Local Ergodicity for Systems with Growth Properties including Multi-dimensional Dispersing Billiards

Pavel Bachurin
Department of Mathematics, University of Toronto,
Toronto, ON, M5S 2E4, Canada
bachurin@math.toronto.edu
Péter Bálint
Mathematical Institute of the Technical University of Budapest
Egry József u. 1, H-1111, Budapest, Hungary and
Courant Institute of Mathematical Sciences, New York University
251 Mercer Street, New York, NY 10012, USA
bp@renyi.hu
Imre Péter Tóth
Alfréd Rényi Institute of the Hungarian Academy of Sciences
Reáltanoda u. 13–15, H-1053, Budapest, Hungary, and
Research group "Stochastics" of the Technical University of Budapest
and the Hungarian Academy of Sciences
mogy@renyi.hu

February 6, 2007

Abstract

We prove local ergodicity of uniformly hyperbolic discrete time dynamical systems with singularities, which satisfy certain regularity conditions and an assumption on the growth of unstable manifolds. We apply the result to prove ergodicity of a class of multi-dimensional dispersing billiards.

Introduction

In this paper we study local ergodicity of hyperbolic systems with singularities. Our goal is to understand a *multi-dimensional, uniformly* hyperbolic setting, in particular *multi-dimensional dispersing billiards*.

Theorems that guarantee abundance of stable and unstable manifolds, designed for local ergodicity purposes (Hopf chain construction), are often referred to as *fundamental theorem*s in the literature. We mention theorems as there are (at least) two different types of statements of this sort.

The one which we concentrate on in this paper is Sinai's argument from [S], worked out for the case of two dimensional dispersing billiards in its original formulation. Sinai's approach uses uniform hyperbolicity strongly, and, in the original exposition, also two dimensional geometry. One of our main goals with the present paper is to show that this beautiful theorem by Sinai can be suitably generalized to a *multi-dimensional* uniformly hyperbolic setting. In particular, we can prove this way local ergodicity for a reasonable class of multi-dimensional dispersing billiards (see below).

The other formulation of the fundamental theorem is adapted to non-uniformly hyperbolic situations. This case, of course, needs much more care and more complicated constructions – eg. regular coverings of local neighborhoods with parallelopipeds, the Chernov-Sinai Ansatz etc. We do not consider the issue of this "non-uniformly hyperbolic" fundamental theorem here, we refer to the literature instead, see [SCh], [KSSz], [Ch1], [LW], [BChSzT2], [B] and refereces therein.

Our exposition states a set of assumptions based on which the fundamental theorem – and thus local ergodicity – can be proved. Though our main application is definitely the multi-dimensional dispersing billiard situation, we believe that there are many other classes of systems that satisfy our assumptions.

To see in what extent our results are new in the billiard case we recall some recent history. In [BChSzT1] a pathological behaviour of singularities in multi-dimensional (semi-)dispersing billiards was found. This discovery calls for a reconsideration of earlier proofs of ergodicity in multi-dimensional semi-dispersing billiards. The papers [BChSzT2] and [B] handle the problem for certain special cases: [BChSzT2] deals with the case of algebraic scatterers, while [B] treats *strictly* dispersing billiards with highly smooth scatterers, and a strong condition on the complexity of the singularities. The present paper also contributes to this work: in Theorem 4.4 we prove ergodicity for a broader class of billiards, requiring only C^3 smoothness of the scatterers, but still requiring strict dispersing, and a weaker assumption about complexity of singularities.

Actually, the systems we are interested in are more than just ergodic. In his paper [Ch3], Chernov gave sufficient conditions for exponential decay of correlations in hyperbolic systems with singularities. In that paper, ergodicity is one of the assumptions, and other assumptions – regularity conditions and growth of unstable manifolds – are stronger than those usually needed in proofs of ergodicity. No wonder, that until recently, all the systems for which the result of [Ch3] could be applied, were known in advance to be ergodic.

