash S_{j,0}$, therefore $S_{i,2}$ equals the union of $S_{i,1}$ and a set of horizontal diameter less than κ^2. By induction, $S_{i,k+1}$ equals the union of $S_{i,k}$ and a set of horizontal diameter less than κ^{k+1}. This shows that each S_i is well defined, and the same occurs for each U_i.

Step 3 (Bowen–Sinaĭ refinement). To destroy non-trivial intersections, refine $\mathscr{Z}$ as follows. For Z_i, let $I_i = \{j : Z_i^* \cap Z_j^* \neq \emptyset\}$. For $j \in I_i$, let $\mathscr{E}_{ij}$ = cover of Z_i by rectangles (see Figure 15):

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

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

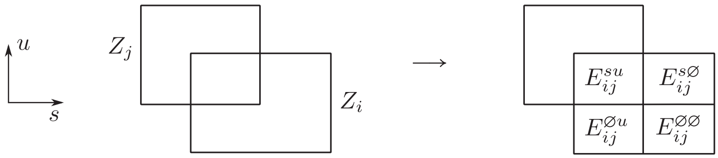

FIGURE 15. $\mathscr{E}_{ij} = \{E^{su}_{ij}, E^{s\emptyset}_{ij}, E^{\emptyset u}_{ij}, E^{\emptyset\emptyset}_{ij}\}$ is a cover of Z_i by rectangles.

$$E^{\emptyset u}_{ij} := \overline{\{x \in Z^*_i : W^s(x, Z_i) \cap Z_j = \emptyset,\, W^u(x, Z_i) \cap Z^*_j \neq \emptyset\}},$$
$$E^{\emptyset\emptyset}_{ij} := \overline{\{x \in Z^*_i : W^s(x, Z_i) \cap Z_j = \emptyset,\, W^u(x, Z_i) \cap Z_j = \emptyset\}}.$$

Hence $\mathscr{R}$:= cover defined by $\{\mathscr{E}_{ij} : Z_i \in \mathscr{Z}, j \in I_i\}$ is a Markov partition for *f*, and the induced $\pi : \Sigma \to M$ is a finite-to-one continuous surjection that is one-to-one on $\{x \in M : f^n(x) \in \bigcup_{R\in\mathscr{R}} R^*, \forall n \in \mathbb{Z}\}$.

3.1.5. *The method of successive approximations for flows.* Ratner applied the method of successive approximations to three-dimensional Anosov flows [**Rat69**]. Later she extended it to higher-dimensional Anosov flows [**Rat73**], and Bowen used it for Axiom A flows [**Bow73**]. Below, we follow Bowen's construction. As usual, the main difficulty when dealing with flows is the presence of discontinuities for the Poincaré return map.

Consider a proper section $\mathscr{C}$. Since the stable/unstable directions of φ project to stable/unstable directions of the Poincaré map *f*, 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 (Step 2 of §3.1.4), proceed as follows.

- Take $L > 0$ large such that for every $x \in R \in \mathscr{R}$ there are $C^+, C^- \in \mathscr{C}$ such that $\varphi^L(W^s_{\text{loc}}(x)) \subset \varphi^{[-\alpha,\alpha]}(C^+)$ and $\varphi^{-L}(W^u_{\text{loc}}(x)) \subset \varphi^{[-\alpha,\alpha]}(C^-)$. The existence of L follows from the uniform hyperbolicity of φ.
- For each such x, take a neighborhood $V \ni x$ small enough such that $\varphi^L(V) \subset \varphi^{[-\alpha,\alpha]}(C^+)$ and $\varphi^{-L}(V) \subset \varphi^{[-\alpha,\alpha]}(C^-)$, and define $f^+_V : V \to C^+$ and $f^-_V : V \to C^-$ by

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

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

For the details see [**Bow73**].

3.1.6. *The method of pseudo-orbits.* Bowen provided an alternative method to construct Markov partitions for Axiom A diffeomorphisms [**Bow08**]. His idea was to use the

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

theory of pseudo-orbits and shadowing, which explores the expected richness on the orbit structure of uniformly hyperbolic systems. These notions appeared in the qualitative theory of structural stability for uniformly hyperbolic systems. Indeed, Anosov considered a version of pseudo-orbits for flows, which he called ε-trajectories, and used them to prove that Anosov flows are structurally stable; see [**Ano70**, Theorem 1].

Let $f : M \to M$ be an invertible map. An orbit of f is a sequence $\{x_n\}_{n\in\mathbb{Z}}$ such that $f(x_n) = x_{n+1}$ for all $n \in \mathbb{Z}$, while a *pseudo-orbit* is a sequence $\{x_n\}_{n\in\mathbb{Z}}$ such that $f(x_n) \approx x_{n+1}$ for all $n \in \mathbb{Z}$. In other words, a pseudo-orbit is an orbit up to small errors at each iteration. This is exactly what a computer returns when we try to iterate a map: due to roundoff errors, the sequence is not a real orbit but just a pseudo-orbit. Since a hyperbolic matrix remains hyperbolic after a small perturbation, Theorem 2.2 holds for pseudo-orbits, as we will see below: changing $f(x)$ to some nearby y, we can represent f in the Lyapunov charts Ψ_x and Ψ_y and still obtain a small perturbation of a hyperbolic matrix. This is the main tool to introduce the symbolic model. To maintain consistency with Part one, §2, we continue to assume that M is a surface.

3.1.6.1. *Pseudo-orbits.* Recall the definition of Lyapunov charts from §2.1.5: for $\varepsilon > 0$ small enough, we let $Q = \varepsilon^{3/\beta}$ and define, for each $x \in M$, its Lyapunov chart $\Psi_x : [-Q, Q]^2 \to M$. Recall that Ψ_x is 2-Lipschitz and its inverse is $2\mathscr{L}$-Lipschitz. The splitting $E^s \oplus E^u$ is continuous, so there is $\delta = \delta(\varepsilon) > 0$ such that if $d(x, y) < \delta$ then $\|\Psi_y^{-1} \circ \Psi_x - \mathrm{Id}\|_{1+\beta/2} < \varepsilon^3$, where the norm is taken in $[-Q, Q]^2$. (The composition $\Psi_y^{-1} \circ \Psi_x$ is well defined in $[-Q, Q]^2$. To see this, fix $\varepsilon > 0$ small enough so that each Ψ_x is well defined in the larger domain $[-10\mathscr{L}Q, 10\mathscr{L}Q]^2$. Taking $\delta = \delta(\varepsilon) > 0$ small enough, if $d(x, y) < \delta$ then $\Psi_x([-Q, Q]^2) \subset B(x, 4Q) \subset B(y, 5Q) \subset \Psi_y([-10\mathscr{L}Q, 10\mathscr{L}Q]^2)$.)

ε-overlap. Two Lyapunov charts Ψ_x, Ψ_y are said to *ε-overlap* if $d(x, y) < \delta$. When this happens, we write $\Psi_x \overset{\varepsilon}{\approx} \Psi_y$.

Hence, if two points are close enough, the charts they define are essentially the same. This notation is somewhat redundant for uniformly hyperbolic systems, but we prefer to state it as above because it helps understand the symbolic model and the difficulties when we consider non-uniformly hyperbolic systems.

If $\Psi_{f(x)} \overset{\varepsilon}{\approx} \Psi_y$, then we can write f in the Lyapunov charts Ψ_x and Ψ_y as $f_{x,y} = \Psi_y^{-1} \circ f \circ \Psi_x$. Since $f_{x,y} = \Psi_y^{-1} \circ \Psi_{f(x)} \circ f_x =: g \circ f_x$, where $g := \Psi_y^{-1} \circ \Psi_{f(x)}$ is a small perturbation of the identity, the map $f_{x,y}$ is again a small perturbation of a hyperbolic matrix. The same reasoning happens if $\Psi_{f^{-1}(y)} \overset{\varepsilon}{\approx} \Psi_x$, in which case $f_{x,y}^{-1} = \Psi_x^{-1} \circ f^{-1} \circ \Psi_y$, the representation of f^{-1} in the Lyapunov charts Ψ_x and Ψ_y, is also a small perturbation of a hyperbolic matrix. This is summarized in the next theorem, which is the version of Theorem 2.2 in the present context.

THEOREM 3.1. *For all $\varepsilon > 0$ small enough, if $\Psi_{f(x)} \overset{\varepsilon}{\approx} \Psi_y$, then $f_{x,y}$ is well defined in $[-Q, Q]^2$ and can be written as $f_{x,y}(v_1, v_2) = (Av_1 + h_1(v_1, v_2), Bv_2 + h_2(v_1, v_2))$,* *where:*

(1) $|A|, |B^{-1}| < \lambda$ *(cf. Lemma 2.1);*

(2) $\|h_i\|_{1+\beta/3} < \varepsilon$ *for* $i = 1, 2$.

If $\Psi_{f^{-1}(y)} \overset{\varepsilon}{\approx} \Psi_x$ *then a similar statement holds for* $f_{x,y}^{-1}$.

Observe that, in contrast to Theorem 2.2, above we can only control the $C^{1+\beta/3}$ norm. This slight decrease is necessary to keep the estimates of size ε.

Edge. We write $\Psi_x \overset{\varepsilon}{\to} \Psi_y$ if $\Psi_{f(x)} \overset{\varepsilon}{\approx} \Psi_y$ and $\Psi_{f^{-1}(y)} \overset{\varepsilon}{\approx} \Psi_x$.

Conditions $\Psi_{f(x)} \overset{\varepsilon}{\approx} \Psi_y$ and $\Psi_{f^{-1}(y)} \overset{\varepsilon}{\approx} \Psi_x$ are called *nearest neighbor conditions*.

Pseudo-orbit. A sequence of Lyapunov charts $\{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ is called a pseudo-orbit if $\Psi_{x_n} \overset{\varepsilon}{\to} \Psi_{x_{n+1}}$ for all $n \in \mathbb{Z}$.

We point out that classically a pseudo-orbit is a sequence of points instead of Lyapunov charts, but so far these notions coincide, since $\Psi_{x_n} \overset{\varepsilon}{\to} \Psi_{x_{n+1}}$ is equivalent to $d(f(x_n), x_{n+1}) < \delta$ and $d(f^{-1}(x_{n+1}), x_n) < \delta$.

3.1.6.2. *Graph transforms.* Assume that $\Psi_x \overset{\varepsilon}{\to} \Psi_y$. Since $f_{xy}^{\pm 1}$ are perturbations of hyperbolic matrices, we can proceed as in §2.1.6 and define graph transforms between admissible manifolds at Ψ_x and Ψ_y. First, we need to redefine admissibility. For instance, we can no longer require $F(0) = 0$, since this property is not preserved by the maps $f_{x,y}^{\pm 1}$ (unless $f(x) = y$). For ease of exposition, we continue not to prescribe the precision in this definition.

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

Let $\mathscr{M}_x^s, \mathscr{M}_x^u$ be the space of all s, u-admissible manifolds at Ψ_x respectively, and introduce metrics on $\mathscr{M}_x^{s/u}$ as before. Assume that $\Psi_x \overset{\varepsilon}{\to} \Psi_y$.

Graph transforms $\mathscr{F}_{x,y}^s, \mathscr{F}_{x,y}^u$. The stable graph transform $\mathscr{F}_{x,y}^s : \mathscr{M}_y^s \to \mathscr{M}_x^s$ is the map that sends $V^s \in \mathscr{M}_y^s$ to the unique $\mathscr{F}_{x,y}^s[V^s] \in \mathscr{M}_x^s$ with representing function F such that $\Psi_x\{(t, F(t)) : |t| \le Q\} \subset f^{-1}(V^s)$. Similarly, the unstable graph transform $\mathscr{F}_{x,y}^u : \mathscr{M}_x^u \to \mathscr{M}_y^u$ is the map that sends $V^u \in \mathscr{M}_x^u$ to the unique $\mathscr{F}_{x,y}^u[V^u] \in \mathscr{M}_y^u$ with representing function G such that $\Psi_y\{(G(t), t) : |t| \le Q\} \subset f(V^u)$.

Again, the hyperbolicity of $f_{x,y}^{\pm 1}$ implies the following result.

THEOREM 3.2. *If* $\Psi_x \overset{\varepsilon}{\to} \Psi_y$, *then* $\mathscr{F}_{x,y}^s$ *and* $\mathscr{F}_{x,y}^u$ *are well-defined contractions.*

Consequently, each pseudo-orbit has local stable and unstable manifolds.

Stable/unstable manifolds. The stable manifold of $\underline{v} = \{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ 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 $\{V_n\}_{n\geq 0}$ with $V_n \in \mathscr{M}^s_{x_n}$. The *unstable manifold* of $\underline{v}$ 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 $\{V_n\}_{n\leq 0}$ with $V_n \in \mathscr{M}^u_{x_n}$.

The observations at the end of §2.1.6 can be repeated verbatim. In particular, $V^s[\underline{v}]$ only depends on the future $\{\Psi_{x_n}\}_{n\geq 0}$, while $V^u[\underline{v}]$ only depends on the past $\{\Psi_{x_n}\}_{n\leq 0}$.

Shadowing. $\underline{v} = \{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ is said to shadow x if $f^n(x) \in \Psi_{x_n}([-Q, Q]^2)$, for all $n \in \mathbb{Z}$.

THEOREM 3.3. (Shadowing lemma) *Every pseudo-orbit $\underline{v}$ shadows a unique point* $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$.

This follows from the hyperbolicity of each $f^{\pm 1}_{x_n,x_{n+1}}$.

3.1.6.3. *Construction of a Markov partition.* We now explain how Bowen used the above tools to construct a Markov partition for f. The construction involves two codings, the first being usually infinite-to-one and the second finite-to-one. We divide the construction into three steps. Let $f : M \to M$ be Axiom A, and let $L > 1$ be a Lipschitz constant for $f^{\pm 1}$.

Step 1 (Coarse graining). Fix a finite subset $X \subset \Omega(f)$ that is $(\delta/2L)$-dense in $\Omega(f)$, and let $\mathscr{A} = \{\Psi_x : x \in X\}$. Let $\mathscr{G} = (V, E)$ be the oriented graph with vertex set $V = \mathscr{A}$ and edge set $E = \{\Psi_x \overset{\varepsilon}{\to} \Psi_y\}$, and let (Σ, σ) be the TMS defined by $\mathscr{G}$. Observe that an element of Σ is a pseudo-orbit.

Step 2 (Infinite-to-one extension). Using the shadowing lemma, define a map $\pi : \Sigma \to \Omega(f)$ by

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

The map π has the following properties.

