Труды Москов. Матем. Общ.
Том 72 (2011), вып. 2

Trans. Moscow Math. Soc.
2011, Pages 193–217
S 0077-1554(2012)00196-4
Article electronically published on January 12, 2012

ON C^2-STABLE EFFECTS OF INTERMINGLED BASINS OF ATTRACTORS IN CLASSES OF BOUNDARY-PRESERVING MAPS

V. A. KLEPTSYN AND P. S. SALTYKOV

Abstract. In the spaces of boundary-preserving maps of an annulus and a thickened torus, we construct open sets in which every map has intermingled basins of attraction, as predicted by I. Kan.

Namely, the attraction basins of each of the boundary components are everywhere dense in the phase space for such maps. Moreover, the Hausdorff dimension of the set of points that are not attracted by either of the components proves to be less than the dimension of the phase space itself, which strengthens the result following from the argument due to Bonatti, Diaz, and Viana.

1. Introduction

Studying attractors is one of the most important topics in the theory of dynamical systems. It is worth noting that there are many nonequivalent definitions of what an attractor of a dynamical system is; various definitions can be found, for example, in [GI96] (see also [G01]). However, all so far known examples in which different definitions give different attractors are not typical.

One existing viewpoint is that this situation is generic. Namely, all definitions of attractors for a (metrically) typical dynamical system give the same set, splitting into finitely many components each of which is the maximal attractor of some neighborhood of itself. In particular, the Palis conjecture (e.g., see [Pa00] and [Pa05, Sec. 2.7]) claims that there exist finitely many SRB measures, that the union of their basins of attraction has full measure, and that the attractor components supporting the SRB measures are stochastically stable for the dynamical systems belonging to some everywhere dense set.

There is also a point of view, contrary in spirit, that there exist typical (in a different, topological sense) examples of the opposite: systems whose attractor behavior is chaotic. In particular, Ruelle's conjecture [Ru01] asks whether there exist typical dynamical systems in which there is no convergence of time averages for the initial data in a set of positive Lebesgue measure. (Thus, part of the conclusions of the Palis conjecture cannot hold for such systems; see also [T08].)

An attractor often splits into several components each of which has its own basin of attraction. For a Milnor attractor (see [I08]), these components are closed invariant subsets whose complement in the attractor is closed as well.

When studying attractors and their basins as well as the basins of attraction of their individual components, it is natural to consider the set of points that do not tend to the attractor. We refer to this set as the *exceptional set*. By definition, the exceptional set

2010 *Mathematics Subject Classification.* Primary 37C70; Secondary 37D25.

Key words and phrases. Dynamical system, attractor, stability, partially hyperbolic skew product, Hölder rectifying map.

Supported in part by RFBR grant no. 10-01-00739-a and joint RFBR–CNRS grant no. 10-01-93115-NTsNI_a.

is "small". In particular, the "smallness" of the exceptional set for a Milnor attractor means that it has zero Lebesgue measure.

It is natural to expect that the boundaries of attraction basins of individual components are "good" sets. However, Kan [K94] produced an example of a self-map of an annulus for which the Milnor attractor consists of both components of the boundary of the annulus and the attraction basin of each of the boundary circles is metrically everywhere dense in the annulus (see Theorem 1 below). This effect is known as *intermingled basins of attraction*.

Kan's example is a partially hyperbolic skew product over the angle tripling map of the circle. In the same paper, he announced the stability of the effect of intermingled basins of attraction under small perturbations of this example in the class of boundary-preserving $C^{1+\alpha}$-smooth self-maps of the annulus and the stability of a similar example in a class of diffeomorphisms of a thickened torus. To prove this stability, he intended to use the partial hyperbolicity of the example constructed: every sufficiently close map has smooth central leaves, and hence its dynamics is conjugate to some ("rectified") skew product close to the original one.

However, Kan has not published a proof of this claim, apparently owing to the difficulties encountered in the suggested scheme. Namely, although each individual leaf remains smooth under small perturbations, the central foliation as a whole may fail to be absolutely continuous. Therefore, even knowing how points Lebesgue typical after the conjugation behave, one cannot a priori say anything about the behavior of a Lebesgue typical point under the original perturbed map. Indeed, the conjugation may take a set of measure zero to a set of full measure and vice versa. Moreover, a set that meets each central leaf in a set of measure zero or even in finitely many points may prove to be a set of full measure. This effect was dubbed "Fubini foiled". Such examples were constructed for an arbitrary foliation with smooth leaves in the illustrative paper [Mi97] and for a central foliation in the paper [SW00]. In view of the preceding, one cannot conclude that intermingled basins of attraction occur in the original system even having established this for the rectified perturbed map.

Kan's conjecture was proved by Bonatti, Diaz, and Viana [BDV05]; although the example analyzed in Sec. 11.1.1 of their book is a skew product, all the arguments used in the proof (Pesin's theory and distortions of the images of a transversal) survives C^2-small perturbations. Smooth skew products with base a circle and fiber a closed interval were also studied in [BM07], where the effect of intermingled basins of attraction was established for a wide class of skew products. Finally, note the paper [BLR] by Bleher, Lyubich, and Roeder, where the basins of attraction are described for a specific dynamical system on a cylinder arising from the hierarchical Ising model.

The present paper deals with the proof of Kan's conjecture (both for an annulus and for a thickened torus) with the use of the rectification technique and Hölder estimates. It turns out that the use of this technique permits one to strengthen the result due to Bleher, Lyubich, and Roeder by additionally showing that the Hausdorff dimension of the exceptional set is strictly less than the dimension of the phase space in both cases. (Anyway, we think that the very possibility of such a proof—in connection with numerous situations where the rectification technique is applied together with Hölder estimates; e.g., see [GI99, G06, Os10, GIKN05]—is of interest in itself.)

Namely, Gorodetskiĭ [GI99] (see also [G06]) discovered that, under a perturbation of a partially hyperbolic skew product, the central foliation and hence the rectifying map prove to satisfy the *Hölder condition*. Moreover, Ilyashenko and Negut [IN10] proved that the Hölder exponent tends to 1 as the perturbation tends to zero. This permits us to prove the desired assertion: instead of intermingling itself, we establish stronger properties (such as an estimate of the Hausdorff dimension of the exceptional set), which, however, practically do not become worse under Hölder conjugation.

1.1. **Outline of the paper.** The main results, Theorems 1 and 2, are stated in Sec. 2. Section 3 presents the ideas of the proof. Section 3.2 states some results (Theorems 3 and 4) for Hölder skew products, and Theorems 1 and 2 are derived from these results in Sec. 3.3. In Sec. 4, we recall some auxiliary material, namely, the probabilistic and dynamical large deviation theorems and the special ergodic theorem. Sections 5 and 6 deal with the proof of Theorems 3 and 4, respectively.

2. Statement of the main results

The following assertions strengthening the conclusion of Kan's conjecture (and of the Bonatti–Diaz–Viana theorem) are our main results. (The second part of Theorem 1 was already proved in [IKS08], but we present it here to make the exposition complete.)

Theorem 1. *There exists an open set V in the space of boundary-preserving C^2 maps of the annulus $X^K = \mathcal{S}^1 \times [0,1]$ onto itself such that every map $f \in V$ has the following properties*:

(i) *The orbit of Lebesgue almost every initial point $p \in X^K$ tends to one of the two components of the boundary,*

$$f^n(p) \xrightarrow[n\to\infty]{} A_0^K := \mathcal{S}^1 \times \{0\} \tag{1}$$

or

$$f^n(p) \xrightarrow[n\to\infty]{} A_1^K := \mathcal{S}^1 \times \{1\}. \tag{2}$$

Moreover, the exceptional set, that is, the set of points whose iterates tend to neither A_0^K nor A_1^K, is of Hausdorff dimension less than 2 and separated from 2 uniformly on V.

(ii) *The attraction basins $\mathcal{B}_j^K := \{p \in X^K \mid f^n(p) \xrightarrow[n\to\infty]{} A_j^K\}$, $j = 0, 1$, are metrically everywhere dense*; *i.e., $\mu_{\mathrm{Leb}}(\mathcal{B}_0^K \cap U) > 0$ and $\mu_{\mathrm{Leb}}(\mathcal{B}_1^K \cap U) > 0$ for each open set $U \subset X^K$.*

Theorem 2. *There exists an open set V in the space of C^2 diffeomorphisms of the thickened torus $X^T = \mathbb{T}^2 \times [0,1]$ onto itself such that every diffeomorphism $f \in V$ has the following properties*:

(i) *The set of points x whose iterates tend to neither of the boundary components $A_0^T = \mathbb{T}^2 \times \{0\}$ and $A_1^T = \mathbb{T}^2 \times \{1\}$ has zero Lebesgue measure. Moreover, the Hausdorff dimension of this set is strictly less than 3 and is separated from 3 uniformly on V.*

(ii) *The attraction basins $\mathcal{B}_j^T := \{p \in \mathbb{T}^2 \times [0,1] \mid f^n(p) \xrightarrow[n\to\infty]{} A_j^T\}$, $j = 0, 1$, are metrically everywhere dense in $\mathbb{T}^2 \times [0,1]$.*

Note that the statement of Theorem 2 looks more natural than that of Theorem 1: the condition that the boundary is preserved is automatically satisfied for the class of diffeomorphisms of the thickened torus, but it should be stipulated explicitly for maps of the annulus.

3. Kan's construction and Hölder skew products

3.1. **Kan's idea.** The first step corresponds to the scheme suggested by Kan [K94]. Namely, we seek the desired domain V for Theorems 1 and 2 in a sufficiently small neighborhood of soft skew products over the circle doubling map and a linear Anosov diffeomorphism, respectively.

Mainly, the argument in both cases is completely similar, and we begin to carry it out for both cases simultaneously. Let B be either the circle S^1 or the torus $\mathbb{T}^2$; let μ be

Lebesgue measure on B, and let $T\colon B \to B$ be the circle doubling map $T(y) = 2y \bmod 1$ or a linear Anosov diffeomorphism

$$T((y_1, y_2)) = M\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad M \in SL(2, \mathbb{Z}), \quad |\operatorname{tr} M| > 2,$$

of the two-dimensional torus, respectively. Finally, let $X := B \times [0,1]$, $A_0 = B \times \{0\}$, and $A_1 = B \times \{1\}$.

First, consider one skew product

$$F\colon B \times [0,1] \to B \times [0,1], \quad F(y,x) = (T(y), f_y(x)), \tag{3}$$

where $f_y(x)$ is a family of diffeomorphisms of $[0,1]$ depending on the parameter $y \in B$.

Suppose that this skew product satisfies the following two conditions.

(U1) The central ("along-the-leaf") Lyapunov exponents of the boundary components $x = 0$ and $x = 1$ with respect to (invariant) Lebesgue measure μ are negative,

$$\lambda_0 = \int_B \log(f_y)'(0)\, d\mu(y) < 0 \quad \text{and} \quad \lambda_1 = \int_B \log(f_y)'(1)\, d\mu(y) < 0. \tag{4}$$

(U2) There exist periodic points $y_{(0)}$ and $y_{(1)}$ of the base B,

$$y_{(0)} = T^{k_0}(y_{(0)}), \quad y_{(1)} = T^{k_1}(y_{(1)}),$$

such that the compositions

$$f_{T^{k_0-1}(y_{(0)})} \circ f_{T^{k_0-2}(y_{(0)})} \circ \dots \circ f_{y_{(0)}} \quad \text{and} \quad f_{T^{k_1-1}(y_{(1)})} \circ f_{T^{k_1-2}(y_{(1)})} \circ \dots \circ f_{y_{(1)}}$$

over the respective periods each have exactly two fixed points, $x = 0$ and $x = 1$, all these points are hyperbolic, the point $x = 0$ is attracting for the first composition and repelling for the second composition, and the point $x = 1$ is repelling for the first composition and attracting for the second composition.