However, in the recent manuscript [BT], two of the present authors managed to check all the conditions in [Ch3] – except for ergodicity – for a class of multi-dimensional billiards for which no independent proof of ergodicity is known. This naturally raises the question, whether the other assumptions of [Ch3] imply ergodicity in any sense.

In this paper we give a partially positive answer: we prove *local* ergodicity. Due to the local nature of the assumptions, there is no chance to get more. In this way, the assumptions of [Ch3] give exponential correlation decay on ergodic components.

The assumptions we need for ergodicity are basically a subset of Chernov's assumptions in [Ch3], e.g. our main assumption about the growth of unstable manifolds (Assumption A8) is a simplified version of the growth property required in [Ch3]. Only a few of our assumptions are a little stronger than those in [Ch3]. Namely, the equivalence of measures (Assumption A2) and alignment (Assumption A5) are not required in [Ch3], and absolute continuity (Assumption A7) is required in a somewhat different form. However, most of these properties are known in case of the systems of interest.

This way, we can prove Theorem 4.4 about ergodicity of a class of multi-dimensional dispersing billiards with C^3 smooth scatterers. The result of [Ch3] then gives exponential correlation decay for the same sytems. Before this, in more than two dimensions only those dispersing billiards were known to be ergodic, for which the scatterers exhibit the stronger regularity properties required in [BChSzT2] or [B].

The paper is organized as follows. In Section 1 we list some basic notions and our assumptions, while in Section 2 we draw a few immediate consequences of these assumptions. In Section 3 we prove Theorem 3.1, which is the adoptation of Sinai's original fundamental theorem for the multi-dimensional case. Then in Section 4 we perform the Hopf chain construction based on the fundamental theorem to prove Theorem 4.1 about local ergodicity. Finally, as a corollary, we prove Theorem 4.4 about ergodicity of certain multi-dimensional dispersing billiards.

1 Assumptions and preliminaries

Assumption A1 (The Dynamical system)**.**
We consider the dynamical system (M, T, μ) where the phase space M is a compact Riemannian manifold of dimension d_M, possibly with boundary. The dynamics T is defined on a full (Riemannian) measure subset of M. Namely, there are two closed subsets $\Gamma_1, \Gamma_{-1} \subset M$ with Riemannian measure zero, such that $T : M \setminus \Gamma_1 \to M \setminus \Gamma_{-1}$ is a diffeomorphism. μ is a T-invariant probability measure on M. If M has boundary, then we assume $\partial M \subset \Gamma_1, \Gamma_{-1}$.

We will use the notation ρ for the distance and m for the induced Lebesgue (Riemannian) measure on M. If V is a submanifold of M, ρ_V and m_V will denote the induced metric and measure on V.

We will call Γ_1 the singularity set of T, and Γ_{-1} the singularity set of T^{-1}.

For integer $n \neq 0$ we introduce the notation Γ_n for the n-step singularity set:

$$\Gamma_n = \begin{cases} T^{-n+1}\Gamma_1 & \text{if } n \geq 1 \\ T^{n+1}\Gamma_{-1} & \text{if } n \leq -1 \end{cases}.$$

We say that T^n is singular in the set

$$\Gamma^{(n)} = \begin{cases} \bigcup_{k=1}^{n-1} \Gamma_k & \text{if } n \geq 1 \\ \bigcup_{k=1}^{-n-1} \Gamma_{-k} & \text{if } n \leq -1 \end{cases}.$$

Remark 1.1. *In known examples Γ_1 is either a finite collection of 1 codimensional submanifolds, or $\Gamma_1 = \mathcal{S} \cup \tilde{\Gamma}$, where $\tilde{\Gamma}$ is a countable collection of 1 codimensional submanifolds that accumulates on $\mathcal{S}$ which is a finite union of 1 codimension submanifolds. In the literature (see e.g.[BT]) this $\mathcal{S}$ is often referred to as the "primary singularity set". Often one can roughly write $\mathcal{S} = T^{-1}\partial M \cup \partial M$. Note however, that for convenience we have chosen not to define T on Γ_1.*

It is also worth noting that we do not have any smoothness requirements on the singularity. If T has unbounded and anisotropic derivatives, it may happen that, though Γ_1 consists of smooth and compact submanifolds, the sets Γ_n, $n \geq 2$ behave irregularly. This happens in the case of multi-dimensional dispersing billiards, see [BChSzT1, BChSzT2]. This phenomenon does not conflict our assumptions.