- π is surjective. For every $x \in \Omega(f)$, choose $\{x_n\}_{n\in\mathbb{Z}} \subset X$ such that $d(f^n(x), x_n) < \delta/2L$. Then $\underline{v} = \{\Psi_{x_n}\}_{n\in\mathbb{Z}}$ is a pseudo-orbit, since

$$\begin{aligned} d(f(x_n), x_{n+1}) &\leq d(f(x_n), f^{n+1}(x)) + d(f^{n+1}(x), x_{n+1}) \\ &\leq Ld(x_n, f^n(x)) + d(f^{n+1}(x), x_{n+1}) < \delta/2 + \delta/2L < \delta, \end{aligned}$$

  and similarly $d(f^{-1}(x_{n+1}), x_n) < \delta$. Clearly, $\pi(\underline{v}) = x$.
- $\pi \circ \sigma = f \circ \pi$. This follows from the shadowing lemma, since if $\underline{v}$ shadows x then $\sigma(\underline{v})$ shadows $f(x)$.
- π is usually infinite-to-one. Imagine, for example, that for some $x \in \Omega(f)$ there are $x_n, y_n \in X$ such that $d(f^n(x), x_n) < \delta/2L$ and $d(f^n(x), y_n) < \delta/2L$. Any choice of $z_n \in \{x_n, y_n\}$ defines a pseudo-orbit $\{\Psi_{z_n}\}_{n\in\mathbb{Z}}$ that shadows x, hence $\pi^{-1}(x)$ has cardinality at least $2^{\mathbb{Z}}$, which is uncountable.

The third property above is, in general, unavoidable. Hence, (Σ, σ, π) is *not* a symbolic model for *f*. But the Markov structure of Σ induces, via π, a cover of $\Omega(f)$ satisfying a (symbolic) Markov property.

The Markov cover $\mathscr{Z}$. Let $\mathscr{Z} := \{Z(v) : v \in \mathscr{A}\}$, where

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

In other words, $\mathscr{Z}$ is the family defined by the natural partition of Σ into cylinder at the zeroth position. In general, each $Z(v)$ is fractal. Admissible manifolds allow us to define *invariant fibers* inside each $Z \in \mathscr{Z}$. Let $Z = Z(v)$.

s/u-fibers in $\mathscr{Z}$. Given $x \in Z$, let $W^s(x, Z) := V^s[\underline{v}] \cap Z$ be the s-fiber of x in Z for some (any) $\underline{v} = \{v_n\}_{n\in\mathbb{Z}} \in \Sigma$ such that $\pi(\underline{v}) = x$ and $v_0 = v$. Similarly, let $W^u(x, Z) := V^u[\underline{v}] \cap Z$ be the u-fiber of x in Z.

Observe that, while $V^{s/u}[\underline{v}]$ are smooth manifolds, $W^{s/u}(x, Z)$ is usually fractal. Below we collect the main properties of $\mathscr{Z}$.

PROPOSITION 3.4. *The following statements hold for all* $\varepsilon > 0$ *small enough.*

(1) *Covering property.* $\mathscr{Z}$ *is a cover of* $\Omega(f)$.
(2) *Product structure. For every* $Z \in \mathscr{Z}$ *and every* $x, y \in Z$*, the intersection* $W^s(x, Z) \cap W^u(y, Z)$ *consists of a single point, and this point belongs to Z.*
(3) *Symbolic Markov property. If* $x = \pi(\underline{v})$ *with* $\underline{v} = \{v_n\}_{n\in\mathbb{Z}} \in \Sigma$*, then*

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

Statement (1) follows from the surjectivity of π. To prove (2), we define a Smale bracket $[\cdot, \cdot]_Z$ for each $Z \in \mathscr{Z}$ as follows. Write $Z = Z(v)$, and let $x = \pi(\underline{v})$, $y = \pi(\underline{w})$ where $\underline{v} = \{v_n\}_{n\in\mathbb{Z}}$, $\underline{w} = \{w_n\}_{n\in\mathbb{Z}} \in \Sigma$ with $v_0 = w_0 = v$. Then $W^s(x, Z) \cap W^u(y, Z)$ consists of a unique point $z = \pi(\underline{u})$ where $\underline{u} = \{u_n\}_{n\in\mathbb{Z}}$ is defined by:

$$u_n = \begin{cases} v_n & \text{if } n \geq 0, \\ w_n & \text{if } n \leq 0. \end{cases}$$

The equality $z = \pi(\underline{u})$ follows from the shadowing lemma. Observe that $z \in Z$. We write $z =: [x, y]_Z$. Finally, part (3) follows from the Markov structure of Σ. At this point, it is also important to show that the above definitions are compatible among the elements of $\mathscr{Z}$.

LEMMA 3.5. *The following statements hold for all* $\varepsilon > 0$ *small enough.*

(1) *Compatibility. If* $x, y \in Z(v_0)$ *and* $f(x), f(y) \in Z(v_1)$ *with* $v_0 \overset{\varepsilon}{\to} v_1$ *then* $f([x, y]_{Z(v_0)}) = [f(x), f(y)]_{Z(v_1)}$.
(2) *Overlapping charts properties. If* $Z = Z(\Psi_x)$, $Z' = Z(\Psi_y) \in \mathscr{Z}$ *with* $Z \cap Z' \neq \emptyset$ *then:*
    (a) $Z \subset \Psi_y([-Q, Q]^2)$;
    (b) *if* $x \in Z \cap Z'$ *then* $W^{s/u}(x, Z) \subset V^{s/u}(x, Z')$;
    (c) *if* $x \in Z$, $y \in Z'$ *then* $V^s(x, Z)$ *and* $V^u(y, Z')$ *intersect at a unique point.*

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

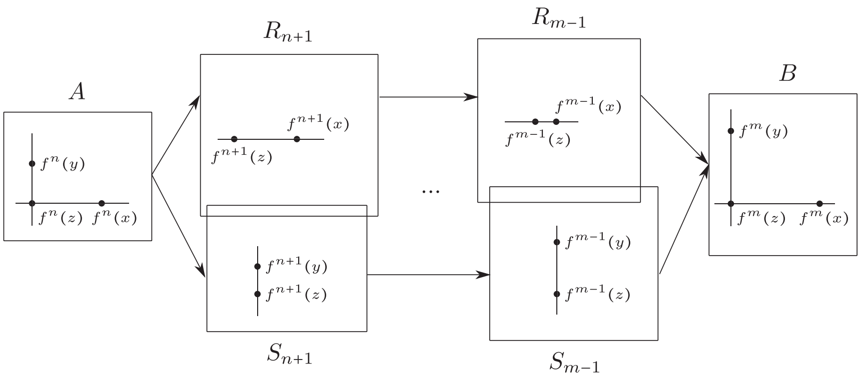

FIGURE 16. The diamond argument.

Statement (1) also follows from the Markov structure of Σ, while (2) follows from the fine control we have on the Lyapunov charts inside each rectangle $[-Q, Q]^2$. Lemma 3.5 allows us to consider Smale brackets of different intersecting rectangles.

Step 3 (Bowen–Sinaĭ refinement). We repeat verbatim Step 3 from §3.1.4. The resulting partition $\mathscr{R}$ is a Markov partition. By §3.1.2, we obtain a symbolic model $(\widehat{\Sigma}, \widehat{\sigma}, \widehat{\pi})$, where $(\widehat{\Sigma}, \widehat{\sigma})$ is the TMS defined by the graph $\widehat{\mathscr{G}} = (\widehat{V}, \widehat{E})$ with vertex set $\widehat{V} = \mathscr{R}$ and edge set $\widehat{E} = \{R \to S : f(R^*) \cap S^* \neq \emptyset\}$. Let $\widehat{\pi} : \Sigma \to \Omega(f)$ be defined for $\underline{R} = \{R_n\}_{n\in\mathbb{Z}} \in \Sigma$ by

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

Then $\widehat{\pi}$ is a finite-to-one surjection that is one-to-one on the set $\{x \in \Omega(f) : f^n(x) \in \bigcup_{R\in\mathscr{R}} R^*, \forall n \in \mathbb{Z}\}$.

We remark that the method of pseudo-orbits also works for uniformly expanding maps, in which case the TMS is one-sided. To do this, assume for simplicity that $f : M \to M$ satisfies $d(f(x), f(y)) \geq \kappa^{-1} d(x, y)$. Define Lyapunov charts simply by $\Psi_x := \exp_x$, then define an edge $\Psi_x \overset{\varepsilon}{\to} \Psi_y$ if and only if $d(f(x), y) \ll 1$, and prove that each pseudo-orbit $\{\Psi_{x_n}\}_{n\geq 0}$ shadows a single point $x \in M$. Now implement Steps 1–3.

3.1.6.4. *Bowen relation.* We explain why $\widehat{\pi}$ is finite-to-one. The proof follows Bowen [**Bow78b**], and is sometimes referred as the *diamond argument*, as justified by Figure 16; see also [**Adl98**, Lemma 6.7]. Bowen's idea was to investigate the quotient map $\pi : \Sigma \to M$. For us, this argument will be extremely useful to prove, in §3.2, that the coding obtained for non-uniformly hyperbolic maps is finite-to-one. For uniformly hyperbolic maps, this argument is not essential, but we stress that the method is of interest in its own right and introduces a relation, nowadays called the *Bowen relation*, that precisely characterizes the loss of injectivity of π.

Consider the triple $(\widehat{\Sigma}, \widehat{\sigma}, \widehat{\pi})$ as above. Define a relation in $\mathscr{R}$ by $R \sim S$ if and only if $R \cap S \neq \emptyset$. Assume that $R \sim S$. If $x \in R$ and $y \in S$, let $[x, y]$ be their Smale bracket, which is well defined by part (2)(c) of Lemma 3.5. Let $\underline{R} = \{R_n\}_{n\in\mathbb{Z}}, \underline{S} = \{S_n\}_{n\in\mathbb{Z}} \in \Sigma$.

Bowen relation. We say that $\underline{R} \approx \underline{S}$ if $R_n \sim S_n$ for all $n \in \mathbb{Z}$.

This clearly defines an equivalence relation on Σ, with $\widehat{\pi}(\underline{R}) = \widehat{\pi}(\underline{S})$ if and only if $\underline{R} \approx \underline{S}$. Let $N = \#\mathscr{R}$.

LEMMA 3.6. *The following statements hold for all $\varepsilon > 0$ small enough.*

(1) *If $R_n \to \cdots \to R_m$ and $S_n \to \cdots \to S_m$ are paths on $\widehat{\mathscr{G}}$ such that $R_n = S_n$, $R_m = S_m$ and $R_k \sim S_k$ for $k = n, \ldots, m$, then $R_k = S_k$ for $k = n, \ldots, m$.*

(2) *$\widehat{\pi}$ is everywhere at most N^2-to-one, that is, for every x we have $\#\widehat{\pi}^{-1}(x) \le N^2$.*

Proof. The original reference is [**Bow78b**, pp. 13–14]. As mentioned by Bowen himself, statement (2) was pointed out by Brian Marcus. Write $A = R_n = S_n$ and $B = R_m = S_m$. For part (1), choose x, y such that $f^k(x) \in R_k^*$ and $f^k(y) \in S_k^*$ for $k = n, \ldots, m$. Define z by the equality $f^n(z) = [f^n(x), f^n(y)]$. Since $R_k \sim S_k$, we have $f^k(z) = [f^k(x), f^k(y)]$ for $k = n, \ldots, m$. Noting that $f^n(x), f^n(y) \in A^*$ and $f^m(x), f^m(y) \in B^*$, we have that $f^n(z) \in A^*$ and $f^m(z) \in B^*$. Now we use the Markov property.

- The Markov property for the stable direction at $f^n(x) \in A$ implies that $f^k(z) \in W^s(f^k(x), R_k)$ for $k = n, \ldots, m$. Indeed, we can prove inductively that $f^k(z) \in (W^s(f^k(x), R_k))^*$, the interior of $W^s(f^k(x), R_k)$ in the relative topology of R_k. In particular, $f^k(z) \in R_k^*$ for $k = n, \ldots, m$.
- Applying the same argument for the unstable direction of $f^m(y) \in B$, we obtain that $f^k(z) \in S_k^*$ for $k = n, \ldots, m$.

Hence $f^k(z) \in R_k^* \cap S_k^*$ and so $R_k = S_k$, which proves (1). Now we prove (2). If some x has more than N^2 pre-images, then there are two of them, say $\underline{R}$ and $\underline{S}$, and two indices $n < m$ such that $(R_n, \ldots, R_m) \ne (S_n, \ldots, S_m)$ with $R_n = S_n$ and $R_m = S_m$. Since $f^k(x) \in R_k \cap S_k$, we have $R_k \sim S_k$ for $k = n, \ldots, m$. This contradicts part (1). □

3.2. *Symbolic dynamics for non-uniformly hyperbolic systems.* We finally arrive at the core of the discussion, showing how to employ the method of pseudo-orbits for non-uniformly hyperbolic systems. As mentioned in the introduction, Katok was the first to use the theory of pseudo-orbits for hyperbolic measures [**Kat80**], and constructed what are now called *Katok horseshoes*. Restricted to $C^{1+\beta}$ surface diffeomorphisms, his construction provides finite Markov partitions that approximate the topological entropy. For that, he used Pesin theory on subsets where the parameters vary continuously. This approach is not genuinely non-uniformly hyperbolic, because it discards regions with bad behavior of such parameters. In this section we explain how this difficulty was solved by Sarig [**Sar13**]. The starting step is to control the hyperbolicity parameters more effectively, as explained in §2.2. We will now use it to present a finer theory of pseudo-orbits, that in particular provides symbolic models for non-uniformly hyperbolic systems. The idea for construction of the symbolic model is similar to Bowen's method described in §3.1.6, but instead of Lyapunov charts we use (double) Pesin charts as vertices of the TMS. In order to code all points with some non-uniform hyperbolicity, we will invariably need countably many such charts. Hence, while for uniformly hyperbolic systems the TMS has finitely many states, now it will have countably many.

In the following sections we will restrict ourselves to $C^{1+\beta}$ surface diffeomorphisms. Later we explain how to perform the construction in other settings, which include higher-dimensional diffeomorphisms, flows, and billiard maps. We will emphasize five main ingredients in the proof:

- ε-overlap;
- ε-double charts;
- coarse graining;
- improvement lemma;
- inverse theorem.

The first two are discussed in the next subsection, and the others in the subsequent subsections.