Then, on the one hand, the conditions imposed on the Lyapunov exponents imply that the behavior of a typical point in a small neighborhood of the boundary component $x = 0$ (or $x = 1$) is described by the corresponding "linear" dynamics. Hence the smaller this neighborhood, the larger is the fraction of points tending to $x = 0$. On the other hand, since an arbitrarily small domain contains points whose y-coordinate coincides with or is very close to one of the preimages of $y_{(0)}$ (or $y_{(1)}$), which form an everywhere dense set, it follows that every domain contains both points attracted by A_0 and points attracted by A_1. (A slightly more complicated argument shows that the iterates of almost every initial point in this skew product tend to either A_0 or A_1.)

Note that since we are interested in a small neighborhood of F, its elements can be thought of as perturbations of F. Next, assume that the maps f_y are sufficiently close to the identity map. Then the corresponding skew product is *partially hyperbolic*. (For a skew product over the circle doubling map, we have in mind the meaning of the term as applied to noninvertible maps; see below.) A standard argument (see [HPS77] and [GI99]) shows that an arbitrary small perturbation $\widetilde{F}$ of the map F has smooth central leaves close to those of the original product. Moreover, the dynamics on the set of central leaves proves to be conjugate to the original map T of the base.

Thus, the map $\widetilde{F}$ proves to be conjugate (by an x-coordinate-preserving conjugation $\Psi\colon (y,x) \mapsto (\psi(y,x), x)$, where ψ is a map constant on the central leaves and conjugating the dynamics on these leaves with the original dynamics on the base) to some skew product of the form

$$G\colon (y,x) \mapsto (T(y), \tilde{f}_y(x)) \tag{5}$$

with leaf maps $\tilde{f}_y$ that are smooth and close to the original ones and with the same map on the base.

After that, one would like (and this exactly was Kan's plan) to complete the proof by saying that the conditions imposed on the family of maps f_y are open and hence are also satisfied by the small perturbation $\tilde{f}_y$ of the original family. Unfortunately, this argument breaks down, because, as was already mentioned, although each individual central leaf proves to be smooth, the central foliation may even fail to be absolutely continuous, and hence so may the rectifying map Ψ.

Therefore, we need a more careful analysis both of the properties to be proved and of the behavior of the conjugating map.

3.2. **Estimates of the Hausdorff dimension.** To carry out this analysis, we first note that the central foliation of $\widetilde{F}$ satisfies the Hölder condition, as was shown in [GI99, Theorem 3] (see also [G06]). Moreover, it was proved in [IN10] that, for perturbations of skew products over the circle doubling map or over a linear diffeomorphism of the torus, the Hölder exponents of the central foliation tend to 1 uniformly in a neighborhood of the map F as this neighborhood shrinks. Finally, for each fixed Hölder exponent, the corresponding constant remains uniformly bounded in a sufficiently small neighborhood of F.

To make the exposition self-contained, let us give the main idea of the proof of these assertions in [IN10]. Both the central-stable and central-unstable foliations are invariant under the dynamics; the Hölder condition holds automatically for "remote" points. On the other hand, for a sufficiently small perturbation of the skew product, the α-Hölder property does not become worse as the corresponding image or preimage is taken; this permits one to extend the condition from the remote points to all pairs of points. Finally, the central foliation is the intersection of the central-stable and central-unstable foliations and hence satisfies the α-Hölder condition as well.

Owing to the above-mentioned results, we can prove Theorems 1 and 2 by obtaining stronger assertions for the skew product (5), which are nevertheless preserved (or, more precisely, do not become significantly worse) under Hölder conjugations. These assertions include estimates of the Hausdorff dimension of the exceptional set and related estimates; we state the corresponding assertions below (see Theorems 3 and 4).

Furthermore, $\tilde{f}_y$ proves to be a Hölder function of y owing to the same Hölder property of the conjugation. Hence we work in the class of Hölder skew products.

Consider the space of skew products with fiber $[0,1]$. We equip the space $\mathrm{Diff}^2([0,1])$ of diffeomorphisms with the metric

$$d(f,g) = \max(\mathrm{dist}_{C^2}(f,g), \mathrm{dist}_{C^2}(f^{-1},g^{-1}))$$

and treat skew products as continuous maps of the base B into $\mathrm{Diff}^2([0,1])$ with the corresponding metric $d_{\mathcal{H}}(F,\tilde{F}) = \max_{y\in B} d(f_y,\tilde{f}_y)$ of the space of continuous maps.

Definition 1. For any C and α, we say that a skew product is of *Hölder class* (C,α) if it satisfies the Hölder condition with exponent α and constant C as a map of B into $\mathrm{Diff}^2([0,1])$. We denote the set of all Hölder skew products of class (C,α) by $\mathcal{H}^2_{(C,\alpha)}$.

It is easily seen that $\mathcal{H}^2_{(C,\alpha)}$ is a closed subset of the set of all skew products.

In further reasoning, we also need the following two technical conditions:

(U3) After the change of coordinates $\Phi\colon (0,1)\to\mathbb{R}$, $z=\Phi(x)=\ln(x/(1-x))$, the derivatives of the maps $\tilde{f}_y$ prove to be strictly greater than 1 (but not separated from 1 in general),

$$\forall y\in B,\ \forall z\in\mathbb{R}\quad (\Phi\circ\tilde{f}_y\circ\Phi^{-1})'(z) > 1. \tag{6}$$

(U3′)

$$\forall y \in B \quad \frac{(\tilde{f}_y)''(0)}{2(\tilde{f}_y)'(0)} - 1 + (\tilde{f}_y)'(0) > 0, \quad -\frac{(\tilde{f}_y)''(1)}{2(\tilde{f}_y)'(1)} - 1 + (\tilde{f}_y)'(1) > 0. \tag{7}$$

Examples of skew products satisfying all conditions (U1)–(U3′) are given at the end of the subsection.

The first of the desired results for skew products is given by the following theorem.

Theorem 3 (Estimate of the Hausdorff dimension of the exceptional set)**.** *Let $C > 0$ and $\alpha \in (0,1)$ be arbitrary, and let $G(y,x) = (T(y), \tilde{f}_y(x))$ be a Hölder skew product of class (C,α) satisfying conditions* (U1) *and* (U3)*. Then the Hausdorff dimension of the set of points (x,y) that tend to neither A_0 nor A_1 is strictly less than the total dimension of the phase space* (*which is* 2 *and* 3 *for products over the circle and the torus, respectively*)*.*

Moreover, if the system additionally satisfies condition (U3′)*, then the dimension of the exceptional set is separated from the dimension of the phase space uniformly in a sufficiently small neighborhood of the original skew product in the space $\mathcal{H}^2_{(C,\alpha)}$ of Hölder skew products.*

Remark 1. Condition (U3′) in Theorem 3 is needed to ensure that condition (U3) (which is the expansion property in the new coordinates in a neighborhood of the boundary $A_0 \cup A_1$) is preserved under small perturbations (see Proposition 2 below).

Remark 2. For C^3-smooth systems, a sufficient condition for condition (U3) to hold is that the Schwarzian derivative of $\tilde{f}_y$ be negative. (This negativeness was used in [BM07].) This is related to the fact that maps with negative Schwarzian derivative increase the cross-ratio of four points and that the distance in the new coordinates is the logarithm of the cross-ratio in the old coordinates,

$$\begin{aligned}\Phi(x_1) - \Phi(x_2) &= \ln\left(\frac{x_1}{1-x_1}\right) - \ln\left(\frac{x_2}{1-x_2}\right)\\ &= \ln\left(\frac{x_1}{1-x_1} : \frac{x_2}{1-x_2}\right) = \ln[0:1:x_1:x_2].\end{aligned}$$

It remains to state an assertion that would imply that the attraction basins are metrically dense. Here we cannot use an estimate of the Hausdorff dimension directly. Indeed, we wish to prove that the points tending to A_0, as well as the points tending to A_1, form a set of positive measure in any neighborhood. Thus, the measure of the set of points that do *not* tend to A_0 should be positive, and hence the complement of the basin of attraction of A_0 should be of full Hausdorff dimension. (Clearly, the same argument holds for the basin of attraction of A_1.)

However, it turns out that one can cover the set of points that cease to tend to one or the other boundary component by balls whose total measure can be estimated by the sum of a converging geometric series. Since a Hölder conjugation with exponent close to 1 can only slightly change the common ratio of such a series, it follows that the estimate for the sum will remain almost unchanged; in particular, this sum will be smaller than the total measure of the neighborhood. An assertion about such a cover is stated in the next theorem, from which Corollary 1 and Remark 3 below derive the desired assertion that the basins of attraction are metrically dense.

Let $U_r(p)$ be the ball of radius r centered at p in the base B. From now on in this section, all balls are understood as balls in the base. Next, let λ_u be the Lyapunov exponent of expansion of the system in the base; we set $\lambda_u = 2$ for the case of the circle doubling map and take λ to be the modulus of the eigenvalue of M greater than 1 in modulus for the case of a linear Anosov diffeomorphism $M\colon \mathbb{T}^2 \to \mathbb{T}^2$.

Definition 2. Consider a skew product $G(y,x) = (T(y), \tilde{f}_y(x))$. Let a point $P = (p, x)$ belong to the attraction basin $\mathcal{B}_0 := \{Q \mid G^n(Q) \xrightarrow[n\to\infty]{} A_0\}$ of the boundary A_0. We say that *the basin* $\mathcal{B}_0$ *is* (c_1, c_2)*-horizontally dense* if the following conditions are satisfied:

- $c_1 > 0$, $0 < c_2 < \dim B$.
- For all sufficiently large n, the set of points that lie in the "horizontal neighborhood"
$$U_n := U_{\lambda_u^{-n}}(p) \times \{x\}$$
of P and whose images do not tend to A_0 can be covered by a union over $m \geq (1+c_1)n$ of unions of at most $\lambda_u^{c_2(m-n)}$ balls of radius λ_u^{-m}.

Theorem 4 (Strengthened metric density)**.** *Let* $C > 0$ *and* $\alpha \in (0,1)$ *be arbitrary, and let* $G(y,x) = (T(y), \tilde{f}_y(x))$ *be a Hölder skew product of class* (C, α) *satisfying conditions* (U1) *and* (U2)*. Then there exist constants* $c_1 > 0$ *and* $c_2 < \dim B$ *such that the basin* $\mathcal{B}_0$ *is* (c_1, c_2)*-horizontally dense for each point* $P = (p, x)$ *in the stable manifold of the point* $(y_{(0)}, 0)$ (*where* $y_{(0)}$ *is the point in condition* (U2)).

Moreover, the constants c_1 *and* c_2 *can be chosen uniformly in a small neighborhood of the original system in the space* $\mathcal{H}^2_{(C,\alpha)}$.

Corollary 1. *Under the assumptions of Theorem* 4*, in every neighborhood of each point of the stable manifold* $W^s(y_{(0)}, 0)$ (*respectively,* $W^s(y_{(1)}, 1)$) *the set* $\mathcal{B}_0$ (*respectively,* $\mathcal{B}_1$) *of points whose iterates tend to* A_0 (*respectively, to* A_1) *is of positive Lebesgue measure.*

Proof of Corollary 1. Note that it suffices to prove that the measure of a "horizontal" section of such a set, i.e., the measure of the intersection of the set $\{q \in B \mid G^n(q,x) \to A_0\}$ with an arbitrarily small neighborhood of p, is positive. Indeed, all points $(p, \bar{x})$ with $\bar{x} < x$ belong to $W^s(y_{(0)}, 0)$ as well, and hence the horizontal sections of $\mathcal{B}_0$ passing through them will have positive measure. After that, an application of the Fubini theorem to the horizontal sections completes the proof.