Assumption A2 (Equivalence of measures)**.**
We assume that the invariant measure μ is equivalent to the Lebesgue (Riemannian) measure m on M.

This assumption will only be used in Section 4 for the Hopf chain construction. It is possible that it could be relaxed somewhat, but we don't investigate this issue here. The assumption is in coherence with (actually implies) the fact that Γ_1 and Γ_{-1} have Riemannian measure zero.

Assumption A3 (Uniform hyperbolicity)**.**
We assume that there are two families of cone fields C^u_x and C^s_x in the tangent planes $\mathcal{T}_x M$, $x \in M$ and there exists a constant $\Lambda > 1$ with the following properties:

- $DT(C^u_x) \subset C^u_{Tx}$ *and* $DT(C^s_x) \supset C^s_{Tx}$ *whenever* DT *exists;*
- $|DT(v)| \geq \Lambda |v| \quad \forall v \in C^u_x$*;*
- $|DT^{-1}(v)| \geq \Lambda |v| \quad \forall v \in C^s_x$*;*
- *these families of cones are continuous on M, their axes have the same dimensions across the entire M which we denote by d_u and d_s, respectively;*
- $d_u + d_s = \dim M$*;*

- *the angles between C_x^u and C_x^s are uniformly bounded away from zero:*

  $\exists\ \delta_{us} > 0$ *such that* $\forall x \in M$ *and for any* $dw_1 \in C_x^u$ *and* $dw_2 \in C_x^s$ *one has*

$$\sphericalangle(dw_1, dw_2) \geq \delta_{us}.$$

The C_x^u are called the unstable cones whereas the C_x^s are called the stable ones.

The property that the angle between stable and unstable cones is uniformly bounded away from zero is called **transversality**. Since we have demanded that the cone fields be defined and continuous on all of M (which is compact), this transversality is automatic, still we wish to stress its importance and use the constant δ_{us}.

When investigating unstable manifolds, we will use the notation $d = d_u$.

Remark 1.2. *In the literature uniform hyperbolicity is often used in a weaker form, namely so that only the existence of a $c > 0$ and a $\Lambda > 1$ is assumed so that $|DT^n(v)| \geq c\Lambda^n|v| \quad \forall v \in C_x^u$ and $|DT^{-n}(v)| \geq c\Lambda^n|v| \quad \forall v \in C_x^s$. Throughout the paper, all our arguments could easily be modified to fit this setting as well.*

To formulate further assumptions, we need the notion of stable and unstable manifolds.

Definition 1.3. *A d_u-dimensional submanifold γ^u is a u-manifold if, for any point $x \in \gamma^u$, the tangent plane of γ^u at x belongs to the unstable cone C_x^u. s-manifolds are defined analogously. A special subclass of u-manifolds (s-manifolds) is of particular interest:*

A submanifold $W^u \subset M$ of dimension d_u is an unstable manifold if T^{-n} is defined and smooth on it for all $n \geq 0$, and for any $x, y \in W^u$ $\rho(T^{-n}x, T^{-n}y) \to 0$ exponentially fast.

A submanifold $W^s \subset M$ of dimension d_s is a stable manifold if T^n is defined and smooth on it for all $n \geq 0$, and for any $x, y \in W^s$ $\rho(T^n x, T^n y) \to 0$ exponentially fast.

When we consider a small (piece of a) stable manifold, we will use the phrase "local stable" manifold (LSM), to emphasize that we think of a *connected* submanifold on which T^n is defined for all $n > 0$, or a subset of such a connected manifold. That is, several components which can be separated from each other temporarily, but are eventually brought back together by T are not allowed. Similarly, the unstable manifolds under consideration will be local unstable manifolds (LUM).