3.2.1. *Preliminaries.* Let M be a closed connected smooth Riemannian surface, let $f : M \to M$ be a $C^{1+\beta}$ diffeomorphism, and fix $\chi > 0$. Recall the definitions of the non-uniformly hyperbolic locus NUH_χ and Pesin charts Ψ_x from §2.2.2. One important part of the construction will be to restrict the domains of Pesin charts, tuning them properly. Since we want to end up with countably many of them, we choose their sizes from a countable set. Fix $\varepsilon > 0$ small enough, and let $I_\varepsilon := \{e^{-\varepsilon n/3} : n \geq 0\}$. We redefine $Q(x)$ as follows.

Parameter $Q(x)$. For each $x \in \mathrm{NUH}_\chi$, define $Q(x)$ to be the largest element of I_ε that is less than or equal to $\varepsilon^{3/\beta}\|C(f(x))^{-1}\|_{\mathrm{Frob}}^{-12/\beta}$.

In other words, we truncate $Q(x)$ to I_ε. Now define the parameters q, q^s, q^u and the non-uniformly hyperbolic locus NUH_χ^* as in §§2.2.5 and 2.2.6. Observe that $q, q^s, q^u \in I_\varepsilon$. To obtain a finite-to-one coding, we need a recurrence assumption on the parameter q, so we define a subset of NUH_χ^* as follows.

The non-uniformly hyperbolic locus $\mathrm{NUH}_\chi^\#$.

$$\mathrm{NUH}_\chi^\# = \Big\{x \in \mathrm{NUH}_\chi^* : \limsup_{n\to+\infty} q(f^n(x)) > 0 \text{ and } \limsup_{n\to-\infty} q(f^n(x)) > 0\Big\}.$$

It is important to notice that this recurrence assumption is harmless for measures, since an analogue of Lemma 2.7 holds: if μ is an f-invariant probability measure supported on NUH_χ^*, then it is supported on $\mathrm{NUH}_\chi^\#$. This follows from the Poincaré recurrence theorem (we leave the details to the reader). We now state the main result we wish to discuss.

THEOREM 3.7. *Let $f : M \to M$ be as above. For every $\chi > 0$, there exist a TMS (Σ, σ) and a Hölder continuous map $\pi : \Sigma \to M$ such that:*

(1) *$\pi \circ \sigma = f \circ \pi$;*
(2) *$\pi[\Sigma^\#] = \mathrm{NUH}_\chi^\#$;*
(3) *the restriction $\pi\restriction_{\Sigma^\#}: \Sigma^\# \to \mathrm{NUH}_\chi^\#$ is finite-to-one.*

Remember the definition of the recurrent set $\Sigma^\#$ at the beginning of §3. Theorem 3.7 does not rely on any measure, and instead provides a single symbolic model that codes

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

all χ-hyperbolic measures at the same time. Theorem 3.7 is a strengthening of Theorem 1.3 established by Sarig in [**Sar13**]. The main difference between Theorem 3.7 and [**Sar13**, Theorem 1.3] is that Sarig's construction relies on Lyapunov regularity, and as a consequence he only obtains an inclusion of the form $\pi[\Sigma^\#] \supset \mathrm{NUH}^\#_\chi$. But rehearsing the same arguments of his proof inside the non-uniformly hyperbolic loci we have defined here provides the above statement. This observation grew from ongoing work with Buzzi and Crovisier [**BCL**], in which we need a more intrinsic construction to make it work for three-dimensional flows. The proof of Theorem 3.7 makes essential use of the low dimension of M: since the bundles E^s, E^u are one-dimensional, we are able to apply arguments of bounded distortion. If M has dimension larger than two, then E^s, E^u can both have dimension larger than one, and there is no a priori reason for them to satisfy bounded distortion estimates. Nevertheless, building on his previous work [**BO18**], Ben Ovadia was able to obtain a result similar to Theorem 3.7 that works in any dimension [**BO19**]; for more see §3.2.7.

3.2.1.1. *ε-overlap of Pesin charts.* We start by establishing an analogue of Theorem 3.1. In the uniformly hyperbolic situation, the map $x \mapsto \Psi_x$ is continuous, hence the norm of $\Psi_y^{-1} \circ \Psi_x$ can be controlled by $d(x, y)$. For non-uniformly hyperbolic systems, the maps $x \in \mathrm{NUH}_\chi \mapsto e^s_x, e^u_x$ are not necessarily continuous, so even though $d(x, y) \ll 1$ we can still have $\Psi_y^{-1} \circ \Psi_x$ with large norm, if the behavior of $C(x)$ and $C(y)$ are very different. Therefore, we only allow overlaps when, in addition to taking nearby points, their matrices C are close. For the definition, we allow Pesin charts to have different domains: for each $\eta \in I_\varepsilon$, define $\Psi^\eta_x := \Psi_x \restriction_{[-\eta,\eta]^2}$.

ε-overlap. Two Pesin charts $\Psi^{\eta_1}_{x_1}, \Psi^{\eta_2}_{x_2}$ are said to *ε-overlap* if $\eta_1/\eta_2 = e^{\pm\varepsilon}$ and $d(x_1, x_2) + \|\widetilde{C(x_1)} - \widetilde{C(x_2)}\| < (\eta_1\eta_2)^4$. When this happens, we write $\Psi^{\eta_1}_{x_1} \overset{\varepsilon}{\approx} \Psi^{\eta_2}_{x_2}$.

This notion was introduced in [**Sar13**]. It constitutes the first main ingredient in the proof of Theorem 3.7. The definition is strong enough to guarantee that the hyperbolicity parameters of x_1 and x_2 are almost the same; see [**Sar13**, Lemma 3.3] or [**LM18**, Proposition 3.4]. We are now able to recover Theorem 3.1.

THEOREM 3.8. *For all $\varepsilon > 0$ small enough, if $\Psi^\eta_{f(x)} \overset{\varepsilon}{\approx} \Psi^{\eta'}_y$, then $f_{x,y}$ is well defined on $[-10Q(x), 10Q(x)]^2$ and can be written as $f_{x,y}(v_1, v_2) = (Av_1 + h_1(v_1, v_2), Bv_2 + h_2(v_1, v_2))$ where:*

(1) $|A|, |B^{-1}| < e^{-\chi}$ *(cf. Theorem 2.6);*

(2) $\|h_i\|_{1+\beta/3} < \varepsilon$ *for $i = 1, 2$.*

If $\Psi^{\eta'}_{f^{-1}(y)} \overset{\varepsilon}{\approx} \Psi^\eta_x$ then a similar statement holds for $f^{-1}_{x,y}$.

This is [**Sar13**, Proposition 3.4].

3.2.1.2. *ε-double charts.* Now that we have a good notion of overlap between Pesin charts, we want to define graph transforms. The approach will be similar to §2.2.7, where we defined stable and unstable graph transforms using different scales q^s and q^u. In that

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

subsection, we also gave a dynamical explanation of the recursive equations in Lemma 2.8(2) that q^s, q^u satisfy. To extend this to Pesin charts, we consider two scales for each chart, one that controls the stable direction and another that controls the unstable direction. Hence we *do not* work with Pesin charts alone, but instead consider different objects, called *double charts*, and use them to define stable and unstable graph transforms. This idea, also introduced in [**Sar13**], is the second main ingredient in the proof of Theorem 3.7.

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 $0 < p^s, p^u \leq Q(x)$.

Intuitively, just as $q^{s/u}(x)$ are choices for the sizes of local stable/unstable manifolds at x, the parameters p^s/p^u represent candidates for the sizes of local stable/unstable manifolds of pseudo-orbits. To make sense of this, let us first define transitions between ε-double charts.

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) $p^s = \min\{e^\varepsilon q^s, Q(x)\}$ and $q^u = \min\{e^\varepsilon p^u, Q(y)\}$.

Condition (GPO1) allows us to represent f nearby x by Pesin charts at x and y, and similarly for f^{-1}. Condition (GPO2) is a greedy algorithm that chooses the local hyperbolicity parameters as large as possible, and is the counterpart of Lemma 2.8(2) for pseudo-orbits.

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 such that $\Psi_{x_n}^{p_n^s,p_n^u} \overset{\varepsilon}{\to} \Psi_{x_{n+1}}^{p_{n+1}^s,p_{n+1}^u}$ for all $n \in \mathbb{Z}$.

This definition is much stronger than the one given in §3.1.6. Observe that if $\underline{v}$ is an ε-gpo then by (GPO2) we have that

$$p_0^s = \inf\{e^{\varepsilon n} Q(x_n) : n \geq 0\} \quad \text{and} \quad p_0^u = \inf\{e^{\varepsilon n} Q(x_{-n}) : n \geq 0\}.$$

These equations are very similar to the definitions of $q^{s/u}$; see §2.2.6.

3.2.1.3. *Graph transforms.* To finally define graph transforms, it remains to strengthen the notion of admissibility. Let $v = \Psi_x^{p^s,p^u}$ be an ε-double chart.

Admissible manifolds. An *s-admissible manifold* at v is a set of the form $V^s = \Psi_x\{(t, F(t)) : |t| \leq p^s\}$, where $F : [-p^s, p^s] \to \mathbb{R}$ is a $C^{1+\beta/3}$ function such that:

(AM1) $|F(0)| \leq 10^{-3}(p^s \wedge p^u)$;

(AM2) $|F'(0)| \leq \frac{1}{2}(p^s \wedge p^u)^{\beta/3}$;

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

Similarly, a *u-admissible manifold* at v is a set of the form $V^u = \Psi_x\{(G(t), t) : |t| \leq p^u\}$ 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 $[-p^u, p^u]$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Note that $p^{s/u}$ control the domains of the representing functions, and $p^s \wedge p^u$ controls their behavior at 0. Let $\mathscr{M}_v^s$, $\mathscr{M}_v^u$ be the space of all s, u-admissible manifolds at v, which are metric spaces with the same metrics as before. For each edge $v \overset{\varepsilon}{\to} w$, define the stable graph transform $\mathscr{F}_{v,w}^s : \mathscr{M}_w^s \to \mathscr{M}_v^s$ and the unstable graph transform $\mathscr{F}_{v,w}^u : \mathscr{M}_v^u \to \mathscr{M}_w^u$ as before. An analogue of Theorem 3.2 holds, and we can similarly define stable and unstable manifolds for every ε-gpo $\underline{v}$.

Stable/unstable manifolds. The *stable manifold* of an ε-gpo $\underline{v} = \{v_n\}_{n\in\mathbb{Z}}$ is the unique s-admissible manifold $V^s[\underline{v}] \in \mathscr{M}_{v_0}^s$ defined by

$$V^s[\underline{v}] := \lim_{n\to\infty} (\mathscr{F}_{v_0,v_1}^s \circ \cdots \circ \mathscr{F}_{v_{n-1},v_n}^s)[V_n]$$

for some (any) sequence $\{V_n\}_{n\geq 0}$ with $V_n \in \mathscr{M}_{v_n}^s$. The *unstable manifold* of $\underline{v}$ is the unique u-admissible manifold $V^u[\underline{v}] \in \mathscr{M}_{v_0}^u$ defined by

$$V^u[\underline{v}] := \lim_{n\to-\infty} (\mathscr{F}_{v_{-1},v_0}^u \circ \cdots \circ \mathscr{F}_{v_n,v_{n+1}}^u)[V_n]$$

for some (any) sequence $\{V_n\}_{n\leq 0}$ with $V_n \in \mathscr{M}_{v_n}^u$.

These manifolds are genuine Pesin invariant manifolds; see [**Sar13**, Proposition 6.3]. In particular, if $y, z \in V^s[\underline{v}]$ then $s(y)/s(z) = e^{\pm\text{const}}$, and similarly for $V^u[\underline{v}]$. We point out that if F is the representing function of $V^s[\underline{v}]$ and F_n is the representing function of $(\mathscr{F}_{v_0,v_1}^s \circ \cdots \circ \mathscr{F}_{v_{n-1},v_n}^s)[V_n]$ for $n \geq 0$, then $\|F - F_n\|_{C^1} \to 0$ as $n \to \infty$, and the same holds for the representing function of $V^u[\underline{v}]$. This follows from the Arzelà–Ascoli theorem, since the $C^{1+\beta/3}$ norm of representing functions is uniformly bounded; see, for example, part (2) in the proof of [**Sar13**, Proposition 4.15]. We also define 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 $n \in \mathbb{Z}$.

The shadowing lemma is still valid, again with $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$; see [**Sar13**, Theorem 4.16(1)].

3.2.2. *Coarse graining.* The third main ingredient in the proof of Theorem 3.7 involves choosing countably many charts that shadow all orbits of interest. For uniformly hyperbolic systems, a sufficiently dense set of points is enough. For non-uniformly hyperbolic systems, the construction is more elaborate. Firstly, the definition of ε-overlap depends on $\|\widetilde{C(x)} - \widetilde{C(y)}\|$, which depends on a comparison between $s(x), u(x), \alpha(x)$ and $s(y), u(y), \alpha(y)$. Secondly, nearest neighbor conditions do not follow from control at x and y. Fortunately, all parameters involved in the construction belong to a precompact set, so there are countable dense subsets.

THEOREM 3.9. *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 an ε-gpo $\underline{v} \in \mathscr{A}^{\mathbb{Z}}$ 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 $\mathrm{NUH}_\chi^\#$.*

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

The original statement is [Sar13, Theorem 4.16]. The discreteness property (1) says that Pesin blocks require only finitely many ε-double charts, while the relevance property (3) guarantees that none of the ε-double charts is redundant.

Sketch of proof. Define $X := M^3 \times \mathrm{GL}(2, \mathbb{R})^3 \times (0, 1]$. For each $x \in \mathrm{NUH}^*_\chi$, let $\Gamma(x) = (\underline{x}, \underline{C}, \underline{Q}) \in X$ with

$$\underline{x} = (f^{-1}(x), x, f(x)), \quad \underline{C} = (C(f^{-1}(x)), C(x), C(f(x))), \quad \underline{Q} = Q(x).$$

Let $Y = \{\Gamma(x) : x \in \mathrm{NUH}^*_\chi\}$. For each triple $\underline{\ell} = (\ell_{-1}, \ell_0, \ell_1) \in \mathbb{N}_0^3$, define

$$Y_{\underline{\ell}} := \{\Gamma(x) \in Y : e^{\ell_i} \le \|C(f^i(x))^{-1}\| < e^{\ell_i+1}, -1 \le i \le 1\}.$$

Then $Y = \bigcup_{\underline{\ell}\in\mathbb{N}_0^3} Y_{\underline{\ell}}$, and each $Y_{\underline{\ell}}$ is precompact in X by definition. For each $j \ge 0$, choose a finite set $Y_{\underline{\ell}}(j) \subset Y_{\underline{\ell}}$ such that for every $\Gamma(x) \in Y_{\underline{\ell}}$ there exists $\Gamma(y) \in Y_{\underline{\ell}}(j)$ such that:

(a) $d(f^i(x), f^i(y)) + \|\widetilde{C(f^i(x))} - \widetilde{C(f^i(y))}\| < e^{-8(j+2)}$ for $-1 \le i \le 1$;
(b) $Q(x)/Q(y) = e^{\pm\varepsilon/3}$.