Let n be sufficiently large; let us find a lower bound for the measure of the set of points of the form (q, x), $q \in U_{\lambda_u^{-n}}(p)$, tending to A_0. Namely, the measure of a ball of radius λ_u^{-m} differs from the measure of a ball of radius λ_u^{-n} by the factor $\lambda_u^{-(m-n)\cdot\dim B}$, and we can use the conclusion of the theorem to obtain

$$\begin{aligned}
\frac{\mu\{q \in U_{\lambda_u^{-n}}(p) \mid G^n(q,x) \to A_0\}}{\mu(U_{\lambda_u^{-n}}(p))} &\geq 1 - \sum_{m=(1+c_1)n}^{\infty} \lambda_u^{-(m-n)\cdot\dim B} \cdot \lambda_u^{c_2(m-n)} \\
&\geq 1 - \sum_{m=(1+c_1)n}^{\infty} \lambda_u^{-(m-n)\cdot(\dim B - c_2)} = 1 - \frac{\lambda_u^{-c_1 n(\dim B - c_2)}}{1 - \lambda_u^{-(\dim B - c_2)}}.
\end{aligned} \tag{8}$$

In particular, we obtain a positive lower bound for the desired measure for all sufficiently large n, which completes the proof. □

Remark 3. The conclusion of Theorem 4 (and hence Corollary 1) remains valid under Hölder "horizontal" conjugations of the form $\Psi(y,x) = (\psi(y,x), x)$ with Hölder exponents sufficiently close to 1.

Proof. Under such a conjugation with Hölder exponent α, the image of a ball of radius λ_u^{-n} contains a ball of radius $c\lambda_u^{-n/\alpha}$, and the images of balls of radius λ_u^{-m} can be covered by balls of radius $C\lambda_u^{-\alpha m}$. This readily implies that, for α sufficiently close to 1, the conclusion of Theorem 4 remains valid. (The geometric series in the estimate (8) remains converging, and its common ratio does not change much.)

On the other hand, since such a conjugation preserves the horizontal foliation (and the possibility to apply the Fubini theorem to it), we see that Corollary 1 remains valid as well. □

3.3. **Derivation of the main results from the theorems for skew products.** Let us rigorously carry out the argument outlined at the end of Sec. 3.1 and derive the main results for smooth maps, namely, Theorems 1 and 2, from Theorems 3 and 4.

Proposition 1. (i) *Let* $F\colon (y,x) \mapsto (2y \bmod 1, f_y(x))$ *be a boundary-preserving* C^2*-smooth map of the annulus* $\mathbb{S}^1 \times [0,1]$ *onto itself. Suppose that* F *is a skew product over the circle doubling map and satisfies conditions* (U1), (U2), (U3), *and* (U3′) *and that the maps* f_y *are sufficiently close to the identity map. Then the conclusions of Theorem* 1 *are satisfied in some neighborhood of* F *in the space of boundary-preserving* C^2*-smooth maps of the annulus onto itself.*

(ii) *Let* $F\colon (y,x) \mapsto (T(y), f_y(x))$ *be a* C^2*-smooth partially hyperbolic diffeomorphism of the thickened torus* $\mathbb{T}^2 \times [0,1]$ *onto itself. Suppose that* F *is a skew product over a linear Anosov diffeomorphism and satisfies conditions* (U1), (U2), (U3), *and* (U3′). *Then the conclusions of Theorem* 2 *are satisfied in some neighborhood of* F *in the space of boundary-preserving* C^2*-smooth diffeomorphisms of the thickened torus onto itself.*

Proof. In both cases, the maps f_y are sufficiently close to the identity map, and hence F is *partially hyperbolic* (in case (i), in the meaning of the term as applied to noninvertible maps).

Namely, in case (i), the lift of F to the universal covering $\mathbb{R} \times [0,1]$ over the annulus admits a *dominated splitting* (see [HPS77]), and the dynamics is uniformly expanding in the dominating direction. By projecting the corresponding foliations (the strongly unstable foliation and the central foliation) back onto $\mathbb{S}^1 \times [0,1]$, we see that the vertical segments are central leaves and that the central foliation can be found by iterating the operation of taking the full F-preimage of a foliation, starting from a foliation close to the vertical foliation (in the sense of the cone condition). For completeness, note that the strongly unstable manifold of the point $p \in \mathbb{S}^1 \times [0,1]$ is not uniquely determined; it depends on the choice of the sequence of preimages. The strongly stable manifold is discrete and consists of the preimages of the given point, i.e., of the points whose dynamics "merges" with the dynamics of the given point.

The cone condition corresponding to the dominated splitting is preserved under small perturbations, and hence an arbitrary map $\widetilde{F}$ sufficiently C^2-close to F possesses a central foliation as well.

By virtue of the above-mentioned Gorodetskiĭ result [G06, GI99], the central leaves of any small perturbation $\widetilde{F}$ satisfy the Hölder condition, and the dynamics on the set of central leaves is conjugate to the original dynamics on the base; namely, there exists a *Hölder* map $\psi\colon \mathbb{S}^1 \times [0,1] \to \mathbb{S}^1$ semiconjugating $\widetilde{F}$ with the circle doubling map, and moreover, $\psi(y,x) = \psi(\bar{y},\bar{x})$ if and only if (y,x) and $(\bar{y},\bar{x})$ lie in a common central leaf $\widetilde{F}$.

Likewise, in case (ii), the map F is partially hyperbolic, its central leaves are compact, and hence an arbitrary close diffeomorphism $\widetilde{F}$ has compact central leaves as well. Just as in case (i) (see [GI99]), the dependence of these leaves on the point satisfies the Hölder condition, and the dynamics on the set of these leaves proves to be conjugate to the original map $T\colon \mathbb{T}^2 \to \mathbb{T}^2$ of the base; namely, there exists a Hölder map $\psi\colon \mathbb{T}^2 \times [0,1] \to \mathbb{T}^2$ semiconjugating $\widetilde{F}$ with T for which $\psi(y,x) = \psi(\bar{y},\bar{x})$ if and only if the points (y,x) and $(\bar{y},\bar{x})$ lie on a common central leaf.

Finally, according to Ilyashenko and Negut [IN10]), the Hölder exponent of the dependence of the central leaves on the point (or, which is the same, the Hölder exponent of ψ) tends to 1 as the perturbed map $\tilde{F}$ tends to the original map; this is true in both cases (i) and (ii). Moreover, the Hölder condition holds uniformly; namely, for each $\alpha < 1$, the α-Hölder constant of ψ is uniformly bounded for maps $\widetilde{F}$ in a sufficiently small neighborhood of the original map F.

Thus, in both cases (i) and (ii), the map $\Psi(y,x) = (\psi(y,x), x)$ is a Hölder conjugation of every map $\widetilde{F}$ in a sufficiently small neighborhood of F with some skew product $G(y,x) \to (Ty, \tilde{f}_y(x))$. Moreover, since the leaves satisfy the Hölder condition uniformly, we can take the Hölder exponent α and constant C to be the same for all $\widetilde{F}$ in a sufficiently small neighborhood of F, and the "rectified" skew product G will lie in a small neighborhood of F in the space $\mathcal{H}^2_{(C,\alpha)}$.

For the new skew product in a sufficiently small (in the sense of the metric $d_{\mathcal{H}}$) neighborhood of the original skew product, Theorem 3 guarantees that the Hausdorff dimension of the set

$$E := \{(y,x) \mid G^n(y,x) \not\to A_0 \cup A_1\}$$

is strictly less than the dimension of the phase space and is even separated from the latter by some constant $d_0 < \dim X$.

On the other hand, the Falconer lemma claims that the Hausdorff dimension changes at most by the factor $1/\alpha$ under a Hölder conjugation with exponent α. Set

$$d' := \frac{d_0 + \dim X}{2}, \quad \alpha' = \frac{d_0}{d'}.$$

Then $\alpha' < 1$, and hence the conjugation Ψ satisfies the Hölder condition with exponent at least $\alpha' < 1$ in some neighborhood of F. Thus, for every map $\widetilde{F}$ in this neighborhood, the Hausdorff dimension of the set of points that do not tend to the boundary can be estimated as

$$\begin{aligned} (9) \quad & \dim_H\{(y,x) \mid \widetilde{F}^n(y,x) \not\to A_0 \cup A_1\} \\ & \qquad \le \frac{1}{\alpha'} \cdot \dim_H\{(y,x) \mid G^n(y,x) \not\to A_0 \cup A_1\} \le d_0 \cdot \frac{d'}{d_0} = d' \end{aligned}$$

and hence is strictly less than the dimension of the phase space. This proves conclusion (i) in Theorems 1 and 2.

Next, for each point in $W^s(y_{(0)}, 0)$ (respectively, $W^s(y_{(1)}, 1)$), the measure of the set of points tending to A_0 or A_1 is positive by Corollary 1 and Remark 3.

On the other hand, by condition (U2), both of these stable manifolds are dense in $B \times [0,1]$ (because they are everywhere dense for the skew product G conjugate to $\widetilde{F}$). Thus, each of the attraction basins $B(A_0)$ and $B(A_1)$ is metrically everywhere dense in $B \times [0,1]$, and we arrive at conclusion (ii) in Theorems 1 and 2. □

Proposition 2. *Conditions* (U3) *and* (U3′) *remain simultaneously valid under $d_{\mathcal{H}}$-small perturbations of the skew product.*

Proof. It is easily seen that condition (U3′) is open, as is the expansion condition (6) in condition (U3) for any compact domain (i.e., domain separated from 0 and 1 in the x-coordinate).

Now consider a neighborhood of A_0. (The case of a neighborhood of A_1 is completely similar.) Hadamard's lemma gives $\tilde{f}_y(x) = x h_y(x)$, where h_y is a C^1 map; moreover, by virtue of the explicit integral formula

$$h_y(x) = \int_0^1 (\tilde{f}_y)'(tx)\, dt$$

for h_y, the change in the maps h_y under C^2-small perturbations $\tilde{f}_y$ is C^1-small.

Now let us write out the derivative in the coordinates $z = \Phi(x)$. Since $\Phi'(x) = \frac{1}{x(1-x)}$, we have

$$(g_y)'(z) = \frac{(\tilde{f}_y)'(x)\Phi'(\tilde{f}_y(x))}{\Phi'(x)} = \frac{(xh_y'(x) + h_y(x)) \cdot x(1-x)}{xh_y(x) \cdot (1 - xh_y(x))}$$
$$= \frac{1-x}{1-xh_y(x)} \cdot \left(1 + x \cdot \frac{h_y'(x)}{h_y(x)}\right) = 1 + \left(\frac{h_y'(x)}{h_y(x)} + h_y(x) - 1\right)x + O(x^2).$$

Accordingly, if the coefficient of x is positive, then the derivative g_y' is greater than 1 at the point $z = \Phi(x)$ for sufficiently small x. Moreover, it is easily seen that the length of the interval on which this estimate holds is separated from zero (for C^1-small perturbations h_y corresponding to C^2-small perturbations $\tilde{f}_y$). Finally, a uniform continuity argument shows that the existence of such intervals is guaranteed by the strict inequality for $x = 0$,

$$\forall y \in B \quad \frac{h_y'(0)}{h_y(0)} + h_y(0) - 1 > 0. \tag{10}$$

However, condition (10) is precisely equivalent to the part concerning the derivatives at zero in inequalities (7) in condition (U3′); $h_y(0) = (\tilde{f}_y)'(0)$, $h_y'(0) = \frac{1}{2}(\tilde{f}_y)''(0)$, and hence

$$\frac{h_y'(0)}{h_y(0)} + h_y(0) - 1 > 0 \quad \Leftrightarrow \quad \frac{(\tilde{f}_y)''(0)}{2(\tilde{f}_y)'(0)} - 1 + (\tilde{f}_y)'(0) > 0.$$

In a similar way, one can consider a neighborhood of the point $x = 1$. Thus, if condition (U3′) is satisfied, then condition (U3) is preserved under small perturbations. □

3.4. **Examples.** It remains to give examples of smooth skew products over the circle doubling map and a linear Anosov diffeomorphism of the torus such that conditions (U1), (U2), (U3), and (U3′) are satisfied. Let $f_{a,t}(x)$ be the time t map of the phase flow corresponding to the vector field $v_a(x) = x(1-x)(x-a)\frac{\partial}{\partial x}$, depending on a real parameter a (see Figure 1).