Assumption A4 (Existence of invariant manifolds)**.**
We assume that for μ-a.e. point x of M there exists a unique maximal local stable and a unique maximal local unstable manifold, both of which contain x as an interior point.

Remark 1.4. *The μ-a.e. existence of invariant manifolds follows from uniform hyperbolicity and some regularity assumption on the* frist step *singularity set Γ_1. For example*

$$m(\{\, x \in M \mid \rho(x, \Gamma_1) < \varepsilon \,\}) \leq C\varepsilon^q \quad \text{for some } q > 0 \tag{1}$$

would do, see [Ch2] and references therein. In the setting of multi-dimensional dispersing billiards such a regularity assumption holds for Γ_1 itself, it is only the higher order singularities Γ_n, $n \geq 2$ for which irregularities start to show up, see Remark 1.1 above. We could have assumed (1) directly, however, we prefer to state assumption A4 as this is the one we need.

We use the notation $W^s(x)$ ($W^u(x)$) for the maximal local stable (unstable) manifold passing through $x \in M$. We will denote by $r^s(x)$ ($r^u(x)$) the radius of the largest d_s-dimensional (d_u-dimensional) ball centered at x contained in $W^s(x)$ ($W^u(x)$), in the inner geometry of the submanifold.

In case $r^s(x) > R$, $W^s_R(x)$ is the stable ball of radius R around x. That is: $W^s_R(x) = \{y \in W^s(x) \mid \rho_{W^s(x)}(x,y) \leq R\}$. $W^u_R(x)$ is defined similarly.

If W is a LSM, we denote its inner radius with $r(W)$, that is, $r(W) = \sup_{x \in W} r^s(x)$. The inner radius of an unstable manifold is denifed similarly.

Our next assumption, called alignment, is a key regularity property that allows us to compare the distance of a point form a singularity when measured in two different ways, along the unstable direction, or in the entire manifold. We demand such a property for Γ_{-1}, the singularity set of T^{-1}. This is not a usual property to assume, but due to the nature of the growth propery we will use (Assumption A8), this is what we need.

Our assumption basically means that unstable manifolds are uniformly transversal to components of Γ_{-1}. That is, there is a global constant $C_a < \infty$ so that the distance of any point x from Γ_{-1}, when measured along the unstable manifold, is at most C_a times the distance in the phase space. More precisely, this is only required if the unstable manifold containing x is long enough to reach that far in every direction. Note, however, that unstable manifolds terminate on Γ_{-1}, thus technically it is enough to compare the inner radius of the maximal unstable manifold passing through x to the distance of x form Γ_{-1}, when measured in the full phase space.

We give two versions of this assumption. The first, simplified version is more transparent, but not general enough to cover billiards.

Assumption A5 (Alignment)**.**
We assume that the dynamical system satisfies one of the following two properties:

1. ***Alignment, simplified version***

   There is a global constant $C_a < \infty$ such that for any $x \in M$ we have $\rho(x, \Gamma_{-1}) \geq \frac{1}{C_a} r^u(x)$. Similarly, $\rho(x, \Gamma_1) \geq \frac{1}{C_a} r^s(x)$ is assumed.

2. ***Alignment, general version***

   Consider any $x \in M$ and $y \in \Gamma_{-1}$. We assume that

   - *either $\rho(x,y) \geq \frac{1}{C_a} r^u(x)$,*
   - *or the image of y under T^{-1} is also in Γ_{-1} in the following sense (recall from Assumption A1 that T^{-1} is not defined on Γ_{-1}): Take any continuous curve*

γ connecting x to y and avoiding Γ_{-1} – that is, $\gamma : [0,1] \to M$, $\gamma(0) = x$, $\gamma(1) = y$, $\gamma(t) \notin \Gamma_{-1}$ for $t < 1$. We require that $y^- = \lim_{t\to 1} T^{-1}\gamma(t)$ exists and $y^- \in \Gamma_1 \cap \Gamma_{-1}$.