We define the countable set of ε-double charts as follows.

The alphabet $\mathscr{A}$. Let $\mathscr{A}$ be the countable family of $\Psi_x^{p^s,p^u}$ such that:

(CG1) $\Gamma(x) \in Y_{\underline{\ell}}(j)$ for some $(\underline{\ell}, j) \in \mathbb{N}_0^3 \times \mathbb{N}_0$;
(CG2) $0 < p^s, p^u \le Q(x)$ and $p^s, p^u \in I_\varepsilon$;
(CG3) $e^{-j-2} \le p^s \wedge p^u \le e^{-j+2}$.

This alphabet satisfies (1) and (2) but not necessarily (3). We can easily reduce it to a sub-alphabet $\mathscr{A}'$ satisfying (1)–(3) as follows. Call $v \in \mathscr{A}$ relevant if there is $\underline{v} \in \mathscr{A}^{\mathbb{Z}}$ with $v_0 = v$ such that $\underline{v}$ shadows a point in $\mathrm{NUH}^\#_\chi$. Since $\mathrm{NUH}^\#_\chi$ is f-invariant, every v_i is relevant. Then $\mathscr{A}' = \{v \in \mathscr{A} : v \text{ is relevant}\}$ is discrete because $\mathscr{A}' \subset \mathscr{A}$, and it is sufficient and relevant by definition. □

Referring to the steps of §3.1.6.3, we have just completed Step 1. Now let $\mathscr{G} = (V, E)$ be the oriented graph with vertex set $V = \mathscr{A}$ and edge set $E = \{v \overset{\varepsilon}{\to} w\}$, and let (Σ, σ) be the TMS generated by $\mathscr{G}$. The proof of sufficiency actually gives more: if $x \in \mathrm{NUH}^\#_\chi$ then there is a *recurrent* ε-gpo $\underline{v} \in \Sigma^\#$ that shadows x. By the shadowing lemma, $\pi : \Sigma \to M$ defined by $\{\pi(\underline{v})\} := V^s[\underline{v}] \cap V^u[\underline{v}]$ is an infinite-to-one extension of f such that $\pi[\Sigma^\#] \supset \mathrm{NUH}^\#_\chi$.

3.2.3. *Improvement lemma.* The fourth main ingredient in the proof of Theorem 3.7 is an important fact that will imply the reverse inclusion $\pi[\Sigma^\#] \subset \mathrm{NUH}^\#_\chi$ as well as the inverse theorem in the next subsection. We start by observing that points of $\pi[\Sigma^\#]$ do have stable and unstable directions. If $\underline{v}$ is an ε-gpo, then for $x \in V^u[\underline{v}]$ we can take e^u_x to be the tangent vector (any tangent vector) to $V^u[\underline{v}]$ at x. This direction indeed contracts in the past and expands in the future; see [Sar13, Proposition 6.3(2)] and claim (i) in the proof of [Sar13, Proposition 6.5]. Similarly, for $x \in V^s[\underline{v}]$ we can take the tangent vector (any tangent vector) to $V^s[\underline{v}]$ at x to be the stable direction e^s_x.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

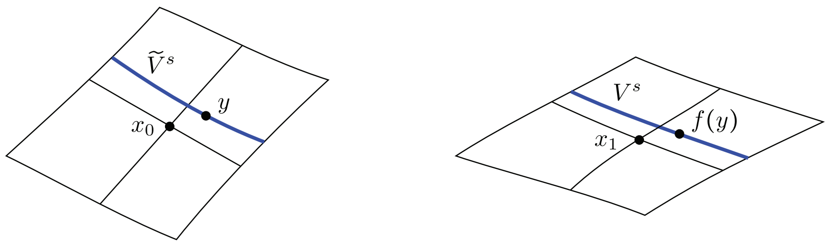

FIGURE 17. Improvement lemma: the graph transform improves ratios.

Fix $\underline{v} \in \Sigma^\#$, and let $x = \pi(\underline{v})$. Since $\{x\} = V^s[\underline{v}] \cap V^u[\underline{v}]$, e^s_x, e^u_x are defined. If we prove that $s(x), u(x) < \infty$, then (NUH3) holds and automatically (NUH1)–(NUH2) hold as well, implying that $x \in \mathrm{NUH}_\chi$. The proof that $s(x), u(x) < \infty$ is very delicate, since a priori there is no reason for x to have hyperbolic behavior as good as the behavior of the centers of the ε-double charts of $\underline{v}$. But the notions of ε-overlap and admissibility are so strong that indeed $s(x), u(x) < \infty$. The proof of this fact relies on a general philosophy that f improves smoothness along the unstable direction, and f^{-1} improves smoothness along the stable direction. In terms of graph transforms, the ratios of s, u parameters improve. We call this an *improvement lemma*. The heuristics for such improvement can be explained as follows. Assume that $v \overset{\varepsilon}{\to} w$, where $v = \Psi^{p^s_0, p^u_0}_{x_0}$ and $w = \Psi^{p^s_1, p^u_1}_{x_1}$, let $V^s \in \mathscr{M}^s_w$ and $\widetilde{V}^s = \mathscr{F}^s_{v,w}[V^s]$, and fix a point $y \in \widetilde{V}^s$. In particular $f(y) \in V^s$; see Figure 17. Assuming that $s(y) < \infty$, we want to compare the ratios $s(f(y))/s(x_1)$ and $s(y)/s(x_0)$. Proceeding as in the proof of Lemma 2.1, we have $s(y)^2 = 2 + Cs(f(y))^2$, where $C = \|df e^s_y\|^2 e^{2\chi}$. By the ε-overlap, we also have $s(x_0)^2 \approx 2 + Cs(x_1)^2$, and so $s(y)^2/s(x_0)^2 \approx (2 + Cs(f(y))^2)/(2 + Cs(x_1)^2)$. If the initial ratio is $s(f(y))^2/s(x_1)^2 = K \gg 1$, then the new ratio is $s(y)^2/s(x_0)^2 \approx (2 + KCs(x_1)^2)/(2 + Cs(x_1)^2) < K$. The same occurs if $K \ll 1$, in which case the new ratio becomes greater than K.

LEMMA 3.10. (Improvement lemma) *For $\varepsilon > 0$ small enough, and for $\xi \geq \sqrt{\varepsilon}$, if $s(f(y))/s(x_1) = e^{\pm\xi}$, then $s(y)/s(x_0) = e^{\pm(\xi - Q(x_0)^{\beta/4})}$.*

Hence the ratio improves whenever it is outside $[e^{-\sqrt{\varepsilon}}, e^{\sqrt{\varepsilon}}]$; see [Sar13, Lemma 7.2] for a proof. Here is the first main consequence of this important lemma.

COROLLARY 3.11. $\pi[\Sigma^\#] \subset \mathrm{NUH}_\chi$, *that is, if $x = \pi(\underline{v})$ for $\underline{v} \in \Sigma^\#$, then $s(x)$, $u(x) < +\infty$.*

The proof can be found in [Sar13, §7.1]. We summarize it as follows. Let $v_{n_k} = v$ for infinitely many $n_k > 0$. Since $v = \Psi^{p^s, p^u}_z$ is relevant, there is $V^s \in \mathscr{M}^s_v$ such that $s(y) < \infty$ for every $y \in V^s$. Starting with $V_{n_k} = V^s$, pull it back through the graph transforms to get $\widetilde{V}^s_k \in \mathscr{M}^s_{v_0}$. If k is large enough, then the original ratio $s(y)/s(z)$ passes through sufficiently many improvements $Q(z)$ so that $s(w_k)/s(x_0) = e^{\pm\xi}$ for all $w_k \in \widetilde{V}^s_k$, for some fixed $\xi \geq \sqrt{\varepsilon}$. Since the representing functions of $\widetilde{V}^s_k$ converge to the representing function of $V^s[\underline{v}]$ in the C^1 norm, we conclude that $s(x) < \infty$. Similarly, we obtain that $u(x) < \infty$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

3.2.4. *Inverse theorem.* The fifth and final main ingredient in the proof of Theorem 3.7 is the inverse theorem. To understand its importance, recall that once we have constructed an infinite-to-one coding π, the next step is to apply a Bowen–Sinaĭ refinement to the Markov cover $\mathscr{Z}$ induced by π. Since Σ has countably many states, the Markov cover is countable, so that after refining the resulting partition could be uncountable (for example, the refinement of the dyadic intervals in [0, 1] is the point partition). One condition that guarantees a countable refinement is *local finiteness*: $\mathscr{Z}$ is locally finite if every $Z \in \mathscr{Z}$ only intersects finitely many other $Z' \in \mathscr{Z}$. The understanding of intersections $Z \cap Z'$ comes from an *inverse problem*: if $\pi(\underline{v}) = x$, how is $\underline{v}$ defined in terms of x? The next theorem (essentially) answers this question.

THEOREM 3.12. (Inverse theorem) *Let $\underline{v} = \{\Psi_{x_n}^{p_n^s,p_n^u}\}_{n\in\mathbb{Z}} \in \Sigma^\#$, and let $\pi(\underline{v}) = x$. Then the following properties hold for all $n \in \mathbb{Z}$.*

(1) *Control of x:* $\mathrm{dist}(x_n, f^n(x)) < \mathrm{const}$.

(2) *Control of α:* $\sin\alpha(x_n)/\sin\alpha(f^n(x)) = e^{\pm\mathrm{const}}$.

(3) *Control of s, u:* $s(x_n)/s(f^n(x)) = u(x_n)/u(f^n(x)) = e^{\pm\mathrm{const}}$.

(4) *Control of p^s, p^u:* $p_n^s/q^s(f^n(x)) = p_n^u/q^u(f^n(x)) = e^{\pm\mathrm{const}}$.

In particular, $x \in \mathrm{NUH}_\chi^\#$, and so $\pi[\Sigma^\#] = \mathrm{NUH}_\chi^\#$.

The above theorem is not the original reference [**Sar13**, §6], since it did not make use of q, q^s, q^u. Instead, the original statement considered two ε-gpos $\underline{v}, \underline{w} \in \Sigma^\#$ and compared their parameters directly. The constant appearing in the theorem is of the order of $\sqrt[3]{\varepsilon}$. The assumption $\underline{v} \in \Sigma^\#$ is essential to guarantee parts (3) and (4), since the proof uses that the trajectories visit a Pesin block infinitely often. Theorem 3.12 states that each coordinate of $\underline{v}$ is uniquely defined 'up to bounded error'. Below we explain how to get the estimates for $n = 0$.

Control of x and α. Let F, G be the representing functions of V^s, V^u. Since $F(0), G(0) \approx 0$, the graphs of F, G intersect close to the origin $(0, 0)$. Applying Ψ_{x_0}, we conclude that $d(x_0, x) \ll 1$. To control the angle, we use that $\|F'\|_{C^0}, \|G'\|_{C^0} \ll 1$ and so the graphs of F, G intersect almost perpendicularly. Applying Ψ_{x_0}, we get that $\alpha(x_0) \approx \alpha(x)$; see Figure 18.

Control of s, u. The proof is identical to the proof of Corollary 3.11.

Observe that parts (2)–(3) imply that $Q(x_n)/Q(f^n(x)) = e^{\pm\mathrm{const}}$ for all $n \in \mathbb{Z}$.

Control of p^s, p^u. We prove the first estimate $p_0^s/q^s(x) = e^{\pm\mathrm{const}}$ (the second is analogous). The idea is to observe that p_0^s and $q^s(x)$ are defined as infima of comparable sequences. Indeed, by definition we have

$$q^s(x) = \inf\{e^{\varepsilon n} Q(f^n(x)) : n \geq 0\},$$

and by (GPO2) we have

$$p_0^s = \inf\{e^{\varepsilon n} Q(x_n) : n \geq 0\}.$$

Since $Q(x_n)/Q(f^n(x)) = e^{\pm\mathrm{const}}$ for all $n \geq 0$, it follows that $p_0^s/q^s(x) = e^{\pm\mathrm{const}}$.

The above argument differs from the original one, and does not require any maximality assumption on q^s. Indeed, it comes for free from the recurrence of $\underline{v}$.

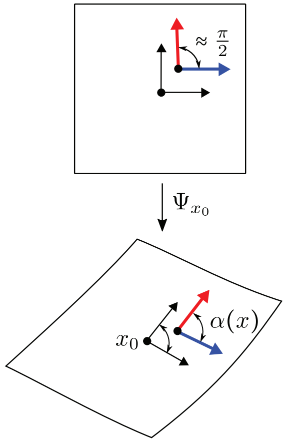

FIGURE 18. Control of x and α.

3.2.5. *Bowen–Sinaĭ refinement.* Local finiteness allows us to refine the Markov cover of $\mathrm{NUH}_\chi^\#$ and obtain a countable Markov partition. As mentioned in §3.1.1, the notion of Markov cover/partition for non-uniformly hyperbolic systems will be weaker than the one for uniformly hyperbolic systems, since we are not able to control the topology of the rectangles. Indeed, due to Theorem 3.12, the Markov cover is the projection of the canonical partition of $\Sigma^\#$ (instead of Σ) into cylinders at the zeroth position, hence rectangles will not have the regularity property. This loss of topological control compels us to perform the refinement in a more abstract way, and then check that the resulting partition still generates a finite-to-one symbolic extension. The reader who is already comfortable with the uniformly hyperbolic situation will not see much difficulty in the adaptations.

The Markov cover $\mathscr{Z}$. Let $\mathscr{Z} := \{Z(v) : v \in \mathscr{A}\}$, where

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

Restricted to $\Sigma^\#$, cylinders are neither closed nor empty, so the same occurs for $Z(v)$. Nevertheless, *s*/*u*-fibers are still well defined in $\mathscr{Z}$.