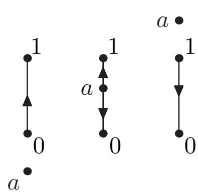

FIGURE 1. The vector field v_a for $a < 0$, $a \in (0,1)$, and $a > 1$.

Proposition 3. *Take some periodic points*

$$y_{(0)} = T^{k_0}(y_{(0)}), \quad y_{(1)} = T^{k_1}(y_{(1)})$$

of the base and a smooth function $a\colon B \to \mathbb{R}$ *such that*

- $a(T^j(y_{(0)})) > 1$, $j = 1, 2, \ldots, k_0$;
- $a(T^j(y_{(1)})) < 0$, $j = 1, 2, \ldots, k_1$;
- $0 < \int_B a(y)\, d\mu(y) < 1$.

Then the smooth skew product

$$F(y,x) = (T(y), f_y(x)), \quad f_y(x) = f_{a(y),t}(x) \tag{11}$$

satisfies conditions (U1), (U2), (U3), *and* (U3′) *for each* $t > 0$.

Moreover, one can make the diffeomorphisms f_y *as close as desired to the identity diffeomorphism by taking a sufficiently small* t.

Proof of Proposition 3. We rewrite the vector field v_a in the coordinate $z = \Phi(x)$ and see that it acquires the form $u_a(z) = (\Phi^{-1}(z) - a)\frac{\partial}{\partial z}$. In particular, since the function $\Phi^{-1}(z) - a$ is monotone increasing, it follows that the derivative of the time t map $g_{a,t}$ corresponding to the field $u(z)$ is everywhere greater than unity for each $t > 0$, whence condition (U3) follows.

Moreover, one can readily verify that the time t map $f_{a,t}$ corresponding to the field v_a (for which $g_{a,t} = \Phi \circ f_{a,t} \circ \Phi^{-1}$) satisfies condition (U3′). Indeed, let $z(\theta)$ and $x(\theta)$ be the solutions of the equations $\dot{x} = v_a(x)$ and $\dot{z} = u_a(z)$ with the initial conditions $x(0) = x_0$ and $z(0) = z_0$ respectively. Then

$$g'_{a,t}(z_0) = \int_0^t (u_a)'(z(\theta))\,d\theta = \int_0^t \frac{\partial(\Phi^{-1}(z) - a)}{\partial z}\bigg|_{z(\theta)}\,d\theta$$
$$= \int_0^t x(\theta)(1 - x(\theta))\,d\theta = t \cdot x_0 + O(x_0^2).$$

On the other hand, the left-hand sides of inequalities (7) in condition (U3′) are the coefficients of x in the expansions of the derivative at the points $x = 0$ and $x = 1$. The preceding computation shows that this coefficient for the maps $f_{a,t}$ at the point $x = 0$ is equal to t, and inequality (7) is satisfied. In a similar way, one can consider the case of $x = 1$.

Next, $f'_{a,t}(0) = e^{-at}$ and $f'_{a,t}(1) = e^{-t(1-a)}$, whence it follows that the central Lyapunov exponents for the circles $x = 0$ and $x = 1$ are

$$\lambda_0 = -t \cdot \int_B a(y)\,dy, \quad \lambda_1 = -t \cdot \int_B (1 - a(y))\,dy,$$

respectively. Hence condition (U1) is satisfied owing to the constraint

$$0 < \int_B a(y)\,dy < 1.$$

Furthermore, the vector field v_a is directed upwards everywhere on the interval $(0, 1)$ for $a < 0$ and downwards everywhere for $a > 1$ (see Figure 1). This, together with the conditions imposed on $a(T^j(y_{(0)}))$ and $a(T^j(y_{(1)}))$, implies that condition (U2) holds.

Thus, the skew product $F(y, x) = (T(y), f_{a(y),t}(x))$ satisfies all conditions (U1), (U2), (U3), and (U3′).

Finally, by letting t tend to zero, we obtain a skew product whose leaf maps are arbitrarily close to the identity map. □

Since the corresponding skew product is partially hyperbolic provided that the maps f_y are sufficiently close to the identity map, we see that this example, together with Proposition 1, completes the derivation of Theorems 1 and 2 from Theorems 3 and 4.

The remaining part of the paper deals with the proof of Theorems 3 and 4.

4. Tools: large deviations and the special ergodic theorem

The main tools in the proofs of Theorems 3 and 4 are the so-called *special ergodic theorem* and the probabilistic and dynamical versions of the *large deviation theorem*.

Namely, let a smooth dynamical system $f\colon X \to X$ with a Sinai–Ruelle–Bowen measure m be given. For a Lebesgue almost every point $x \in X$ and every continuous function

ϕ, one has

$$\frac{1}{n}\sum_{j=0}^{n-1}\phi(f^j(x)) \to \int_X \phi\, dm =: \bar{\phi}.$$

We say that the *special ergodic theorem* holds for this system if the set of points at which the time averages "strongly deviate" from the spatial averages is not only of zero Lebesgue measure but also has Hausdorff dimension smaller than the dimension of the phase space,

$$\forall\phi\in C(X)\ \ \forall\varepsilon>0 \quad \dim_H\left\{x\in X \;\middle|\; \limsup_{n\to\infty}\left|\frac{1}{n}\sum_{j=0}^{n-1}\phi(f^j(x))-\bar{\phi}\right|\geq\varepsilon\right\}<\dim X.$$

To the best of the authors' knowledge, the idea of the special ergodic theorem is due to Ilyashenko. (In a similar situation, the Hausdorff dimension of the set of points with large deviations of the time averages was studied by Gurevich and Tempel′man [GT02].) In particular, the special ergodic theorem holds for the circle doubling map [IKS08] and for linear Anosov diffeomorphisms of the two-dimensional torus [S10].

Moreover, one can prove that the special ergodic theorem holds for every dynamical system for which the (dynamical) large deviation theorem holds. The proof of this is beyond the framework of the present paper (and is the subject of [KR]). Recall that the large deviation theorem in probability theory permits one to estimate the probability of the event that the arithmetic mean of independent identically distributed random variables strongly deviates from their expectation (e.g., see Varadhan's survey [Var08]).

Theorem (On large deviations). *Let $\xi_1,\xi_2,\dots,\xi_n,\dots$ be a sequence of independent identically distributed random variables with $E\xi=0$ and with finite exponential moment, $Ee^{a|\xi|}<\infty$ for some small $a>0$. Then for each $\alpha>0$ there exist $C,\delta>0$ such that*

$$P\Big(\frac{\xi_1+\dots+\xi_n}{n}>\alpha\Big)<Ce^{-\delta n} \tag{12}$$

for each $n>0$. In other words, the probability that the arithmetic mean goes past a given level $\alpha>0$ exponentially decays with increasing n.

This statement of the problem can naturally be transferred to dynamical systems. Namely, one says that the (dynamical) large deviation theorem holds for a smooth dynamical system $f\colon X\to X$ with an SRB-measure m if

$$\forall\phi\in C(X)\ \ \forall\varepsilon>0\ \ \exists C,\beta>0\colon \forall n\in\mathbb{N}$$

$$\operatorname{Leb}\left\{x\colon \left|\frac{1}{n}\sum_{j=0}^{n-1}\phi(f^j(x))-\int_X\phi\, dm\right|\geq\varepsilon\right\}<Ce^{-\beta n}. \tag{13}$$

The question as to whether the large deviation theorem holds for various dynamical systems was studied by Lai-Sang Young, who proved (see [Y90, Theorem 2(2) and item (ii) in the subsequent discussion] and [Y03, Sec. 3.3]) that, in particular, the estimate (4) holds for hyperbolic dynamical systems.

Theorem (Lai-Sang Young, [Y90]). *The large deviation principle* (4) *holds for hyperbolic dynamical systems.*

Thus, the special ergodic theorem holds for such systems as well (see [KR]).

We finish this section by noting that the estimate of the Hausdorff dimension of the set of points with given large deviation proves to be uniform with respect to small perturbations of the function ϕ to be averaged.

5. Estimate of the Hausdorff dimension of the exceptional set

In this section, we prove Theorem 3.

5.1. **Two types of exceptional points.** For each $\varepsilon > 0$, set

$$\varphi_0^\varepsilon(y) := \max_{0\le x\le\varepsilon} \frac{\tilde f_y(x)}{x}, \quad \varphi_1^\varepsilon(y) := \max_{1-\varepsilon\le x\le 1} \frac{1-\tilde f_y(x)}{1-x}. \tag{14}$$

Then we can estimate how the distance to the boundary component A_0 (or A_1) from a point (y,x) sufficiently close to it changes per iteration of G: for any $x \in [0,\varepsilon]$ and $y \in B$, we have

$$\tilde f_y(x) \le \varphi_0^\varepsilon(y)\cdot x.$$

By using the version of Hadamard's lemma uniform in y, for $\varepsilon \to 0$ we readily see that

$$\varphi_0^\varepsilon \rightrightarrows (\tilde f_y)'(0), \quad \varphi_1^\varepsilon \rightrightarrows (\tilde f_y)'(1). \tag{15}$$

Hence

$$\tilde\lambda_0 = \int_B \log\varphi_0^{\varepsilon_0}(y)\,dy < 0, \quad \tilde\lambda_1 = \int_B \log\varphi_1^{\varepsilon_0}(y)\,dy < 0 \tag{16}$$

for each sufficiently small $\varepsilon_0 > 0$.

Take and fix an $\varepsilon_0 > 0$ such that, in addition to condition (16), no point can "jump" over the strip $\Pi := B \times [\varepsilon_0, 1-\varepsilon_0]$ in one iteration,

$$\forall y \in B \quad \tilde f_y(\varepsilon_0) < 1-\varepsilon_0, \quad \tilde f_y(1-\varepsilon_0) > \varepsilon_0.$$

Note that the points tending to neither A_0 nor A_1 are divided into two types. Namely, a point either resides in one of the strips $\Pi_0 = B\times[0,\varepsilon_0]$ and $\Pi_1 = B\times[1-\varepsilon_0,1]$ starting from some iteration but nevertheless does not tend to A_0 or A_1 respectively, or it can visit the strip Π infinitely many times. We refer to these points as exceptional points of the first and second type, respectively (subdividing the points of the first type into two subtypes according to the strip visited) and denote the corresponding sets by

$$\begin{aligned}
E_1^0 &= \{(y,x) \mid G^n(y,x) \not\to A_0,\ \exists n_0\colon \forall n \ge n_0\ \ G^n(y,x) \in \Pi_0\},\\
E_1^1 &= \{(y,x) \mid G^n(y,x) \not\to A_1,\ \exists n_0\colon \forall n \ge n_0\ \ G^n(y,x) \in \Pi_1\},\\
E_2 &= \{(y,x) \mid \forall n_0\ \exists n > n_0\colon G^n(y,x) \in \Pi\}.
\end{aligned}$$

5.2. **Estimate of the Hausdorff dimension of the set of exceptional points belonging to the first type.**

Proposition 4. *Under the assumptions of Theorem* 3*, the Hausdorff dimension of the set of exceptional points of the first type* (*that is, of the set* $E_1 = E_1^0 \cup E_1^1$) *is strictly less than the dimension of the phase space and, moreover, is separated from the latter uniformly in a sufficiently small neighborhood of the skew product* (5) *in the space* $\mathcal{H}^2_{(C,\alpha)}$ *of Hölder skew products.*

Proof. Let us prove the claim for the points in E_1^0. (The argument for E_1^1 is similar.)

Let Y be the projection of E_1^0 onto the base B; since $E_1^0 \subset Y\times[0,1]$, we see that, to prove Proposition 4, it suffices to show that the Hausdorff dimension $\dim_H(Y)$ is smaller than the dimension of the base B (and write out the corresponding uniform estimates in a small neighborhood of the original skew product).