The analogous property for $r^s(x)$ and Γ_1 is also assumed.

The simplified version of this assumption obviously implies the general, so we will use the general one in our proof.

In applications this property is typically a consequence of some smoothness of the manifolds, plus a uniform lower bound on their angles at the intersection points. The heuristic meaning of the general version is that as an unstable manifold evolving under T hits a component of $\Gamma_1 \cap \Gamma_{-1}$ (e.g. ∂M), they only need to be properly aligned at their first encounter. If the "image under T" (as a one-sided limit) of that component is also in Γ_{-1}, alignment there is no longer required, at least not from that side.

Remark 1.5. *This assumption is general enough to cover many interesting examples, including (multi-dimensional) dispersing billiards with no corner points. However, we mention two examples where it does not hold: the baker's map, and billiards with corner points.*

Assumption A6 (Distortions)**.**
Consider a LUM W that lies in one connected component of $M \setminus \Gamma^{(n)}$, i.e., the map T^n is smoothly defined on W. In particular, $W' = T^nW$ is a LUM as well. Given $x \in W$ consider the n-step unstable Jacobian $J^u_n(x)$, i.e., the Jacobian (w.r. to the Lebesgue measures on W and W') of the map T^n restricted to W at the point x. We assume there exists a function $\varphi(\cdot)$ with $\varphi(s) \to 0$ as $s \to 0$, such that, given any W described above and any $x, y \in W$, we have $\log \frac{J^u_n(x)}{J^u_n(y)} \le \varphi(\rho_{W'}(T^nx, T^ny))$.

We assume similar distortion bounds for T^{-n} when restricted to suitable LSMs.

Assumption A7 (Absolute continuity)**.**
Let γ_1, γ_2 be two u-manifolds (cf. Definition 1.3). Note this implies that both γ_1 and γ_2 are d_u-dimensional submanifolds of M, uniformly transversal to stable manifolds. Let them be so small that any LSM W^s intersects both γ_1 and γ_2 in at most one point. Let $\gamma_1' = \{x \in \gamma_1 \mid W^s(x) \cap \gamma_2 \neq \emptyset\}$. Then we define a map $h : \gamma_1' \to \gamma_2$ by sliding along stable manifolds. This map is often called the holonomy map. We assume that it is absolutely continuous with respect to the Lebesgue measures m_{γ_1} and m_{γ_2}.

Remark 1.6. *Similarly to assumptions A2 and A4, the assumption on absolute continuity is only used in Section 4, i.e. it is not needed for the proof of the fundamental theorem itself.*

For $\delta_0 > 0$ we say that a LUM W is a δ_0-LUM if $\mathrm{diam} W \le \delta_0$. In our arguments we will only use a fixed δ_0 that will be chosen to be sufficiently small. For a δ_0-LUM W, the components of T^nW are not necessarily δ_0-LUM-s, due to the expansion by T. However, they may be chopped into δ_0-LUM-s by omitting a set of measure zero. This may be done e.g. by cutting with a grid of locally flat hypersurfaces that cut a piece of M into cubes of diameter less than δ_0.

Assumption A8 (Growth property).
We assume that there exist $0 < \alpha < 1$ and $\beta > 0$ such that given any δ_0-LUM W, its T-image TW can be decomposed, modulo a set of m_{TW}-measure zero, into (at most) countably many disjoint δ_0-LUM-s $TW_1, TW_2, \ldots$ such that for any $\varepsilon > 0$ we have:

$$m_W(\cup_i \{x \in W_i \mid \rho_{TW_i}(Tx, \partial TW_i) < \varepsilon\}) \leq \\ \leq \alpha \cdot \Lambda \, m_W(\{x \in W \mid \rho_W(x, \partial W) < \varepsilon/\Lambda\}) + \varepsilon\beta\delta_0^{-1} m_W(W). \qquad (2)$$

We also assume the time reversal counterpart of this statement for δ_0-LSMs.