PROPOSITION 3.13. *The following statements are true.*

(1) *Covering property. $\mathscr{Z}$ is a cover of* $\mathrm{NUH}_\chi^\#$.
(2) *Product structure. For every $Z \in \mathscr{Z}$ and every $x, y \in Z$, the intersection $W^s(x, Z) \cap W^u(y, Z)$ consists of a single point, and this point belongs to Z.*
(3) *Symbolic Markov property. If $x = \pi(\underline{v})$ with $\underline{v} = \{v_n\}_{n\in\mathbb{Z}} \in \Sigma^\#$, then*

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

(4) *Local finiteness. For every $Z \in \mathscr{Z}$, the set $\{Z' \in \mathscr{Z} : Z \cap Z' \neq \emptyset\}$ is finite.*

Parts (1)–(3) are analogues of Proposition 3.4. The novelty is part (4), which follows from Theorem 3.9(1) and Theorem 3.12(3): if $v = \Psi_x^{p^s,p^u}$ and $w = \Psi_y^{q^s,q^u}$ satisfy

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

$Z(v) \cap Z(w) \neq \emptyset$ then $p^s/q^s = p^u/q^u = e^{\pm\text{const}}$, hence

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

Inside each $Z \in \mathscr{Z}$, define the Smale bracket $[\cdot, \cdot]_Z$ as before. Lemma 3.5 remains valid; see [**Sar13**, Lemmas 10.7, 10.8, 10.10]. We now refine $\mathscr{Z}$. For $Z, Z' \in \mathscr{Z}$ such that $Z \cap Z' \neq \emptyset$, let $\mathscr{E}_{ZZ'}$ = cover of Z by rectangles:

$$E^{su}_{Z,Z'} = \{x \in Z : W^s(x, Z) \cap Z' \neq \emptyset, W^u(x, Z) \cap Z' \neq \emptyset\};$$
$$E^{s\emptyset}_{Z,Z'} = \{x \in Z : W^s(x, Z) \cap Z' \neq \emptyset, W^u(x, Z) \cap Z' = \emptyset\};$$
$$E^{\emptyset u}_{Z,Z'} = \{x \in Z : W^s(x, Z) \cap Z' = \emptyset, W^u(x, Z) \cap Z' \neq \emptyset\};$$
$$E^{\emptyset\emptyset}_{Z,Z'} = \{x \in Z : W^s(x, Z) \cap Z' = \emptyset, W^u(x, Z) \cap Z' = \emptyset\}.$$

The above definition is simpler than the one for uniformly hyperbolic systems, since we do not take relative interiors nor closures. Let $\mathscr{R}$ be the partition that refines all of $\mathscr{E}_{ZZ'}$. Again due to Theorem 3.9(1) and Theorem 3.12(3), $\mathscr{R}$ and $\mathscr{Z}$ satisfy two additional local finiteness properties:

- for all $R \in \mathscr{R}$, the set $\{Z \in \mathscr{Z} : Z \supset R\}$ is finite;
- for all $Z \in \mathscr{Z}$, the set $\{R \in \mathscr{R} : R \subset Z\}$ is finite.

Inside each $R \in \mathscr{R}$, the Smale brackets $[\cdot, \cdot]_Z$ do not depend on Z, hence we can define $[\cdot, \cdot]$ on R.

LEMMA 3.14. *$\mathscr{R}$ 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*

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

Let $\widehat{\mathscr{G}} = (\widehat{V}, \widehat{E})$ be the graph with $\widehat{V} = \mathscr{R}$ and $\widehat{E} = \{R \to S : f(R) \cap S \neq \emptyset\}$ (compare this definition with the one given in Step 3 of §3.1.6.3). Let $(\widehat{\Sigma}, \widehat{\sigma})$ be the TMS defined by $\widehat{\mathscr{G}}$, and define $\widehat{\pi} : \Sigma \to M$ by

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

In comparison to the previous constructions, we take closures because the R_n are not necessarily closed. A priori, the image of $\widehat{\pi}$ could be much bigger than the image of π. Fortunately, this is not the case: for each $\underline{R} \in \widehat{\Sigma}^\#$, there is an ε-gpo $\underline{v} \in \Sigma^\#$ such that $\widehat{\pi}(\underline{R}) = \pi(\underline{v})$; see the proof of [**Sar13**, Theorem 12.5]. Therefore $\widehat{\pi}[\widehat{\Sigma}^\#] = \pi[\Sigma^\#] = \text{NUH}^\#_\chi$.

We point out that $\widehat{\pi}$ is compatible with the Smale brackets in Σ and $\mathscr{R}$. More specifically, for $\underline{R} = \{R_n\}_{n\in\mathbb{Z}}, \underline{S} = \{S_n\}_{n\in\mathbb{Z}} \in \Sigma$ with $R_0 = S_0$, let $\underline{U} = [\underline{R}, \underline{S}]$ where $\underline{U} = \{U_n\}_{n\in\mathbb{Z}}$ is defined by

$$U_n = \begin{cases} R_n & \text{if } n \geq 0, \\ S_n & \text{if } n \leq 0. \end{cases}$$

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

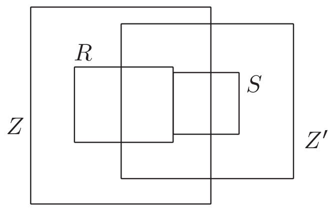

FIGURE 19. The affiliation property $R \sim S$, which might occur even when $R \cap S = \emptyset$.

Then $\widehat{\pi}([\underline{R}, \underline{S}]) = [\widehat{\pi}(\underline{R}), \widehat{\pi}(\underline{S})]$. This is [**BPVL20**, Lemma 4.4], and it is used to study the simplicity of generic fiber-bunched cocycles over non-uniformly hyperbolic diffeomorphisms.

3.2.6. *Affiliation and Bowen relation.* We investigate how $\widehat{\pi} : \widehat{\Sigma}^\# \to \mathrm{NUH}^\#_\chi$ loses injectivity. In the uniformly hyperbolic situation, we saw in §3.1.6.4 that this is characterized by the Bowen relation. Sarig was able to obtain a similar characterization [**Sar13**], which was further explored by Boyle and Buzzi [**BB17**].

Affiliation. Two rectangles $R, S \in \mathscr{R}$ are called *affiliated* if there exist $Z, Z' \in \mathscr{Z}$ such that $Z \supset R$, $Z' \supset S$ and $Z \cap Z' \neq \emptyset$. If this occurs, we write $R \sim S$. See Figure 19.

Affiliation is more complicated than mere non-empty intersection, and it arises from the need to take closures in the definition of $\widehat{\pi}$. If $R \sim S$ as above, then Lemma 3.5 implies that we can take Smale brackets between points of Z and Z'.

Bowen relation. We say that $\underline{R} \approx \underline{S}$ if and only if $R_n \sim S_n$ for all $n \in \mathbb{Z}$.

The following result was implicit in [**Sar13**], as explained in [**BB17**, §8.3].

LEMMA 3.15. *If $\underline{R}, \underline{S} \in \widehat{\Sigma}^\#$, then $\widehat{\pi}(\underline{R}) = \widehat{\pi}(\underline{S})$ if and only if $\underline{R} \approx \underline{S}$.*

Now we apply the diamond argument, as in §3.1.6.4, by choosing $n < m$ such that the rectangle configuration (R, Z, S, Z') of Figure 19 is the same at positions n and m. Introduce

$$N(R) := \#\{(S, Z') \in \mathscr{R} \times \mathscr{Z} : R \sim S \text{ and } Z' \supset S\}.$$

The local finiteness of $\mathscr{R}$ and $\mathscr{Z}$ imply that $N(R) < \infty$ for all $R \in \mathscr{R}$. For fixed $R, S \in \mathscr{R}$ and $n < m$, the pigeonhole principle implies that if there are more than $N(R)N(S)$ paths $R_n \to \cdots \to R_m$ such that $R_n \sim R$ and $R_m \sim S$, then two of them have the same rectangle configuration at iterates n and m, as expressed below.

THEOREM 3.16. *Let $\underline{R} \in \widehat{\Sigma}^\#$, with $R_n = R$ for infinitely many $n < 0$ and $R_m = S$ for infinitely many $m > 0$, and let $x = \widehat{\pi}(\underline{R})$. Then $\#[\widehat{\pi}^{-1}(x) \cap \widehat{\Sigma}^\#] \leq N(R)N(S)$. In particular, the restriction $\widehat{\pi} \restriction_{\widehat{\Sigma}^\#} : \widehat{\Sigma}^\# \to \mathrm{NUH}^\#_\chi$ is finite-to-one.*

The original proof [**Sar13**, Theorem 12.8] has a small error that was corrected in [**LS19**]. This concludes the proof of Theorem 3.7.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

3.2.7. *Higher dimensions.* Recently, building on his previous work [**BO18**], Ben Ovadia obtained a higher-dimensional version of Theorem 3.7 [**BO19**]. Regarding the five main ingredients, the first three can be adapted to higher dimensions, with many extra technical difficulties; see [**BO18**, §§2 and 3]. But the improvement lemma and inverse theorem require different approaches. Indeed, each $x = \pi(\underline{v})$ has many parameters $S(x, v)$ and $U(x, w)$, and in principle it is not clear with which reference parameters they should be compared. Let us explain how to make the estimates for $S(x, v)$. Let $V^s[\underline{v}] = \Psi_{x_0}\{(t, F(t)) : t \in [-p_0^s, p_0^s]^{d_s}\}$ be the stable manifold of $\underline{v} = \{\Psi_{x_0}^{p_n^s, p_n^u}\}_{n\in\mathbb{Z}} \in \Sigma^\#$. Let $x \in V^s[\underline{v}]$, with $x = \Psi_{x_0}(\overline{x}) = \Psi_{x_0}(\overline{t}, F(\overline{t}))$. The candidate for stable subspace at x is $\widetilde{E}_x^s = (d\Psi_{x_0})_{\overline{x}}[H_{\overline{x}}^s]$, where $H_{\overline{x}}^s$ is the subspace tangent to $\{(t, F(t)) : t \in [-p_0^s, p_0^s]^{d_s}\}$ at $\overline{x}$. Observing that the stable subspace at x_0 is $(d\Psi_{x_0})_0[\mathbb{R}^{d_s} \times \{0\}]$, consider a linear transformation $H_{\overline{x}}^s \to \mathbb{R}^{d_s} \times \{0\}$ given by the derivative of the projection $(t, F(t)) \mapsto (t, 0)$ at $\overline{x} = (\overline{t}, F(\overline{t}))$. Applying $d\Psi_{x_0}$, we obtain a linear transformation $\Xi_x : \widetilde{E}_x^s \to E_{x_0}^s$. Since the graph of F is almost horizontal, Ξ_x is almost an isometry. The improvement lemma can be stated as follows: if $v \in \widetilde{E}_{f(y)}^s$ satisfies $S(f(y), v)/S(x_1, \Xi_{f(y)}v) = \exp[\pm\xi]$ for some $\xi \geq \sqrt{\varepsilon}$, then

$$\frac{S(y, w)}{S(x_0, \Xi_y w)} = \exp[\pm(\xi - \tfrac{1}{6}Q(x_0)^{\beta/6})]$$

for $w = df_{f(y)}^{-1}v$; see [**BO18**, Lemma 4.6]. This improvement lemma implies two properties:

- $\pi[\Sigma^\#] \subset \mathrm{NUH}_\chi$ (see Claim 1 in the proof of [**BO18**, Lemma 4.7]);
- the inverse theorem (see [**BO18**, §4]).

3.2.8. *Uniform hyperbolicity versus non-uniform hyperbolicity.* In Table 1 we summarize the main differences between the constructions of symbolic models for uniformly hyperbolic and non-uniformly hyperbolic systems.

3.2.9. *Surface maps with discontinuities.* We now explain the constructions in the contexts of §§2.3 and 2.4. Let us start with surface maps with discontinuities and bounded derivative. As in the previous section, take $I_\varepsilon := \{e^{-\varepsilon n/3} : n \geq 0\}$ and truncate $Q(x)$ to I_ε, then consider ε-overlap and ε-gpo as in §3.2.1. To prove Theorem 3.9, we need to control more parameters: for $\underline{\ell} = (\ell_{-1}, \ell_0, \ell_1)$ and $\underline{k} = (k_{-1}, k_0, k_1)$, define

$$Y_{\underline{\ell},\underline{k}} := \left\{\Gamma(x) \in Y : \begin{array}{ll} e^{\ell_i} \leq \|C(f^i(x))^{-1}\| < e^{\ell_i+1}, & -1 \leq i \leq 1 \\ e^{-k_i-1} \leq d(f^i(x), \mathscr{D}) < e^{-k_i}, & -1 \leq i \leq 1 \end{array}\right\}.$$

Using these precompact sets, proceed as in the proof of Theorem 3.9. Theorem 3.12 works without modification, as well as Theorem 3.7.

As mentioned in §2.3, the prototypical examples for surface maps with discontinuities and bounded derivative are Poincaré return maps of three-dimensional flows with positive speed. Let N be a three-dimensional closed Riemannian manifold, let X be a $C^{1+\beta}$ vector field on N such that $X(p) \neq 0$ for all $p \in N$, and let $\varphi = \{\varphi^t\}_{t\in\mathbb{R}}$ be the flow generated by X. In §2.3 we constructed a global Poincaré section M, equal to the finite union of transverse discs, such that the return time function $\mathfrak{t} : M \to (0, \infty)$ is bounded away

TABLE 1. Comparison between the methods and objects used in the construction of symbolic models for uniformly hyperbolic and non-uniformly hyperbolic systems.

	Uniformly hyperbolic	Non-uniformly hyperbolic
Coding	All points	$\mathrm{NUH}_\chi^\#$
Chart	Lyapunov chart: uniform size	Pesin chart: size Q with $(1/n)\log Q(f^n(x)) \to 0$
Vertices	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)
Rep. function	$F \in C^1$ such that $F(0) \approx 0$ and $\|F'\|_{C^0} \approx 0$	$F \in C^{1+\beta/3}$ with (AM1)–(AM3)
$\pi : \Sigma \to M$	$\{\pi(\underline{v})\} = V^s[\underline{v}] \cap V^u[\underline{v}]$	Same
Cover $\mathscr{Z}$	$Z(v) = \{\pi(\underline{v}) : v_0 = v\}$ closed sets	$Z(v) = \{\pi(\underline{v}) : \underline{v} \in \Sigma^\#, v_0 = v\}$
Refinement	Relative interiors and closures	Set-theoretical refinement
Partition $\mathscr{R}$	Markov with regular rectangles	Markov without control of relative interiors
Graph $(\widehat{V}, \widehat{E})$	$\widehat{V} = \mathscr{R}$, $\widehat{E} = \{R \to S: f(R^*) \cap S^* \neq \emptyset\}$	$\widehat{V} = \mathscr{R}$, $\widehat{E} = \{R \to S: f(R) \cap S \neq \emptyset\}$
$\widehat{\pi} : \widehat{\Sigma} \to M$	$\{\widehat{\pi}(\underline{R})\} = \bigcap_{n\geq 0} f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n)$	$\{\widehat{\pi}(\underline{R})\} = \bigcap_{n\geq 0} \overline{f^n(R_{-n}) \cap \cdots \cap f^{-n}(R_n)}$
Finite-to-one	$\#\widehat{\pi}^{-1}(x) \leq (\#\mathscr{R})^2$, for all $x \in M$	$\#[\widehat{\pi}^{-1}(x) \cap \widehat{\Sigma}^\#] < \infty$, for all $x \in M$

from zero and infinity. Then the Poincaré return map $f : M \to M$ has discontinuities and bounded derivative.