It is easily seen that

$$\begin{aligned}
Y &= \bigcup_{n_0=0}^{\infty} T^{-n_0}(Y_0),\\
Y_0 &= \bigl\{y \mid \exists x\colon (\forall n \ge 0\ G^n(y,x) \in \Pi_0)\ \&\ (G^n(y,x) \not\to A_0 \text{ as } n\to\infty)\bigr\}.
\end{aligned} \tag{17}$$

On the other hand, the map $T\colon B \to B$ is smooth, and passage to the corresponding preimages does not change the Hausdorff dimension. A countable union of sets does not change the Hausdorff dimension either. Hence $\dim_H Y = \dim_H Y_0$.

To estimate $\dim_H Y_0$, we use the special ergodic theorem. Note that if the iterates (y_i, x_i) of some initial point (y, x) remain in the strip Π_0, then

$$x_{j+1} \le x_j \cdot \varphi_0^{\varepsilon_0}(y_j) \tag{18}$$

by the definition of the function $\varphi_0^{\varepsilon_0}$. Consequently,

$$\log x_n \le \log x_0 + \sum_{j=0}^{n-1} \log \varphi_0^{\varepsilon_0}(y_j). \tag{19}$$

By applying the special ergodic theorem for the circle doubling map or a linear Anosov diffeomorphism to the function $\psi := \log \varphi_0^{\varepsilon_0}$, we find that the Hausdorff dimension of the set

$$\widetilde{Y} := \left\{ y \,\middle|\, \varlimsup_{n\to\infty} \frac{\sum_{j=0}^{n-1} \psi(y)}{n} \ge -\frac{|\tilde{\lambda}_0|}{2} \right\} \tag{20}$$

is smaller than $\dim B$. On the other hand,

$$\sum_{j=0}^{n-1} \psi(y_j) \xrightarrow[n\to\infty]{} -\infty$$

for $y \notin \widetilde{Y}$, which excludes the existence of a sequence x_n that satisfies (19) and does not tend to 0. Hence $Y_0 \subset \widetilde{Y}$. It follows that $\dim_H(Y_0) < \dim B$ and $\dim_H(M_1) < \dim X$.

It remains to justify the existence of an estimate uniform in a small neighborhood of the original system. Note that the special ergodic theorem is the only tool used in the preceding argument. A small change in the system in the space $\mathcal{H}^2_{(C,\alpha)}$ of Hölder skew products results in a small change in the function ψ in $C^0(B)$. Hence the uniformness of the estimate for the Hausdorff dimension follows from the uniformness of the estimate in the special ergodic theorem (see [IKS08] and [S10]).

The proof of Proposition 4 is complete. □

5.3. **Estimate of the Hausdorff dimension of the set of exceptional points belonging to the second type.** To estimate the Hausdorff dimension of the set of exceptional points of the second type, consider the sequence of preimages

$$G^{-n}(\Pi) = \{(y, x) \mid G^n(y, x) \in B \times [\varepsilon_0, 1 - \varepsilon_0]\}. \tag{21}$$

We cover each such preimage by neighborhoods and control the size and number of these neighborhoods. More specifically, we prove the following lemma.

Lemma 1 (The main lemma)**.** *There exist constants $d_2 < \dim X$ and $C > 0$ such that, for each n, the full preimage $G^{-n}(\Pi)$ can be covered by at most $C \cdot \lambda_u^{d_2 n}$ neighborhoods of radius λ_u^{-n}.*

This lemma readily implies that

Proposition 5. $\dim_H(E_2) \le d_2$.

Derivation of Proposition 5 *from Lemma* 1. By the definition of E_2,

$$E_2 = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} G^{-n}(\Pi). \tag{22}$$

For each N, the union over $n \geq N$ of covers of the sets $G^{-n}(\Pi)$ by neighborhoods of radius λ_u^{-n} indicated in the main lemma is a cover of E_2. On the other hand, for each $d > d_2$, the d-dimensional volume of such a cover can be estimated as

$$\sum_{n=N}^{\infty} C \cdot \lambda_u^{d_2 n} \cdot (\lambda_u^{-n})^d = \sum_{n=N}^{\infty} C \cdot \lambda_u^{-(d-d_2)n} = \lambda_u^{-(d-d_2)N} \cdot \frac{C}{1-\lambda_u^{-(d-d_2)}}. \tag{23}$$

In particular, the right-hand side tends to zero as $N \to \infty$. Thus, we have proved that the set E_2 has zero d-dimensional volume for all $d > d_2$. Consequently,

$$\dim_H(E_2) \leq d_2 < \dim X. \qquad \square$$

The conclusion of Proposition 4, combined with that of Proposition 5, gives the conclusion of Theorem 3, and so, to prove the theorem, it remains to prove Lemma 1.

We continue the proof for the case of a skew product over the circle doubling map; the case of a skew product over an Anosov diffeomorphism is similar. We need the following result due to Gorodetskiĭ.

Lemma (On the predictability of trajectories)**.** *Let $G\colon Y \times X \to Y \times X$, $G(y,x) = (S(y), f_y(x))$, be a Hölder skew product over an expanding map $S\colon Y \to Y$ with expansion ratio $\lambda > 1$. That is,*

$$\rho(S(y_1), S(y_2)) \geq \lambda \rho(y_1, y_2), \tag{24}$$

$$d_{C^0}(f_{y_1}, f_{y_2}) \leq C_1 \cdot \rho(y_1, y_2)^h \tag{25}$$

for $\rho(y_1, y_2) < \rho_0$. Suppose also that the leaf maps are uniformly Lipschitz,

$$\forall y \ \ \mathrm{Lip}(f_y) \leq L,$$

and the inequality

$$\lambda^h > L$$

(*an analog of the partial hyperbolicity and dominated splitting conditions for Hölder skew products*) *holds. Then there exists a constant $\tilde{C}$ such that, for two arbitrary trajectory segments of length n with initial conditions (y_1, x) and (y_2, x) satisfying*

$$\rho(S^k(y_1), S^k(y_2)) \leq \rho_0 \qquad \forall k = 0, 1, \dots, n, \tag{26}$$

one has the following estimate of the distance along the fiber between the images $(\tilde{y}_1, \tilde{x}_1) = G^n(y_1, x)$ and $(\tilde{y}_2, \tilde{x}_2) = G^n(y_2, x)$:

$$\mathrm{dist}(\tilde{x}_1, \tilde{x}_2) \leq \tilde{C} \cdot \rho(\tilde{y}_1, \tilde{y}_2)^h. \tag{27}$$

In other words, under certain conditions on the system, the net error in the fiber coordinate of an iterate of a point whose initial base coordinate is known approximately is comparable to the error introduced at the last iteration.

A slightly different statement of this lemma can be found in [GI00, Lemma 3.1] and [G01], but we recall the proof to make the exposition self-contained.

Proof. Set $y_j^{(1)} = S^j(y_1)$, $y_j^{(2)} = S^j(y_2)$. Let us estimate the distance between

$$\tilde{x}_1 = (f_{y_{n-1}^{(1)}} \circ \cdots \circ f_{y_1^{(1)}} \circ f_{y_0^{(1)}})(x)$$

and

$$\tilde{x}_2 = (f_{y_{n-1}^{(2)}} \circ \cdots \circ f_{y_1^{(2)}} \circ f_{y_0^{(2)}})(x).$$

To this end, we join them by the sequence of points $t_0 = \tilde{x}_1$, $t_n = \tilde{x}_2$, and

$$t_j = (f_{y_{n-1}^{(1)}} \circ \cdots \circ f_{y_j^{(1)}} \circ f_{y_{j-1}^{(2)}} \circ \cdots \circ f_{y_0^{(2)}})(x).$$

Note that the points t_j and t_{j+1} differ only in one of the maps, $f_{y_j^{(2)}}$ instead of $f_{y_j^{(1)}}$, applied in the middle of the "chain". Hence

$$\operatorname{dist}(t_j,t_{j+1})=\operatorname{dist}((f_{y_{n-1}^{(1)}}\circ\dots\circ f_{y_{j+1}^{(1)}})(f_{y_j^{(1)}}(p_j)),(f_{y_{n-1}^{(1)}}\circ\dots\circ f_{y_{j+1}^{(1)}})(f_{y_j^{(2)}}(p_j))),$$

where

$$p_j=(f_{y_{j-1}^{(2)}}\circ\dots\circ f_{y_0^{(2)}})(x).$$

By the Lipschitz condition along the fiber and the Hölder condition along the base, we have

$$\begin{aligned}\operatorname{dist}((f_{y_{n-1}^{(1)}}\circ\dots\circ f_{y_{j+1}^{(1)}})(f_{y_j^{(1)}}(p_j)),(f_{y_{n-1}^{(1)}}\circ\dots\circ f_{y_{j+1}^{(1)}})(f_{y_j^{(2)}}(p_j)))\\ \le L^{n-j-1}\cdot\operatorname{dist}(f_{y_j^{(1)}}(p_j),f_{y_j^{(2)}}(p_j))\le L^{n-j-1}\cdot C_1\rho(y_j^{(1)},y_j^{(2)})^h.\end{aligned}$$

Finally, the expansion condition implies that $\rho(y_j^{(1)},y_j^{(2)})\le\rho(y_n^{(1)},y_n^{(2)})/\lambda^{n-j}$.

By summing the resulting estimates, we obtain

$$\begin{aligned}\operatorname{dist}(x_1,x_2)&\le\sum_{j=0}^{n-1}\rho(t_j,t_{j+1})\le\sum_{j=0}^{n-1}C_1\frac{L^{n-j-1}\cdot\rho(y_n^{(1)},y_n^{(2)})^h}{(\lambda^{n-j})^h}\\ &\le\frac{C_1}{L}\cdot\rho(y_n^{(1)},y_n^{(2)})^h\cdot\sum_{i=0}^{\infty}\left(\frac{L}{\lambda^h}\right)^i=\tilde C\rho(y_n^{(1)},y_n^{(2)})^h.\qquad\square\end{aligned}$$

We apply this lemma to our skew product and choose some distance δ_0 so as to ensure that $\tilde C\cdot\delta_0^h\le\varepsilon_0/2$, where $\tilde C$ is the same constant as in (27). Next, we divide the circle $\mathbb{S}^1$ into intervals $I_1,\dots,I_{[1/\delta_0]+1}$ of length $\le\delta_0$. Note that the following proposition holds.

Proposition 6. *Let an interval $\tilde I\subset\mathbb{S}^1$ be one of the connected components of the preimage $T^{-n}(I)$, where I is one of the intervals I_j, and let an interval $J\subset[0,1]$ be the vertical component of $G^{-n}\Big(I\times\big[\frac{\varepsilon_0}{2},1-\frac{\varepsilon_0}{2}\big]\Big)$ over some point $y\in\tilde I$,*

$$G^{-n}\Big(I\times\Big[\frac{\varepsilon_0}{2},1-\frac{\varepsilon_0}{2}\Big]\Big)\cap(\{y\}\times[0,1])=\{y\}\times J.\tag{28}$$

Then the rectangle $\tilde I\times J$ covers the connected component $G^{-n}(I\times[\varepsilon_0,1-\varepsilon_0])$ of the preimage in question,

$$G^{-n}(I\times[\varepsilon_0,1-\varepsilon_0])\cap(\tilde I\times[0,1])\subset\tilde I\times J.\tag{29}$$

Proof. This proposition readily follows from the predictability lemma applied to the upper and lower boundaries of $\tilde I\times J$. Indeed, let $J=[x_{\min},x_{\max}]$; then, for each point $y\in\tilde I$, the x-coordinate of the image of $G^n(y,x_{\min})$ does not exceed the sum of the x-coordinate of $G^n(y,x_{\min})$ with $\tilde C\delta_0^h$. But the x-coordinate of $G^n(y,x_{\min})$ is $\varepsilon_0/2$, and $\tilde C\delta_0^h\le\varepsilon_0/2$; hence the point $G^n(y,x_{\min})$ lies below the level ε_0 along the x-axis. Likewise, the image of the segment $G^n(\tilde I\times\{x_{\max}\})$ lies entirely above the level $1-\varepsilon_0$ along the x-axis, which proves the desired inclusion. $\square$

Thus, we have divided the circle into intervals $I_1,\dots,I_{[1/\delta_0]+1}$ of length $\le\delta_0$. Take and fix a point $y_l\in I_l$ in each of these intervals. Then the set $G^{-n}(\mathbb{S}^1\times[\varepsilon_0,1-\varepsilon_0])$ is covered by the union of rectangles $\tilde I_{k,l}\times J_{k,l}$, where $T^{-n}(I_l)=\bigcup_{k=1}^{2^n}\tilde I_{k,l}$ and $J_{k,l}$ is the preimage of the interval $[\varepsilon_0/2,1-\varepsilon_0/2]$ over the corresponding point $\tilde y_{k,l}$, which is the preimage of y_l in $\tilde I_{k,l}$.

Note that the length of any of the rectangles $\tilde I_{k,l}\times J_{k,l}$ along the y-axis does not exceed 2^{-n}. To complete the proof of the main lemma, we divide these rectangles into two classes.

The rectangles in the first class correspond to the points $\tilde{y}_{k,l}$ for which the corresponding interval meets the strip $\Pi' := \mathbb{S}^1 \times [\varepsilon_0/2, 1-\varepsilon_0/2]$ less than Cn times in n iterations, where C is some constant to be chosen later. It turns out (see Proposition A below) that, since A_0 and A_1 are repulsive with respect to taking the preimages of the circle, the fraction of such rectangles is exponentially small for sufficiently small C; namely, there are less than $2^{\beta n}$ such rectangles, where $\beta < 1$.

The intervals corresponding to rectangles of the second class meet the strip Π' at least Cn times in the course of iterations.

Recall that, by assumption (U3), the derivatives of the maps $\tilde{f}_y$ in the coordinate $z = \Phi(x)$ are larger than 1. Hence the preimages for these maps are (nonstrictly) contracting in the same coordinate. Moreover, the contraction rate is separated from 1 in every domain separated from A_0 and A_1. Hence (see Proposition B below) the rectangles of the second class have vertical size not exceeding $2^{-\gamma n}$ for some $\gamma > 0$.

The rectangles of the first class can be replaced by $2^n \cdot 2^{\beta n} = 2^{(1+\beta)n}$ squares of size $2^{-n} \times 2^{-n}$. (In the fiber, we estimate their height by the entire interval $[0,1]$.) The rectangles of the second class can be covered by $2^{(1-\gamma)n} \cdot 2^n = 2^{(2-\gamma)n}$ such squares. Since $\gamma > 0$ and $\beta < 1$, we can set

$$d_2 := \max(1+\beta, 2-\gamma) < 2.$$

This argument proves the main lemma once we prove the following propositions.

Fix a point $(y_0, x_0) \in \Pi'$. Consider the Markov process $(y_j, x_j)_{j=0}^{\infty}$ of taking G-preimages, $(y_{j+1}, x_{j+1}) \in G^{-1}(y_j, x_j)$, in which one of the two preimages is chosen with probability $1/2$ at each step.

Propostion A. *There exist constants $0 < C < 1$, $\beta > 0$, and $C_2 > 0$ such that, for each point (y_0, x_0) in the strip Π', the probability of the event that the sequence of preimages (y_j, x_j), $j = 1, \ldots, n$, for n iterations contains at least Cn preimages lying in the same strip is not less than $1 - C_2 2^{-\beta n}$.*

Propostion B. *Let $(y_0, x_0) \in \Pi'$, and let the sequence of preimages $(y_j, x_j)_{j=0}^{n-1}$ contain at least m points lying in the same strip Π'. Then the length of the interval*

$$J = \tilde{f}_{y_n}^{-1} \circ \cdots \circ \tilde{f}_{y_1}^{-1}([\varepsilon_0/2, 1-\varepsilon_0/2]),$$

which is the corresponding nth preimage of $[\varepsilon_0/2, 1-\varepsilon_0/2]$ over the point y_0, can be estimated by $|J| \le C_4 2^{-\gamma m}$, where C_4 and $\gamma > 0$ are some constants.

Proof of Proposition B. Let us proceed to the coordinate $z = \Phi(x)$. By condition (U3), the maps $\tilde{f}_y$ are (nonstrictly) expanding in this coordinate,

$$\forall y \in \mathbb{S}^1, z \in \mathbb{R} \quad (g_y)'(z) > 1,$$

where $g_y(z) = \Phi \circ \tilde{f}_y \circ \Phi^{-1}(z)$. Hence, when taking the preimages, the length of an interval in the z-coordinate, which is originally equal to $l_0 := \Phi(1-\varepsilon_0/2) - \Phi(\varepsilon_0/2)$, does not increase. Hence, whenever the preimage of the interval $[\varepsilon_0/2, 1-\varepsilon_0/2]$ meets the domain Π', the interval itself is contained in the strip

$$\Pi'' := \{(y,x) \mid x \in \Phi^{-1}(U_{l_0}([\Phi(\varepsilon_0/2), \Phi(1-\varepsilon_0/2)]))\}.$$

On the other hand, since this strip is a compact set on which the derivatives of the inverse maps in the z-coordinate are strictly less than 1, it follows that the corresponding iterations are contracting with contraction ratio $\lambda_z < 1$ separated from 1. Hence the length of the image $\Phi(J)$ in the z-coordinate can be estimated as

$$|\Phi(J)| \le l_0 \cdot \lambda_z^m = l_0 \cdot 2^{-\gamma m},$$

where $\gamma := |\log_2 \lambda_z| > 0$. Finally, since $(\Phi^{-1})'(z) \le 1$ for all z, we obtain the desired estimate of the length of the image of the interval J in the x-coordinate. □

The proof of Proposition A in the next subsection will complete the proof of the main lemma.

5.4. **Preimages landing in the middle strip (Proof of Proposition A).** First, let us prove the following auxiliary assertion.

Propostion C. *For some constants $C_3 > 0$ and $\beta_1 > 0$, the probability of the event that the entire sequence of preimages $(y_1, x_1), \dots, (y_n, x_n)$ lies outside the strip Π' does not exceed $C_3 2^{-\beta_1 n}$. This estimate is uniform with respect to the choice of the initial point $(y_0, x_0) \in \Pi'$.*

Proof. This proposition is actually a consequence of an auxiliary estimate obtained in the special ergodic theorem. Indeed, let $x_1, \dots, x_n < \varepsilon_0/2$. (The case in which the sequence goes beyond the strip upwards can be considered in a similar way.) Then, as we have already seen (see the proof of Proposition 4),

$$\log x_n \ge \log x_1 + \sum_{i=1}^{n-1} -\psi(y_i). \tag{30}$$

On the other hand, $\log x_1 = \log f^{-1}_{y(1)}(x_0) \ge \log \varepsilon_0/2 - \log L$, where L is the Lipschitz constant. Hence if $x_n \le \varepsilon_0/2$, then

$$\sum_{j=1}^{n-1} (-\psi(y_i)) \le \log L. \tag{31}$$

Now note that the function ψ satisfies the Hölder condition $|\psi(y') - \psi(y'')| \le c|y' - y''|^\alpha$. Hence an argument similar to the proof of the predictability lemma shows that, for each preimage y_n of y for which (31) holds, the sum $\sum_{j=0}^{n-1}(-\psi(T^j(q)))$ does not exceed some constant C_ψ independent of n for all points $q \in [y_n - 2^{-n}, y_n + 2^{-n}]$. Indeed,

$$\begin{aligned}\left|\sum_{j=0}^{n-1} (-\psi(T^j(q))) - \sum_{j=0}^{n-1} (-\psi(T^j(y_n)))\right| \\ \le \sum_{j=0}^{n-1} \left|\psi(T^j(q)) - \psi(T^j(y_n))\right| \le \sum_{j=0}^{n} c(2^{-(n-j)})^\alpha \le \frac{c}{1-2^{-\alpha}}.\end{aligned}$$

But the possible preimages of y in n iterations divide the circle into the 2^n intervals considered above, and hence the probability of the event that the sum (31) for the sequence of preimages is less than or equal to $\log L$ does not exceed the Lebesgue measure of the set of points q for which $\sum_{j=0}^{n-1} \psi(T^j(q)) \le C_\psi$.

On the other hand, $\int \psi(y)\, dy > 0$; consequently, by virtue of the exponential decay of measures of large deviations (see [Y90] and [IKS08]), the above-mentioned measure is exponentially small, and hence so is the desired probability. □

Proof of Proposition A. For the Markov process in question, we define (possibly infinite) Markov times $\tau_1, \tau_2, \dots$ as follows: τ_j is the time at which the preimage (y_m, x_m) hits the strip Π' for the jth time if this happens at all and ∞ otherwise. Then it follows from Proposition C that each τ_j is finite almost surely, and moreover, for the conditional probability, one has the estimate

$$P(\tau_j - \tau_{j-1} \ge k \mid \tau_{j-1} = s, (\omega_1, \dots, \omega_s) = (\omega_1^0, \dots, \omega_s^0)) \le C_3 \cdot 2^{-\beta_1 k} \tag{32}$$

for any s and any choice of the preimages $\omega_1^0, \dots, \omega_s^0$ such that $\tau_{j-1}(\omega_1^0, \dots, \omega_s^0) = s$. (Recall that $\omega_i^0 = G^{-i}(y_0, x_0)$.)

Consider a distribution ξ taking positive integer values and defined by the relation $P(\xi \geq k) = \min\{1, C_3 2^{-\beta_1 k}\}$. Then (expanding the probability space $\Omega_0 = \{0,1\}^{\mathbb{N}}$ by multiplication by $[0,1]^{\mathbb{N}}$ to ensure the absence of atoms) one can find independent random variables $\xi_1, \xi_2, \dots$ with the same distribution as ξ satisfying $\xi_j \geq \tau_j - \tau_{j-1}$. (For example, see the proof of Lemma 2.3 in [DK07].) Note that all ξ_i are defined on the same probability space as the Markov process itself and that the distribution ξ is independent of the initial point $(y, x) \in \Pi'$.

To prove the proposition, we should estimate the probability of the event that, in n iterations of the preimage, there are less than Cn hits into the strip Π'. In terms of our process, this is $P(\tau_{[Cn]} \geq n)$. But

$$\tau_{[Cn]} = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_{[Cn]} - \tau_{[Cn]-1}) \leq \xi_1 + \cdots + \xi_{[Cn]}.$$

Hence

$$P(\tau_{[Cn]} \geq n) \leq P(\xi_1 + \cdots + \xi_{[Cn]} \geq n). \tag{33}$$

The last probability can be estimated with the use of the large deviation theorem. Take $C := \frac{1}{E\xi+1}$. Then

$$P(\xi_1 + \cdots + \xi_{[Cn]} \geq n) \leq P(\xi_1 + \cdots + \xi_{[Cn]} \geq [Cn] \cdot (E\xi + 1)) \leq C_2 \cdot \exp(-\beta n) \tag{34}$$

for some $C_2, \beta > 0$. Hence $P(\tau_{[Cn]} \geq n) \leq C_2 \cdot \exp(-\beta n)$, which proves the desired exponential decay. $\square$

The proof of this proposition completes the proof of the main lemma and hence of Theorem 3 for the case of a skew product over the circle doubling map.

The above argument can be transferred with small modifications to the case of dynamics over a linear diffeomorphism of the two-dimensional torus. The modifications mainly amount to the passage to Markov encoding; in particular, the process of random choice of the preimage, which was used in the argument for the circle doubling map, is replaced by a Markov refinement-of-the-past process.