We now relate the hyperbolic properties of φ and f. Let $\widetilde{\chi} > 0$. It is possible to define a non-uniformly hyperbolic locus $\mathrm{NUH}_{\widetilde{\chi}}(\varphi)$ for φ, similar to the definition on page 2609. This is part of an ongoing project with Buzzi and Crovisier [**BCL**]. Let NUH_χ be the non-uniformly hyperbolic locus of f. If $x \in M \cap \mathrm{NUH}_{\widetilde{\chi}}(\varphi)$, then x satisfies (NUH1)–(NUH3) for $\chi := \widetilde{\chi}\inf(\mathfrak{t})$. Indeed, the flow trajectory of x spends at least time $\inf(\mathfrak{t})$ between visits to M; see [**LS19**, Lemma 2.6]. To prove that $x \in \mathrm{NUH}_\chi$, it remains to check (NUH4). Here we encounter a problem: the section M could be chosen in such a bad way that every trajectory of f converges exponentially fast to $\mathscr{D}$, and so (NUH4) never holds. To bypass this difficulty, we need to choose M carefully so that most φ-trajectories define f-trajectories satisfying (NUH4). Unfortunately, we do not know how to construct such a section. What we know, and this is done in [**LS19**], is how to construct one section for each fixed measure (more generally, for a countable set of fixed measures). Fortunately, this is enough for many applications.

Let μ be a $\widetilde{\chi}$-hyperbolic probability measure for φ, and let ν be its projection to M, which is χ-hyperbolic for f. The goal is to choose M so that ν is f-adapted. Consider a one-parameter family of global Poincaré sections $\{M_r\}$, by changing the radii of each disc of M. More specifically, let $M_r = D_r(p_1) \cup \cdots \cup D_r(p_\ell)$, $r \in [a, b]$, such that each M_r is a global Poincaré section for φ and $\mathfrak{t}_r : M_r \to (0, \infty)$ is uniformly bounded away from zero and infinity. Let $f_r : M_r \to M_r$ be the Poincaré return map, and let ν_r be the projection of μ to M_r. The next result is [**LS19**, Theorem 2.8].

THEOREM 3.17. *For Lebesgue-a.e. $r \in [a, b]$, the measure ν_r is f_r-adapted.*

Proof. Let $\mathscr{D}_r$ be the discontinuity set of f_r. It is enough to show that

$$\nu_r\left\{x \in M_r : \liminf_{|n|\to\infty} \frac{1}{|n|} \log d(f_r^n(x), \mathscr{D}_r) < 0\right\} = 0 \quad \text{for a.e.} r \in [a, b].$$

For $\alpha > 0$, let

$$A_\alpha(r) := \left\{x \in M_b : \exists \text{ infinitely many } n \in \mathbb{Z} \text{ such that } \frac{1}{|n|} \log d(f_b^n(x), \mathscr{D}_r) < -\alpha\right\}.$$

It is enough to prove that $\nu_b[A_\alpha(r)] = 0$ for Lebesgue-a.e. $r \in [a, b]$. Let $I_\alpha(x) := \{a \le r \le b : x \in A_\alpha(r)\}$. Since $1_{A_\alpha(r)}(x) = 1_{I_\alpha(x)}(r)$, by Fubini's theorem we have $\int_a^b \nu_b[A_\alpha(r)]dr = \int_{M_b} \mathrm{Leb}[I_\alpha(x)]d\nu_b(x)$, so it is enough to prove that

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

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 $\sum_{n\in\mathbb{Z}} e^{-\alpha|n|} < \infty$, by the Borel–Cantelli lemma we get that $\mathrm{Leb}[I_\alpha(x)] = 0$. □

Combining Theorems 3.7 and 3.17, we obtain the following result, proved in [**LS19**] (see page 2622 for the definition of TMF).

THEOREM 3.18 *Let $\varphi : N \to N$ be as above. For each χ-hyperbolic measure μ, there exist a TMF (Σ_r, σ_r) and $\pi_r : \Sigma_r \to N$ Hölder continuous such that:*

(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* $\pi_r[\Sigma_r^\#]$.

Now consider surface maps with discontinuities and unbounded derivative. In some sense, the definition of $Q(x)$ given on page 30 allows us to concentrate the difficulty in this single parameter, but the statements need to be re-proved using this new definition. Let us see how to get Theorem 3.9. Remember from §2.4.1 that f is well behaved inside each ball $D_x = B(x, \mathfrak{r}(x))$. For $t > 0$, let $M_t = \{x \in M : d(x, \mathscr{D}) \geq t\}$. Since M has finite diameter (we are even assuming it is smaller than one), each M_t is precompact. Fix a countable open cover $\mathscr{P} = \{D_i\}_{i \in \mathbb{N}_0}$ of $M \backslash \mathscr{D}$ such that:

- $D_i := D_{z_i} = B(z_i, \mathfrak{r}(z_i))$ for some $z_i \in M$;
- for every $t > 0$, $\{D \in \mathscr{P} : D \cap M_t \neq \emptyset\}$ is finite.

For $\underline{\ell} = (\ell_{-1}, \ell_0, \ell_1)$, $\underline{k} = (k_{-1}, k_0, k_1)$, $\underline{a} = (a_{-1}, a_0, a_1)$, define

$$Y_{\underline{\ell},\underline{k},\underline{a}} := \left\{ \Gamma(x) \in Y : \begin{array}{ll} e^{\ell_i} \leq \|C(f^i(x))^{-1}\| < e^{\ell_i+1}, & -1 \leq i \leq 1 \\ e^{-k_i-1} \leq d(f^i(x), \mathscr{D}) < e^{-k_i}, & -1 \leq i \leq 1 \\ f^i(x) \in D_{a_i}, & -1 \leq i \leq 1 \end{array} \right\},$$

then proceed as in the proof of Theorem 3.9.

Another feature that requires better control is bounded distortion inside each $V^{s/u}[\underline{v}]$. This is proved in [**LM18**, Proposition 6.2]. In summary, under finer analysis, it is possible to prove all that is needed to obtain Theorem 3.7. This is [**LM18**, Theorem 1.3], and establishes problem #17 in Bowen's notebook [**Bow17**]. The proof actually works under greater generality that covers not only billiard maps but also some situations where the derivative of M and the behavior of $\exp_x$ are more complicated, for instance when $R, \nabla R, \nabla^2 R, \nabla^3 R$ grow at most polynomially fast with respect to the distance to $\mathscr{D}$, *inter alia* when M is a moduli space of curves equipped with the Weil–Petersson metric; see [**BMW12**].

4. *Part three: Applications*

There are two canonical applications of Markov partitions:

- estimating the number of closed orbits;
- establishing ergodic properties of equilibrium measures.

Indeed, the Markov partition generates a finite-to-one extension of $\mathrm{NUH}_\chi^\#$, and it is possible to lift measures without increasing their entropy.

Let X be a set, $\mathscr{A}$ a sigma-algebra, and $T : X \to X$ a measurable $\mathbb{Z}$- or $\mathbb{R}$-action. Given a T-invariant probability measure μ, let $h_\mu(T)$ denote its *Kolmogorov–Sinaĭ entropy*. Given two such systems $(X, \mathscr{A}, T)$ and $(Y, \mathscr{B}, S)$, let $\pi : (X, \mathscr{A}) \to (Y, \mathscr{B})$ be a *surjective* measurable extension map, that is, $\pi \circ T = S \circ \pi$. Assume that $\pi^{-1}(y)$ is finite for *all* $y \in Y$.

Projection of measure. If $\widehat{\mu}$ is a T-invariant probability measure on $(X, \mathscr{A})$, then its projection $\mu = \widehat{\mu} \circ \pi^{-1}$ is an S-invariant probability measure on $(Y, \mathscr{B})$ such that $h_\mu(S) = h_{\widehat{\mu}}(T)$.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

Indeed, by the Abramov–Rokhlin formula, $h_{\widehat{\mu}}(T) - h_\mu(S)$ is equal to the average entropy on the fibers. Since each of them is finite, they do not carry entropy. See [ELW11] for a proof of the Abramov–Rokhlin formula. In general, it is much harder to lift measures without increasing the entropy. In our setting, we can do this.

Lift of measure. If μ is an S-invariant probability measure on $(Y, \mathscr{B})$, then

$$\widehat{\mu} = \int_Y \frac{1}{|\pi^{-1}(y)|}\Bigg(\sum_{x\in\pi^{-1}(y)} \delta_x\Bigg)\, d\mu(y)$$

is a T-invariant probability measure on $(X, \mathscr{A})$ such that $h_\mu(S) = h_{\widehat{\mu}}(T)$.

Firstly, one has to check that $\widehat{\mu}$ is well defined; see, for example, [Sar13, Proposition 13.2]. The preservation of entropy follows again from the Abramov–Rokhlin formula. In particular, if the variational principle holds in $(X, \mathscr{A}, T)$ and $(Y, \mathscr{B}, S)$, then their topological entropies coincide.

The above conclusions also hold if there are sets $X^\# \subset X$ and $Y^\# \subset Y$ such that:

- the restriction $\pi\restriction_{X^\#}: X^\# \to Y^\#$ is finite-to-one;
- every T-invariant probability measure is supported on $X^\#$, and every S-invariant probability measure is supported on $Y^\#$.

It is in this way that invariant measures for non-uniformly hyperbolic systems relate with invariant measures for symbolic spaces. In the next two sections we explain some of the applications.

4.1. *Estimates on the number of closed orbits.* Given two sequences $\{a_n\}_{n\geq 1}$, $\{b_n\}_{n\geq 1}$, let us write $a_n \asymp b_n$ if there are constants $C, n_0 > 1$ such that $C^{-1} \leq a_n/b_n \leq C$ for all $n \geq n_0$, and $a_n \sim b_n$ if $\lim_{n\to\infty}(a_n/b_n) = 1$. Assume that $f : M \to M$ is a transitive uniformly hyperbolic diffeomorphism, and let (Σ, σ, π) be a symbolic model. Given $n \geq 1$, let $\mathrm{Per}_n(f), \mathrm{Per}_n(\sigma)$ denote the number of periodic orbits of period n for f, σ respectively, and write $h = h_{\mathrm{top}}(f) = h_{\mathrm{top}}(\sigma)$. Recall that Σ has finitely many states. Also (Σ, σ) is transitive, and if f is topologically mixing then (Σ, σ) is as well [Bow70a, Proposition 30]. Since π is finite-to-one (indeed, bounded-to-one), $\mathrm{Per}_n(f) \asymp \mathrm{Per}_n(\sigma)$. If $p \geq 1$ is the period of the TMS, then $\mathrm{Per}_{pn}(\sigma) \sim pe^{pnh}$; see, for example, [Kit98, Observation 1.4.3]. Hence $\mathrm{Per}_{pn}(f) \asymp e^{pnh}$. If f is topologically mixing, then $\mathrm{Per}_n(f) \asymp e^{nh}$.

Now let $f : M \to M$ be non-uniformly hyperbolic, where M can have any dimension. Since π is finite-to-one, there is a constant $C > 0$ such that $\mathrm{Per}_n(f) \geq C \times \mathrm{Per}_n(\sigma)$ for all $n \geq 1$. The reverse inequality might not hold, because π only codes trajectories inside NUH_χ, and f can have many more (even uncountably many) periodic orbits outside of NUH_χ. Even if we only count isolated periodic orbits, the growth rate of $\mathrm{Per}_n(f)$ can be superexponential [Kal00].

For a TMS with countably many states, good estimates on $\mathrm{Per}_n(\sigma)$ are related to the existence of measures of maximal entropy, as observed by Gurevič [Gur69, Gur70]. He showed that every transitive TMS admits at most one measure of maximal entropy, and

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

such measure exists if and only if there is $p \geq 1$ such that, for every vertex v,

$$\#\{\underline{v} \in \Sigma : \sigma^{pn}(\underline{v}) = \underline{v}, v_0 = v\} \asymp e^{pnh_{\max}},$$

where $h_{\max} = h_{\max}(\sigma) = \sup\{h_\nu(\sigma) : \nu$ is σ-invariant probability measure on $\Sigma\}$. We add the assumption that f has a χ-hyperbolic measure of maximal entropy, in which case $h = h_{\text{top}}(f) = h_{\max}(\sigma)$. Measures of maximal entropy exist for C^∞ diffeomorphisms [**New89**], although they might not be hyperbolic (for example, the product of an Anosov diffeomorphism and an irrational rotation). For surface diffeomorphisms with positive topological entropy, ergodic measures of maximal entropy are χ-hyperbolic for every $\chi < h_{\text{top}}(f)$, by the Ruelle inequality.

Assuming that f possesses a χ-hyperbolic measure of maximal entropy, repeat the arguments used for uniformly hyperbolic systems inside a transitive component of (Σ, σ). If $p \geq 1$ is the period of this component, then $\text{Per}_{pn}(\sigma) \asymp e^{pnh}$, and so $\text{Per}_{pn}(f) \geq C \times e^{pnh}$ for all $n \geq n_0$. This was proved by Sarig in dimension two [**Sar13**, Theorem 1.1] and by Ben Ovadia in higher dimension [**BO18**, Theorem 1.4].

Let us make some comments on the period p. Assume that f is topologically mixing. Recently, Buzzi was able to use the finite-to-one coding to define a one-to-one coding. Together with a precise counting on the TMS, he concluded that $\text{Per}_n(f) \geq C \times e^{nh}$ for all $n \geq n_0$ [**Buz20**]. His proof works whenever the smooth system has a finite-to-one coding with some extra properties, which are true for diffeomorphisms [**BO18, Sar13**] and billiard maps [**LM18**].