Namely, for a linear Anosov diffeomorphism $M\colon \mathbb{T}^2 \to \mathbb{T}^2$ there exists a Markov partition $Q = \{Q_j\}_{j=1}^k$ into parallelograms whose sides are parallel to the stable and unstable directions, respectively. By iterating and refining this partition, one can ensure that its diameter be arbitrarily small. In particular, one can ensure that the integral of the step function $\tilde{\psi}$ equal to $\sup_{Q_j} \psi = \sup_{Q_j} \log \varphi_0^{\varepsilon_0}$ everywhere on the element Q_j with respect to Lebesgue measure be negative (and that the same condition be true for the estimate of repulsion from the level $x = 1$) and that the estimate provided by an analog of the predictability lemma (see below) be not greater than $\varepsilon_0/2$. Take and fix such a partition. Let $\pi\colon \mathbb{T}^2 \to \Sigma_k := \{1, \dots, k\}^{\mathbb{Z}}$ be the corresponding encoding map that takes each point $y \in \mathbb{T}^2$ to its fate $\omega = (\omega_n) \in \Sigma_k$ determined by the rule

$$\forall y \in \mathbb{T}^2,\ \forall n \in \mathbb{Z}\ \ T^n(y) \in Q_{\omega_n}.$$

Consider the partition of $\mathbb{T}^2$ corresponding to the iterations from the $(-n)$th to nth, $Q^{(n)} = \bigwedge_{j=-n}^{n} T^j(Q)$. The elements of these partitions are parallelograms with sides of the order of λ_u^{-n}; the map π takes them to the cylinders

$$C_{w_{-n}\dots w_n} = \{\omega \mid \omega_i = w_i\ \forall i = -n, \dots, n\}.$$

The predictability lemma for the case of hyperbolic maps also applies to such cylinders (see [GI00, Lemma 3.1]) and gives the same conclusion: if a point y is known to belong to a given cylinder $C_{w_{-n}\dots w_n}$, then the error in the computation of the fiber image $G^n(y, x)$ is comparable to the error introduced at the last iteration. Thus, one can make the estimate for this error arbitrarily small (in particular, smaller than $\varepsilon_0/2$) by choosing a sufficiently fine partition Q.

Then, by analogy with Proposition A for the case of dynamics over the circle doubling map, we intend to prove that, for all points of $\mathbb{T}^2$ except for a fraction of these points exponentially small in n, the corresponding vertical intervals of the preimage $G^{-n}(\Pi')$ in the course of computation of n iterations of the preimage meet Π at least Cn times and hence (by the word-for-word generalization of Proposition B) are exponentially small.

Then an application of Proposition 6 (which can be transferred to this case without any alterations as well) permits one to conclude that the preimage $G^{-n}(\Pi)$ is covered by the desired family of neighborhoods. Indeed, the elements R of the refinement $Q^{(n)}$ can be divided into two classes, those for which there exists a $y \in R$ such that the fiber interval $G^{-n}(\Pi') \cap (\{y\} \times [0,1])$ is exponentially small and those for which there is no such interval. The preimage over the elements of the second type is simply covered by the corresponding parallelepiped $R \times [0,1]$ subdivided into $\text{const} \cdot \lambda_u^n$ cubes of edge length λ_u^{-n}; the preimages over the elements of the first type can be estimated with the help of Proposition 6 (which can be transferred to this case word for word).

Thus, it remains to show that, for all points $y \in \mathbb{T}^2$ except for a fraction exponentially small in n, the interval of the fiber of $G^{-n}(\Pi')$ over y in the course of the computation of n interactions of the preimage meets the strip Π at least Cn times, where C is some constant. Since Lebesgue measure is T-invariant, one can instead estimate the fraction of the images $y' = T^n(y)$ corresponding to such points y (the interval of Π' over which is the starting interval in the sequence of n passages to the preimage).

Now let us use Markov encoding; the encoding map π takes the Lebesgue measure μ to the Markov measure $\nu = \pi_*\mu$ on Σ_k with stationary distribution $p_j = \mu(Q_j)$ and with some transition probability matrix (p_{ij}). The choice of a random point with respect to such a measure splits into the successive choice of a "future" $\{\omega_j\}_{j=0}^{\infty}$ (with initial distribution p_j and transition probability matrix (p_{ij})) followed by the construction of the sequence $\omega_{-1}, \omega_{-2}, \dots$ as a Markov chain with transition matrix $(q_{ij}) = (p_j p_{ji}/p_i)$ in reverse time.

Thus, it suffices to prove that, for any given future sequence $\{\omega_j\}_{j=0}^{\infty}$, the conditional probability of the event that the fiber of Π' over a point y' with this future meets Π less than Cn times in the first n iterations is exponentially small.

Now we can use an argument similar to Proposition C. Namely, when "ascribing the future" (that is, going along the Markov chain ω_{-1}, $\omega_{-2}, \dots$), the corresponding preimages are repelled by the boundaries $x = 0$ and $x = 1$, and the effect of the corresponding iterations can be estimated from below in the logarithmic scale by sums of step functions $(-\tilde{\psi})$ and a similar function for the boundary component $x = 1$; these functions have positive means. By applying the large deviation theorem for Markov chains, we obtain an estimate, similar to Proposition C, stating the exponential decay of the probability of a large interval between two subsequent hits. Finally, an argument similar to the estimate (34), already carried out for Proposition A in the case of dynamics over the circle, completes the proof for the case of a skew product over a linear Anosov diffeomorphism of $\mathbb{T}^2$.

The proof of Proposition A and hence of Theorem 3 is complete.

6. Proof of the strengthened metric density theorem

This section deals with the proof of Theorem 4. Throughout this section, a neighborhood is understood as a "horizontal neighborhood" of the form $U \times \{x\}$, where U is some neighborhood in the base B, and the ball $U_r(p)$ is understood as the ball of radius r centered at p in the base B.

Let a skew product $G(y,x) = (T(y), \tilde{f}_y(x))$ satisfy conditions (U1) and (U2). To prove Theorem 4, we need to show that the set of points lying in the neighborhood $U_n := U_{\lambda_u^{-n}}(p)\times\{x\}$ and not belonging to $\mathcal{B}_0$ can be covered by a union over $m \geq (1+c_1)n$ of unions of at most $\lambda_u^{c_2(m-n)}$ balls of radius λ_u^{-m}.

Let $\tilde{f}_{n,y}$ be the transformation of the fiber over y in n iterations,

$$\tilde{f}_{n,y} := \tilde{f}_{T^{n-1}(y)} \circ \dots \circ \tilde{f}_{T(y)} \circ \tilde{f}_y;$$

in particular, $\tilde{f}_{1,y} = \tilde{f}_y$.

Fix a point (p,x), $p \in W^s(y_{(0)})$. By assumption, $\tilde{f}'_{k_0,y_{(0)}}(0) < 1$, and hence there exist sufficiently small $\varepsilon, \delta > 0$ and $r < 1$ such that $\tilde{f}_{k_0,\bar{p}}(\bar{x}) \leq r\bar{x}$ for all $\bar{p} \in U_\delta(y_{(0)})$ and $\bar{x} \leq \varepsilon$.

For convenience, we replace our family of horizontal neighborhoods. For the case of a skew product over a linear Anosov diffeomorphism, we replace the ball $U_{\lambda_u^{-n}}(p)$ by a parallelogram neighborhood defined as the product of neighborhoods of λ_u^{-n} along the stable and unstable directions; for the case of the circle doubling map, we take a neighborhood that is an interval of radius $\lambda_u^{-n}/2$. We denote the new family of neighborhoods by P_n. Since the neighborhoods in the new family contain those of the old family (up to the shift by 1 in the numbering), it suffices to prove the conclusion of the theorem for the new family of neighborhoods.

Let $y \in P_n$; consider the image of the point (y,x) for n iterations. Since $p \in W^s(y_{(0)})$, it follows that almost the entire sequence of iterates, except for a few at the beginning and at the end (whose number is uniformly bounded by a constant independent of n) splits into parts corresponding to an application of $\tilde{f}_{k_0,q}$ with $q \in U_\delta(y_{(0)})$ and $\bar{x} \leq \varepsilon$. Hence the x-coordinate of the image is exponentially small,

$$\forall y \in P_n \quad \tilde{f}_{y,n}(x) \leq C \cdot r^{n/k_0}. \tag{35}$$

For each $N > n$, consider the set $\overline{M}_N$ of points $y \in P_n$ of the base from which the Nth image of (y,x) is "insufficiently close" to A_0,

$$\overline{M}_N := \{y \in P_n \mid \tilde{f}_{y,N}(x) \geq e^{I(N-n)/2}\}, \tag{36}$$

where $I := \int_B \psi \, d\mu < 0$ and the function $\psi = \log \varphi_0^{\varepsilon_0}$ is defined in the proof of Theorem 3.

Note that, by (35), there exists $c_1 > 0$ such that the sets $\overline{M}_N$ are empty for $N \leq n(1+c_1)$. Indeed, let $L := \max_{(y,x)} \tilde{f}'_y(x)$; then the estimate $\tilde{f}_{y,N}(x) \leq Cr^{n/k_0} \cdot L^{N-n}$ holds for each point $y \in P_n$. Now, for inequality (36) to fail (and accordingly, for $\overline{M}_N$ to be empty) it suffices that the inequality

$$Cr^{n/k_0} \cdot L^{N-n} < e^{I(N-n)/2}$$

be satisfied. In turn, this inequality is equivalent to

$$(N-n)\Big(\log L - \frac{I}{2}\Big) < \frac{|\log r|}{k_0} n - \log C, \tag{37}$$

whence we obtain the desired estimate.

Next, consider the sets $M_N := \overline{M}_N \setminus \bigcup_{j=n}^{N-1} \overline{M}_j$. Note that the union of M_N covers the desired set of points of the form (y,x), $y \in P_n$, not tending to A_0. On the other hand, for sufficiently large n (such that $e^{Ic_1 n/2} < \varepsilon_0$), the last $N-n$ iterations before "going beyond the exponential bound" at the Nth iteration can be estimated for an arbitrary point $y \in M_N$ by analogy with (18)–(19). Namely, let $x_j = \tilde{f}_{y,j}(x)$. Then

$$\log \tilde{f}_{y,N}(x) = \log \tilde{f}_{y,n}(x) + \sum_{j=n}^{N-1} \log\Big(\frac{\tilde{f}_{T^j(y)}(x_j)}{x_j}\Big) \leq \sum_{j=n}^{N-1} \psi(T^j(y));$$

thus if $(y,x) \in M_N$, then, in particular, we have the estimate

$$\sum_{j=n}^{N-1} \psi(T^j(y)) \geq \frac{I}{2}(N-n). \tag{38}$$

The fraction of points of the base B at which (38) holds is exponentially small by Young's theorem mentioned in Sec. 4 (see [Y90]). However, we work not with the entire set B but with its part P_n, and the measure of this part itself is exponentially small with respect to n. Hence we need a slightly more careful argument, which we carry out separately for the case in which the base is the circle and the torus.

For the circle doubling map, the map T^n linearly extends the interval P_n onto the circle and glues together its boundary points. Next, let us divide the original interval P_n into 2^{N-n} intervals $\{R_j\}_{j=1}^{2^{N-n}}$ of length 2^{-N}; then the images $\{T^n(R_j)\}$ of these intervals for n iterations form a partition of the circle.

Now note that if two points y and y' belong to the same interval R_j, then the sums on the left-hand sides in (38) differ by at most some constant C_2 (which follows from an estimate similar to the predictability lemma). Thus if (38) holds at a point y, then for each point y' in the same interval one has

$$\sum_{j=n}^{N-1} \psi(T^j(y')) \geq \frac{I}{2}(N-n) + C_2, \tag{39}$$

and accordingly, for the image $y'' = T^n(y')$ one has

$$\frac{1}{N-n} \sum_{j=0}^{(N-n)-1} \psi(T^j(y'')) \geq \frac{I}{2} + \frac{C_2}{N-n} \geq \frac{2I}{3}. \tag{40}$$

Let us cover the set M_N by the union of intervals R_i that contain at least one point of this set. The number of these intervals can be estimated by Young's theorem. Namely, the image of their union for n iterations is a set of exponentially small measure with respect to $N-n$, because there is a "large deviation" (40) at every point of this set. Thus, the fraction of the intervals chosen is exponentially small; namely, the number of these intervals can be estimated from above as $2^{c_2(N-n)}$, where $c_2 < 1$. We have obtained the desired cover of M_N by at most $2^{c_2(N-n)}$ of length 2^{-m}; the union of these covers over N gives the desired cover of the set of points that do not tend to A_0.

The uniformness of the estimate of the constant c_1 follows from the estimate on N in (37). The uniformness of the estimate of the constant c_2 follows from the uniformness in Young's theorem. (A small change in the skew product results in a small change in the corresponding function ψ.) The proof of Theorem 4 for the case of skew products over the circle doubling map is complete.

In a similar way, one proves the theorem for the base $B = \mathbb{T}^2$. Namely, the image $T^n(P_n)$ is a parallelogram of size $\lambda_u^{-2n} \times 1$ in the stable and unstable directions, respectively.

Let us cut P_n into parallelograms of size $\lambda_u^{-n} \times \lambda_u^{-N}$ in the stable and unstable directions, respectively. The image of such a partition for n iterations is a partition of $T^n(P_n)$ into parallelograms $\{R_i\}$ of size $\lambda_u^{-2n} \times \lambda_u^{-(N-n)}$. For each such parallelogram R_i (of size $\lambda_u^{-2n} \times \lambda_u^{-(N-n)}$), consider an ambient parallelogram $\tilde{R}_i$ of size $C_1 \times \lambda_u^{-(N-n)}$, where C_1 is a constant. If C_1 is sufficiently small, then these parallelograms $\tilde{R}_i$ are pairwise disjoint (see Figure 2). On the other hand, the union of the $\tilde{R}_j$ has measure C_1.

By analogy with the case of the circle doubling map, if points $T^n(y)$ and y'' of the torus are contained in the same parallelogram $\tilde{R}$, then the sum on the left-hand side

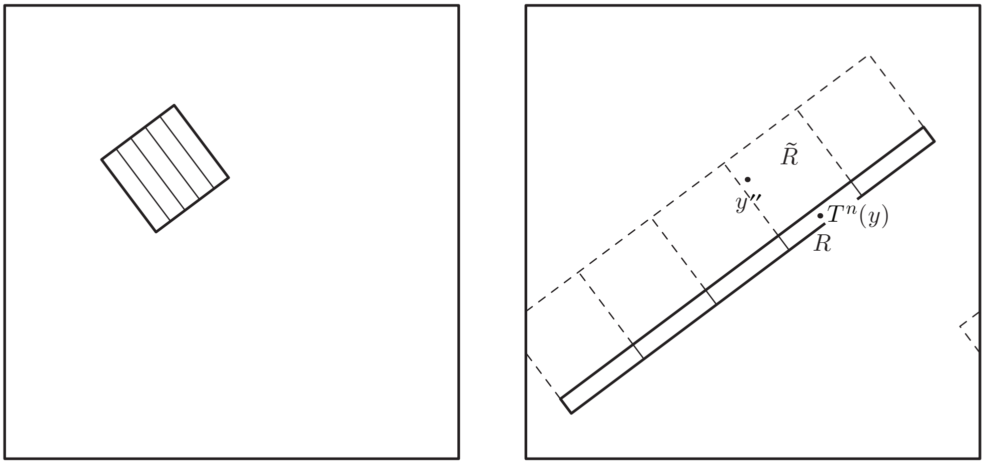

FIGURE 2. The original partition, its image R, and the partition $\tilde{R}$.

in (38) for y and on the left-hand side in (40) for y'' differ at most by some constant C_2,

$$\left|\sum_{j=n}^{N-1} \psi(T^j(y)) - \sum_{j=0}^{(N-n)-1} \psi(T^j(y''))\right| \le C_2.$$

Consider the union of parallelograms $\{R_{i_k}\}$ of the original partition that contain at least one point of the set $T^n(M_N)$. For each such parallelogram, all points y'' of the corresponding parallelogram $\tilde{R}_{i_k}$ satisfy the "large deviation" inequality (40). An application of Young's theorem permits one to conclude that the measure of the union of such parallelograms $\tilde{R}_{i_k}$ is exponentially small. Since the measure of the union of all parallelograms $\tilde{R}_i$ is equal to C_1 (and is independent of n and N), it follows that we have chosen an exponentially small fraction of the set $\{\tilde{R}_i\}$ of all parallelograms, and hence the fraction of parallelograms of the original partition meeting the set M_N is exponentially small as well.

Thus, the set M_N is covered by $\lambda_u^{c(N-n)}$, $c < 1$, parallelograms of size $\lambda_u^{-n} \times \lambda_u^{-N}$. By subdividing this cover in the stable direction, we obtain a cover $\lambda_u^{c_2(N-n)}$, $c_2 = c+1 < 2$, by parallelograms of size $\lambda_u^{-N} \times \lambda_u^{-N}$. The uniformness of the estimate for the constants c_1 and c_2 can be proved in the same way as for skew products over the circle.

The proof of Theorem 4 is complete.

Acknowledgments

The authors are grateful to Yu. S. Ilyashenko for setting the problem and for useful discussions, to M. L. Blank for communicating Lai-Sang Young's result, to C. Bonatti, D. A. Ryzhov, and M. Yu. Lyubich for valuable discussions, and to J. Milnor and A. Bonifant for interest in the present research.

References

[BLR] P. Bleher, M. Lyubich, and R. Roeder, *Lee–Yang zeros for DHL and 2D rational dynamics, I. Foliation of the physical cylinder*, preprint, arXiv:1009.4691 [math.DS], 2010.

[BDV05] C. Bonatti, L. J. Diaz, and M. Viana, *Dynamics beyond uniform hyperbolicity*, Springer, Berlin, 2005. MR2105774 (2005g:37001)

[BM07] A. Bonifant and J. Milnor, *Schwarzian derivatives and cylinder maps*, preprint, 2007. MR2477416 (2010f:37041)

[DK07] B. Deroin and V. Kleptsyn, *Random conformal dynamical systems*, Geometry and Functional Analysis, **17** (2007), no. 4, 1043–1105. MR2373011 (2010j:37012)

[GI96] A. Gorodetski and Yu. Ilyashenko, *Minimal and strange attractors*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **6** (1996), no 6, 1177–1183. MR1409419 (97e:58151)

[GI99] A. S. Gorodetskiĭ and Yu. S. Ilyashenko, *Some new robust properties of invariant sets and attractors of dynamical systems*, Funktsional. Anal. i Prilozhen. **33** (1999), no. 2, 16–30; English transl., Funct. Anal. Appl. **33** (1999), no. 2, 95–105. MR1719330 (2001c:37031)

[GI00] A. S. Gorodetskiĭ and Yu. S. Ilyashenko, *Some properties of skew products over a horseshoe and a solenoid*, Tr. Mat. Inst. Steklova **231** (2000), 96–118; English transl., Proc. Steklov Inst. Math. **231** (2000), no. 4, 90–112. MR1841753 (2002i:37040)

[G01] A. S. Gorodetski, *Minimal attractors and partially hyperbolic invariant sets of dynamical systems*, Cand. Sci. (Phys.–Math.) Dissertation, Faculty of Mechanics and Mathematics, Moscow State University, 2001.

[G06] A. S. Gorodetskiĭ, *Regularity of central leaves of partially hyperbolic sets and applications*, Izv. Ross. Akad. Nauk Ser. Mat. **70** (2006), no. 6, 19–44; English transl., Izv. Math. **70** (2006), no. 6, 1093–1116. MR2285025 (2007k:37033)

[GIKN05] A. S. Gorodetskiĭ, Yu. S. Ilyashenko, V. A. Kleptsyn, and M. B. Nal′skiĭ, *Nonremovability of zero Lyapunov exponents*, Funktsional. Anal. i Prilozhen. **39** (2005), no. 1, 27–38; English transl., Funct. Anal. Appl. **39** (2005), no. 1, 21–30. MR2132437 (2005m:37062)

[GT02] B. M. Gurevich and A. A. Tempel′man, *Hausdorff dimension of the set of generic points for Gibbs measures*, Funktsional. Anal. i Prilozhen. **36** (2002), no. 3, 68–71; English transl., Funct. Anal. Appl. **36** (2002), no. 3, 225–227. MR1935906 (2003j:28026)

[I08] Yu. S. Ilyashenko, *Diffeomorphisms with intermingled attracting basins*, Funktsional. Anal. i Prilozhen. **42** (2008), no. 4, 60–71; English transl., Funct. Anal. Appl. **42** (2008), no. 4, 298–307. MR2492427 (2010b:37045)

[IKS08] Yu. Ilyashenko, V. Kleptsyn, and P. Saltykov, *Openness of the set of boundary preserving maps of an annulus with intermingled attracting basins*, J. Fixed Point Theory Appl. **3** (2008), no. 2, 449–463. MR2434457 (2009h:37043)

[HPS77] M. W. Hirsch, C. C. Pugh, and M. Shub, *Invariant manifolds*, Lect. Notes in Math., vol. 583, Springer, Berlin–New York, 1977. MR0501173 (58:18595)

[IN10] Yu. Ilyashenko and A. Negut, *Hölder properties of perturbed skew products and Fubini regained*, preprint, arXiv:1005.0173 [math.DS], 2010.

[K94] I. Kan, *Open sets of diffeomorphisms having two attractors, each with everywhere dense basin*, Bull. Amer. Math. Soc. **31** (1994), no. 1, 68–74. MR1254075 (94k:58082)

[KR] V. Kleptsyn and D. Ryzhov, *Special ergodic theorem*, a note in preparation.

[Mi97] J. Milnor, *Fubini foiled: Katok's paradoxical example in measure theory*, Math. Intelligencer **19** (1997), no. 2, 30–32. MR1457445 (98g:28005)

[Os10] A. V. Osipov, *Nondensity of the orbital shadowing property in C^1-topology*, Algebra i Analiz **22** (2010), no. 2, 127–163; English transl., St. Petersburg Math. J. **22** (2011), no. 2, 267–292. MR2668126 (2011f:37042)

[Pa00] J. Palis, *A global view of dynamics and a conjecture on the dynamics of finitude of attractors*, Géometrie complexe et systèmes dynamiques (Orsay, 1995), Astérisque **261** (2000), xiii–xiv; 335–347. MR1755446 (2001d:37025)

[Pa05] J. Palis, *A global perspective for non-conservative dynamics*, Ann. Inst. H. Poincaré Anal. Non Linéaire **22** (2005), no. 4, 485–507. MR2145722 (2006b:37037)

[Ru01] D. Ruelle, *Historical behaviour in smooth dynamical systems*, Global analysis of dynamical systems, Inst. Phys., Bristol, 2001, pp. 63–66. MR1858471

[S10] P. S. Saltykov, *A special ergodic theorem for Anosov diffeomorphisms on the 2-torus*, Funktsional. Anal. i Prilozhen. **45** (2011), no. 1,69-78; English transl., Funct. Anal. Appl. **45** (2011), no. 1, 56–63. MR2848741

[SW00] M. Shub and A. Wilkinson, *Pathological foliations and removable zero exponents*, Invent. Math. **139** (2000), no. 3, 495–508. MR1738057 (2001c:37030)

[T08] F. Takens, *Orbits with historic behaviour, or non-existence of averages*, Nonlinearity **21** (2008), no. 3, T33–T36. MR2396607 (2009a:37043)

[Var08] S. R. S. Varadhan, *Large deviations*, Ann. Probab. **36** (2008), no. 2, 397–419. MR2393987 (2009d:60070)

[Y90] Lai-Sang Young, *Large deviations in dynamical systems*, Trans. Amer. Math. Soc. **318** (1990), no. 2, 525–543. MR975689 (90g:58069)

[Y03] Lai-Sang Young, *Entropy in dynamical systems*, Entropy, Princeton Ser. Appl. Math., Princeton Univ. Press, Princeton, NJ, 2003, pp. 313–327. MR2035829 (2004k:37010)

CNRS, Institut de Recherche Mathématique de Rennes (UMR 6625), France
E-mail address: `victor.kleptsyn@univ-rennes1.fr`

Faculty of Mechanics and Mathematics, Moscow State University, Moscow 119992, Russian Federation
E-mail address: `p_saltykov@pochta.ru`

Translated by V. E. NAZAIKINSKII