Now let $\varphi : M \to M$ be a flow and (Σ_r, σ_r) be a TMF. Given $T > 0$, let $\text{Per}_T(\varphi)$ and $\text{Per}_T(\sigma_r)$ denote the number of closed trajectories of φ and σ_r respectively, with minimal period $\leq T$. Note that $(\underline{v}, t) \in \text{Per}_T(\sigma_r)$ if and only if there is a minimal period $n \geq 1$ such that $\underline{v} \in \text{Per}_n(\sigma)$ and $r_n(\underline{v}) \leq T$, hence estimating $\text{Per}_T(\sigma_r)$ is more complicated. For uniformly hyperbolic flows, there are precise estimates.

- Geodesic flows on closed hyperbolic surfaces: in constant curvature, Huber proved that $\text{Per}_T(\varphi) \sim e^T/T$ [**Hub59**]. In variable curvature, Sinaĭ gave the first estimates [**Sin66**], which were later significantly sharpened by Margulis [**Mar69**], who proved that $\text{Per}_T(\varphi) \sim Ce^{Th}/T$ where $C = 1/h$ (C. Toll, unpublished).
- Axiom A flows: Bowen proved that $\text{Per}_T(\varphi) \asymp e^{Th}/T$ [**Bow72a**]. If the flow is topologically weak mixing, Parry and Pollicott proved that $\text{Per}_T(\varphi) \sim e^{Th}/Th$ [**PP83**], and Pollicott and Sharp found an estimate for the error term [**PS01**].

We also mention a result for manifolds with Gromov hyperbolic fundamental group (for example, manifolds that admit a metric with Anosov geodesic flow). Knieper and Coornaert counted free homotopy classes of closed geodesics estimating the growth rate of conjugacy classes in the fundamental group [**CK02, Kni83**].

Now consider non-uniformly hyperbolic flows. For geodesic flows in non-positively curved rank-one manifolds, the following facts are known.

- Knieper showed that $\pi_0(T) \asymp e^{Th}/T$, where $\pi_0(T)$ counts the homotopy classes of simple closed geodesics with length less than T [**Kni97, Kni02**].
- For certain metrics constructed by Donnay [**Don88**] and Burns and Gerber [**BG89**], $\text{Per}_T(\varphi) \sim e^{Th}/Th$ [**Wea**].

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

For the flows in Theorem 3.18, if there exists a measure of maximal entropy then there is $T_0 > 0$ such that $\mathrm{Per}_T(\varphi) \geq C \times e^{Th}/T$ for all $T \geq T_0$ [**LS19**, Theorem 8.1]. This estimate strengthens Katok's bound $\liminf_{T\to\infty}(1/T)\log \mathrm{Per}_T(\varphi) \geq h$; see [**Kat80, Kat82**]. The proof in [**LS19**] uses a dichotomy for TMF; see [**LLS16**, Theorem 4.6].

We end this section by mentioning some results for two-dimensional billiard maps. As seen in §2.4.1, every billiard map preserves an invariant Liouville measure μ_{SRB}. Using the countable Markov partition constructed in [**BSC90**], Chernov proved that $\liminf(1/n)\log \mathrm{Per}_n(f) \geq h_{\mu_{\mathrm{SRB}}}(f)$ [**Che91**, Corollary 2.4]. Better estimates can be obtained using measures of maximal entropy. Recently, Baladi and Demers gave sufficient conditions for periodic Lorentz gases (Sinaĭ billiards with non-intersecting scatterers) to have measures of maximal entropy [**BD20**]. This occurs when the billiard map satisfies two properties.

(1) Finite horizon: there is no trajectory that makes only tangential collisions.

(2) $h_* > s_0 \log 2$.

The second assumption requires some explanation. Part of their work involves defining a topological entropy h_* for finite-horizon Lorentz gases, which is an upper bound for all metric entropies [**BD20**, Theorem 2.3(4)]. Fixing an angle $\theta_0 \approx \pi/2$ and $n_0 > 0$, let $s_0 \in (0, 1)$ be the smallest number such that any orbit of length n_0 has at most $s_0 n_0$ collisions with $|\theta| > \theta_0$. Under conditions (1)–(2), there is an f-adapted measure μ_* such that $h_{\mu_*}(f) = h_*$ [**BD20**, Theorem 2.4]. Using [**LM18**] and [**Buz20**], it follows that $\mathrm{Per}_n(f) \geq C \times e^{nh_*}$ for n sufficiently large.

Here are some examples of billiards satisfying conditions (1) and (2) of [**BD20**]. If $\mathcal{K}_{\min}$ is the minimum curvature of the scatterer boundaries and $\tau_{\min}$ is the minimum free flight time, then $h_* > \log(1 + 2\mathcal{K}_{\min}\tau_{\min})$. Consider the two-parameter family (r, R) of Lorentz gases in $\mathbb{T}^2$ with two discs as scatterers, one centered at the origin $(0, 0)$ with radius R and the other at $(1/2, 1/2)$ with radius r; see Figure 20. Baladi and Demers found a domain in the parameter space for which $\log(1 + 2\mathcal{K}_{\min}\tau_{\min}) \geq \frac{1}{2}\log 2 \geq s_0 \log 2$, hence [**BD20**, Theorem 2.4] applies. There are also numerical experiments dealing with scatterers located in a triangular lattice indicating that $h_* > s_0 \log 2$ whenever the scatterers do not intersect and the billiard has finite horizon [**GB95**].

4.2. *Equilibrium measures.* Let (Y, S), where Y is a complete metric separable space and $S : Y \to Y$ is continuous, and let $\psi : Y \to \mathbb{R}$ be a continuous potential. The following definitions are standard.

Topological pressure. The *topological pressure* of ψ is $P_{\mathrm{top}}(\psi) := \sup\{h_\mu(S) + \int \psi\, d\mu\}$, where the supremum ranges over all S-invariant probability measures for which $\int \psi\, d\mu$ makes sense and $h_\mu(S) + \int \psi\, d\mu \neq \infty - \infty$.

Equilibrium measure. An *equilibrium measure* for ψ is an S-invariant probability measure μ such that $P_{\mathrm{top}}(\psi) = h_\mu(S) + \int \psi\, d\mu$.

A special case occurs when $\psi \equiv 0$: equilibrium measures are measures of maximal entropy. If $\pi : (X, T) \to (Y, S)$ is finite-to-one, then equilibrium measures for ψ lift

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

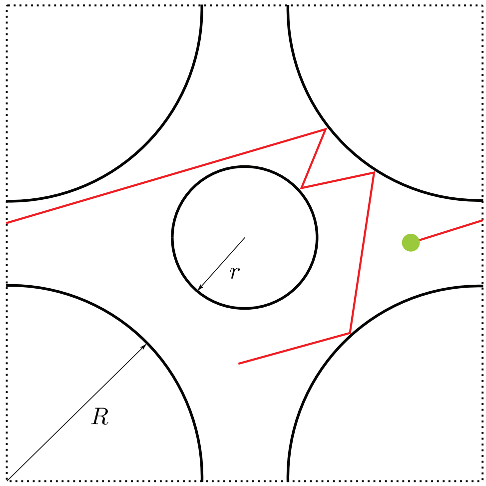

FIGURE 20. If r, R are chosen inside a specific polygon in the parameter space then there is an f-adapted measure of maximal entropy, and $\mathrm{Per}_n(f) \geq C \times e^{nh_*}$ for all large n.

to equilibrium measures for $\widehat{\psi} = \psi \circ \pi$. If π is Hölder continuous, then $\widehat{\psi}$ is Hölder continuous whenever ψ is. In our context, we can apply the thermodynamical formalism for Hölder continuous potentials in TMS to obtain ergodic properties of equilibrium measures of Hölder continuous potentials in uniformly and non-uniformly hyperbolic systems.

Since a transitive TMS with finitely many states has a unique measure of maximal entropy [**Par64**], every uniformly hyperbolic transitive diffeomorphism has a unique measure of maximal entropy [**Bow70a**], equal to the projection of the measure of maximal entropy in (Σ, σ). Prior to this, Gurevič obtained some partial results, using the work of Sinaĭ and of Berg (Berg proved that for hyperbolic toral automorphisms the Haar measure is the only measure of maximal entropy [**Ber69**]). Bowen also showed that every Hölder continuous potential has a unique equilibrium measure [**Bow75b**], and it is either Bernoulli or Bernoulli times a period [**Bow75a**].

Using the same analogy, Bowen and Ruelle proved that Hölder continuous potentials on uniformly hyperbolic flows have unique equilibrium measures [**BR75**]. In this case, equilibrium measures of (Σ_r, σ_r) are related to equilibrium measures of (Σ, σ); see [**BR75**, Proposition 3.1].

For non-uniformly hyperbolic $C^{1+\beta}$ surface diffeomorphisms, Sarig proved that each Hölder continuous potential has at most countably many ergodic hyperbolic equilibrium measures [**Sar13**, Theorem 1.2], and each of them is either Bernoulli or Bernoulli times a period [**Sar11**]. The proof uses that for topologically transitive TMS each Hölder continuous potential has at most one equilibrium measure [**BS03**], and different topologically transitive subgraphs of a TMS have disjoint vertex sets. The same holds for higher-dimensional diffeomorphisms [**BO18**], and for three-dimensional flows [**LS19**]. In the flow case, each such equilibrium measure is either Bernoulli or Bernoulli times a rotation [**LLS16**]. Since geodesic flows cannot have rotational components (they are

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

a particular case of Reeb flows), the following corollary holds: if S is a closed smooth orientable Riemannian surface with non-positive and non-identically zero curvature, then the geodesic flow of S is Bernoulli with respect to its (unique) measure of maximal entropy; see [**LLS16**, Corollary 1.3].

Let us mention some results on the uniqueness of measures of maximal entropy for non-uniformly hyperbolic geodesic flows. The uniqueness referred in the previous paragraph follows from the work of Knieper, who proved it for geodesic flows on closed rank-one manifolds [**Kni98**], and also for geodesic flows on symmetric spaces of higher rank [**Kni05**]. Gelfert and Ruggiero proved the uniqueness for geodesic flows on surfaces without focal points and genus greater than one [**GR19**]. Burns et al. proved the uniqueness of many equilibrium states (including some multiples of the geometric potential and the zero potential) of geodesic flows on rank-one manifolds [**BCFT18**], and there is a recent preprint that obtains similar results for geodesic flows on surfaces without focal points [**CKP20**]. There is also a recent preprint that proves the uniqueness of the measure of maximal entropy for geodesic flows on surfaces without conjugate points [**CKW19**].

Uniqueness of measures of maximal entropy for C^∞ transitive surface diffeomorphisms with positive topological entropy has recently been obtained. Essentially, the results of [**Sar13**] were not able to give uniqueness because it was not clear how to overrule the possibility that two measures in the surface lift to two different transitive components of the TMS. This difficulty was solved by Buzzi, Crovisier, and Sarig [**BCS18**], who showed that if two measures have positive entropy and are homoclinically related (a notion introduced in [**RHRHTU11**]) then they can be lifted to the same transitive component of the TMS. This can be regarded as a version of [**Bow70a**, Proposition 30]. They also prove that if the diffeomorphism is C^∞, then all measures of maximal entropy are homoclinically related (this uses Yomdin's theory), hence there is a unique measure of maximal entropy. It would be interesting to obtain similar results for three-dimensional flows and billiard maps.

Acknowledgements. These notes grew from a series of mini-courses the author has given in the past three years, in various institutions: ICERM (Providence, USA), Jagiellonian University (Cracow, Poland), ICTP (Trieste, Italy), Brigham Young University (Provo, USA), PUC-Chile (Santiago, Chile), University of Warwick (Coventry, UK), and Tsinghua University (Beijing, China). The author would like to thank the institutions and local organizers for their invitations, and Mark Demers and Gerhard Knieper for valuable suggestions. The first version of this paper was prepared while the author was supported by CNPq and by Instituto Serrapilheira, grant Serra-1709-20498. The author is supported by CNPq and Instituto Serrapilheira, grant 'Jangada Dinâmica: Impulsionando Sistemas Dinâmicos na Região Nordeste'.

REFERENCES

[Adl98] R. L. Adler. Symbolic dynamics and Markov partitions. *Bull. Amer. Math. Soc. (N.S.)* **35**(1) (1998), 1–56.

[Ano67] D. V. Anosov. Tangential fields of transversal foliations in U-systems. *Mat. Zametki* **2** (1967), 539–548.

[Ano69] D. V. Anosov. *Geodesic Flows on Closed Riemann Manifolds with Negative Curvature* (*Proc. Seteklov Inst. Math., 90*) (1967). Translated from the Russian by S. Feder. American Mathematical Society, Providence, RI, 1969.

[Ano70] D. V. Anosov. On certain class of invariant sets of smooth dynamical systems (in Russian). *Proc. 5th Int. Conf. on Non-Linear Oscillations*. Vol. 2, 1970, pp. 39–45.

[AW67] R. L. Adler and B. Weiss. Entropy, a complete metric invariant for automorphisms of the torus. *Proc. Nat/Acad. Sci. USA* **57** (1967), 1573–1576.

[AW70] R. L. Adler and B. Weiss. *Similarity of Automorphisms of the Torus* (*Memoirs of the American Mathematical Society, 98*) American Mathematical Society, Providence, RI, 1970.

[Bal95] W. Ballmann. *Lectures on Spaces of Nonpositive Curvature (DMV Seminar, 25)*. Birkhäuser, Basel, 1995. With an appendix by Misha Brin.

[BB17] M. Boyle and J. Buzzi. The almost Borel structure of surface diffeomorphisms, Markov shifts and their factors. *J. Eur. Math. Soc. (JEMS)* **19**(9) (2017), 2739–2782.

[BCFT18] K. Burns, V. Climenhaga, T. Fisher, and D. J. Thompson. Unique equilibrium states for geodesic flows in nonpositive curvature. *Geom. Funct. Anal.* **28**(5) (2018), 1209–1259.

[BCL] J. Buzzi, S. Crovisier, and Y. Lima. χ-universal Markov partitions for three dimensional flows. In preparation (2019).

[BCS18] J. Buzzi, S. Crovisier, and O. Sarig. Measures of maximal entropy for surface diffeomorphisms. *Preprint,* arXiv:1811.02240, 2018.

[BD04] M. Boyle and T. Downarowicz. The entropy theory of symbolic extensions. *Invent. Math.* **156**(1) (2004), 119–161.

[BD20] V. Baladi and M. F. Demers. On the measure of maximal entropy for finite horizon Sinai Billiard maps. *J. Amer. Math. Soc.* **33**(2) (2020), 381–449.

[Ber68] K. R. Berg. Entropy of torus automorphisms. *Topological Dynamics (Symposium, Colorado State Univ., Ft. Collins, CO 1967)*, Benjamin, New York, 1968, pp. 67–79.

[Ber69] K. R. Berg. Convolution of invariant measures, maximal entropy. *Math. Syst. Theory* **3** (1969), 146–150.

[BG89] K. Burns and M. Gerber. Real analytic Bernoulli geodesic flows on S^2. *Ergod. Th. & Dynam. Sys.* **9**(1) (1989), 27–45.

[BH11] K. Burns and B. Hasselblatt. The Sharkovsky theorem: a natural direct proof. *Amer. Math. Monthly* **118**(3) (2011), 229–244.

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

[BO18] S. Ben Ovadia. Symbolic dynamics for non-uniformly hyperbolic diffeomorphisms of compact smooth manifolds. *J. Mod. Dyn.* **13** (2018), 43–113.

[BO19] S. B. Ovadia. Canonically codable points and irreducible codings. *Preprint*, 2019, arXiv:1904.01391.

[Bow70a] R. Bowen. Markov partitions for Axiom A diffeomorphisms. *Amer. J. Math.* **92** (1970), 725–747.

[Bow70b] R. Bowen. Topological entropy and axiom A. *Global Analysis (Berkeley, CA, 1968) (Proceedings of Symposia in Pure Mathematics, XIV)*. American Mathematical Society, Providence, RI, 1970, pp. 23–41.

[Bow71] R. Bowen. Periodic points and measures for Axiom *A* diffeomorphisms. *Trans. Amer. Math. Soc.* **154** (1971), 377–397.

[Bow72a] R. Bowen. The equidistribution of closed geodesics. *Amer. J. Math.* **94** (1972), 413–423.

[Bow72b] R. Bowen. Periodic orbits for hyperbolic flows. *Amer. J. Math.* **94** (1972), 1–30.

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

[Bow78a] R. Bowen. Markov partitions are not smooth. *Proc. Amer. Math. Soc.* **71**(1) (1978), 130–132.

[Bow78b] R. Bowen. *On Axiom A Diffeomorphisms (Regional Conference Series in Mathematics, 35)*. American Mathematical Society, Providence, RI, 1978.

[Bow08] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms (Lecture Notes in Mathematics, 470)*, revised edn. Springer, Berlin. 2008. With a preface by David Ruelle, edited by Jean-René Chazottes.

[Bow17] R. Bowen. Rufus Bowen's Notebook, 2017. Available at https://bowen.pims.math.ca/.

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

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

[BP07] L. Barreira and Y. Pesin. *Nonuniform Hyperbolicity (Encyclopedia of Mathematics and its Applications, 115)*. Cambridge University Press, Cambridge, 2007. Dynamics of systems with nonzero Lyapunov exponents.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

[BP13] L. Barreira and Y. Pesin. *Introduction to Smooth Ergodic Theory (Graduate Studies in Mathematics, 148)*. American Mathematical Society, Providence, RI, 2013.

[BPVL20] L. Backes, M. Poletti, P. Varandas, and Y. Lima. Simplicity of Lyapunov spectrum for linear cocycles over non-uniformly hyperbolic systems. *Ergod. Th. & Dynam. Sys.* **40**(11) (2020), 2947–2969.

[BR75] R. Bowen and D. Ruelle. The ergodic theory of axiom A flows. *Invent. Math.* **29**(3) (1975), 181–202.

[BS02] M. Brin and G. Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, Cambridge, 2002.

[BS03] J. Buzzi and O. Sarig. Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps. *Ergod. Th. & Dynam.* **23**(5) (2003), 1383–1400.

[BSC90] L. A. Bunimovich, Y. G. Sinaĭ, and N. I. Chernov. Markov partitions for two-dimensional hyperbolic billiards. *Uspekhi Mat. Nauk* **45**(3) (1990), 97–134.

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

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

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

[Bur11] D. Burguet. C^2 surface diffeomorphisms have symbolic extensions. *Invent. Math.* **186**(1) (2011), 191–236.

[Buz20] J. Buzzi. The degree of Bowen factors and injective codings of diffeomorphisms. *J. Mod. Dyn.* **16** (2020), 1–36.

[Che91] N. I. Chernov. Topological entropy and periodic points of two-dimensional hyperbolic billiards. *Funktsional. Anal. i Prilozhen* **25**(1) (1991), 50–57.

[CK02] M. Coornaert and G. Knieper. Growth of conjugacy classes in Gromov hyperbolic groups. *Geom. Funct. Anal.* **12**(3) (2002), 464–478.

[CKP20] D. Chen, L.-Y. Kao, and K. Park. Unique equilibrium states for geodesic flows over surfaces without focal points. *Nonlinearity* **33**(3) (2020), 1118–1155.

[CKW19] V. Climenhaga, G. Knieper, and K. War. Uniqueness of the measure of maximal entropy for geodesic flows on certain manifolds without conjugate points. *Preprint*, 2019, arXiv:1903.09831.

[CM06] N. Chernov and R. Markarian. *Chaotic Billiards (Mathematical Surveys and Monographs, 127)*. American Mathematical Society, Providence, RI, 2006.

[Don88] V. J. Donnay. Geodesic flow on the two-sphere. I. Positive measure entropy. *Ergod. Th. & Dynam. Sys.* **8**(4) (1988), 531–553.

[Dow11] T. Downarowicz. *Entropy in Dynamical Systems (New Mathematical Monographs, 18)*. Cambridge University Press, Cambridge, 2011.

[Dua94] P. Duarte. Plenty of elliptic islands for the standard family of area preserving maps. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **11**(4) (1994), 359–409.

[Ebe01] P. Eberlein. Geodesic flows in manifolds of nonpositive curvature. *Smooth Ergodic Theory and its Applications (Seattle, WA, 1999) (Proceedings of Symposia in Pure Mathematics, 69)* American Mathematical Society, Providence, RI, 2001, pp. 525–571.

[ELW11] M. Einsiedler, E. Lindenstrauss, and T. Ward. Entropy in dynamics, 2011. Available at http://maths.dur.ac.uk/tpcc68/entropy/welcome.html.

[Fer99] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Math.* **206**(1–3) (1999), 145–154.

[FHY83] A. Fathi, M.-R. Herman, and J.-C. Yoccoz. A proof of Pesin's stable manifold theorem. *Geometric Dynamics (Rio de Janeiro, 1981) (Lecture Notes in Mathematics, 1007)*. Springer, Berlin, 1983, pp. 177–215.

[Fil19] S. Filip. Notes on the multiplicative ergodic theorem. *Ergod. Th. & Dynam. Sys.* **39**(5) (2019), 1153–1189.

[GB95] P. Gaspard and F. Baras. Chaotic scattering and diffusion in the Lorentz gas. *Phys. Rev. E (3)* **51**(6, part A) (1995), 5332–5352.

[GR19] K. Gelfert and R. O. Ruggiero. Geodesic flows modelled by expansive flows. *Proc. Edinb. Math. Soc. (2)* **62**(1) (2019), 61–95.

[Gur69] B. M. Gurevič. Topological entropy of a countable Markov chain. *Dokl. Akad. Nauk* **187** (1969), 715–718.

[Gur70] B. M. Gurevič. Shift entropy and Markov measures in the space of paths of a countable graph. *Dokl. Akad. Nauk* **192** (1970), 963–965.

[Had98] J. Hadamard. Sur la forme des lignes géodésiques à l'infini et sur les géodésiques des surfaces réglées du second ordre. *Bull. Soc. Math. France* **26** (1898), 195–216.

[Hed39] G. A. Hedlund. The dynamics of geodesic flows. *Bull. Amer. Math. Soc.* **45**(4) (1939), 241–260.

[Hof78] F. Hofbauer. β-shifts have unique maximal measure. *Monatsh. Math.* **85**(3) (1978), 189–198.

[Hof79] F. Hofbauer. On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. *Israel J. Math.* **34**(3) (1979), 213–237.

[Hof81] F. Hofbauer. On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. II. *Israel J. Math.* **38**(1–2) (1981), 107–115.

[Hop39] E. Hopf. Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung. *Ber. Verh. Sächs. Akad. Wiss. Leipzig* **91** (1939), 261–304.

[Hop40] E. Hopf. Statistik der Lösungen geodätischer Probleme vom unstabilen Typus. II. *Math. Ann.* **117** (1940), 590–608.

[Hub59] H. Huber. Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen. *Math. Ann.* **138** (1959), 1–26.

[Kal00] V. Y. Kaloshin. Generic diffeomorphisms with superexponential growth of number of periodic orbits. *Comm. Math. Phys.* **211**(1) (2000), 253–271.

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

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

[Kat82] A. Katok. Entropy and closed geodesics. *Ergod. Th. & Dynam. Sys.* **2**(3–4) (1983), 339–365.

[KH95] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications, 54)*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[Kit98] B. P. Kitchens. *Symbolic Dynamics*. Springer, Berlin, 1998.

[KM95] A. Katok and L. Mendoza. *Dynamical Systems with Non-Uniformly Hyperbolic Behavior*. Cambridge University Press, Cambridge, 1995. Supplement to 'Introduction to the modern theory of dynamical systems'.

[Kni83] G. Knieper. Das Wachstum der Äquivalenzklassen geschlossener Geodätischer in kompakten Mannigfaltigkeiten. *Arch. Math. (Basel)* **40**(6) (1983), 559–568.

[Kni97] G. Knieper. On the asymptotic geometry of nonpositively curved manifolds. *Geom. Funct. Anal.* **7**(4) (1997), pp. 755–782.

[Kni98] G. Knieper. The uniqueness of the measure of maximal entropy for geodesic flows on rank 1 manifolds. *Ann. of Math. (2)* **148**(1) (1998), 291–314.

[Kni02] G. Knieper. Hyperbolic dynamics and Riemannian geometry. *Handbook of Dynamical Systems*. Vol. 1A. North-Holland, Amsterdam, 2002, pp. 453–545.

[Kni05] G. Knieper. The uniqueness of the maximal measure for geodesic flows on symmetric spaces of higher rank. *Israel J. Math.* **149** (2005), 171–183.

[KSLP86] A. Katok, J.-M. Strelcyn, F. Ledrappier, and F. Przytycki. *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities (Lecture Notes in Mathematics, 1222)*. Springer, Berlin, 1986.

[KU07] S. Katok and I. Ugarcovici. Symbolic dynamics for the modular surface and beyond. *Bull. Amer. Math. Soc. (N.S.)* **44**(1) (2007), 87–132.

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

[LM18] Y. Lima and C. Matheus. Symbolic dynamics for non-uniformly hyperbolic surface maps with discontinuities. *Ann. Sci. Éc. Norm. Supér. (4)* **51**(1) (2018), 1–38.

[LS19] Y. Lima and O. M. Sarig. Symbolic dynamics for three-dimensional flows with positive topological entropy. *J. Eur. Math. Soc. (JEMS)* **21**(1) (2019), 199–256.

[Mar69] G. A. Margulis. Certain applications of ergodic theory to the investigation of manifolds of negative curvature. *Funksional. Anal. i Priložen.* **3**(4) (1969), 89–90.

[Mor34] M. Morse. Does instability imply transitivity? *Proc. Natl Acad. Sci. USA* **20**(1) (1934), 46–50.

[MT88] J. Milnor and W. Thurston. On iterated maps of the interval. *Dynamical Systems (College Park, MD, 1986–87) (Lecture Notes in Mathematics, 1342)*. Springer, Berlin, 1988, pp. 465–563.

[New89] S. E. Newhouse. Continuity properties of entropy. *Ann. of Math. (2)* **129**(2) (1989), 215–235.

[Ose68] V. I. Oseledec. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskov. Mat. Obšč.* **19** (1968), 179–210.

[Par64] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.

[Pes76] J. B. Pesin. Families of invariant manifolds that correspond to nonzero characteristic exponents. *Izv. Akad. Nauk SSSR Ser. Mat.* **40**(6) (1976), 1332–1379, 1440.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press

[Pes77a] J. B. Pesin. Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat. Nauk* **32**(4(196)) (1977), 55–112, 287.

[Pes77b] J. B. Pesin. Geodesic flows in closed Riemannian manifolds without focal points. *Izv. Akad. Nauk SSSR Ser. Mat.* **41**(6) (1977), 1252–1288, 1447.

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

[PS01] M. Pollicott and R. Sharp. Error terms for closed orbits of hyperbolic flows. *Ergod. Th. & Dynam. Sys.* **21**(2) (2001), 545–562.

[PT72] J. F. Plante and W. P. Thurston. Anosov flows and the fundamental group. *Topology* **11** (1972), 147–150.

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

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

[RHRHTU11] F. Rodriguez Hertz, M. A. Rodriguez Hertz, A. Tahzibi, and R. Ures. Uniqueness of SRB measures for transitive diffeomorphisms on surfaces. *Comm. Math. Phys.* **306**(1) (2011), 35–49.

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

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

[Shu87] M. Shub. *Global Stability of Dynamical Systems*. Springer, New York, 1987.

[Sin66] J.. G. Sinaĭ. Asymptotic behavior of closed geodesics on compact manifolds with negative curvature. *Izv. Akad. Nauk SSSR Ser. Mat.* **30** (1966), 1275–1296.

[Sin68a] J. G. Sinaĭ. Construction of Markov partitionings. *Funkcional. Anal. i Priložen.* **2**(3) (1968), 70–80 (Loose errata).

[Sin68b] J. G. Sinaĭ. Markov partitions and U-diffeomorphisms. *Funkcional. Anal. i Priložen.* **2**(1) (1968), 64–89.

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

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

[Tak73] Y. Takahashi. Isomorphisms of β-automorphisms to Markov automorphisms. *Osaka J. Math.* **10** (1973), 175–184.

[Wea] B. A. Weaver. Precise asymptotic growth rate of periodic orbits for a class of non-uniformly hyperbolic geodesic flows. *Preprint*, 2011.

[Yoc15] J.-C. Yoccoz. A proof of Jakobson's theorem, 2015, pp. 34. Available at https://www.college-de-france.fr/media/jean-christophe-yoccoz/UPL7416254474776698194%20Jakobson%20jcy.pdf.

[You98] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)* **147**(3) (1998), 585–650.

https://doi.org/10.1017/etds.2020.80 Published online by Cambridge University Press