Now if T is applied to the components $TW_1, TW_2, \ldots$, the images $T(TW_1), T(TW_2), \ldots$ may need to be chopped further to obtain δ_0-LUM-s. Iterating that, we can get for every $n \geq 1$ a partition of $T^n W$ (modulo a set of zero measure) into at most countably many disjoint δ_0-LUM-s. Having such a partion, we can define for every $n \geq 1$ a function on W that measures the distance of a point from the boundary of the containing component after n steps: $r_{W,n}(x) = \rho_{T^n W_j}(T^n x, \partial T^n W_j)$ if $T^n W_j$ is the component of $T^n W$ containing $T^n x$.

With this notiton, (2) can be rewritten as

$$m_W(r_{W,1} < \varepsilon) \leq \alpha\Lambda m_W(r_{W,0} < \varepsilon/\Lambda) + \varepsilon\beta\delta_0^{-1} m_W(W).$$

In [BT] it is shown that all the assumptions listed are satisfied in a class of multi-dimensional dispersing billiards.

2 Basic consequences of the assumptions

2.1 shadowing of unstable manifolds

The (uniform) continuity of the cone fields and the transversality condition (both in Assumption A3) easily imply the following lemma:

Lemma 2.1. *There exist constants $\delta_1 > 0$ and $C_1 = C_1(\delta_{us}) < \infty$ such that for any $x, y \in M$ satisfying $\rho(x,y) < \delta_1$, if $r^u(x) > C_1\rho(x,y)$ and $r^s(y) > C_1\rho(x,y)$, then $W^u(x) \cap W^s(y) \neq \emptyset$.*

from this we may conclude that two unstable manifolds which are sufficiently close to each other and sufficiently large, can surely be connected by a stable manifold which is large enough.

Definition 2.2. *Given two LUMs W' and W, we say that W'* s-shadows *W if for every $x \in W$ with $r^s(x) > 2r(W')$ we have that $W^s_{2r(W')}(x) \cap W' \neq \emptyset$.*

The following lemma is a trivial consequence of Lemma 2.1.

Lemma 2.3. *There exists a $C_{\delta_{us}} < \infty$ with the following property. If $W = W^u_R(z)$ and $W' = W^u_{R'}(z')$ with $R' > 100R$, and $\rho(z, z') < R' C_{\delta_{us}}$, then W' s-shadows W.*

2.2 growth lemma for iterates of T

Lemma 2.4. *If assumptions A6 and A8 are satisfied, then there exists a sufficiently small $\delta_0 > 0$ and global constants $0 < \alpha' < 1$, $\beta' < \infty$ such that for every δ_0-LUM W and every $n \geq 1$, $T^n W$ can be partitioned into δ_0-LUMs (modulo a set of measure zero) so that*

$$m_W(r_{W,n} < \varepsilon) \leq (\alpha'\Lambda)^n m_W(r_{W,0} < \varepsilon/\Lambda^n) + \varepsilon\beta'\delta_0^{-1} m_W(W). \tag{3}$$

Proof. The statement comes from (2) using induction in n, see e.g. [Ch3]. If δ_0 is chosen sufficiently small, then α' is just slightly greater than α, and β' is just slighlty greater than $\beta\left(1 + \frac{1}{\Lambda} + \frac{1}{\Lambda^2} + \dots\right)$. □

3 Growth Lemma implies Fundamental Theorem

Theorem 3.1 (Fundamental Theorem)**.** *Consider a dynamical system satisfying the assumptions A1, A3, A5, A6 and A8. Then, given any $A > 1$, there exists $n_A \in \mathbb{N}$ with the following property. For any $x \in M \setminus \Gamma^{(n_A)}$ there exists a neighborhood $B_R(x)$ such that, for any LUM $W \subset B_R(x)$:*

$$m_W(\{\, y \in W \,|\, r^s(y) > A \cdot r(W) \,\}) > 0.$$

Remark 3.2. *Note that assumptions A2, A4 and A7 are not required, see also Remark 1.6.*

Remark 3.3. *Note that this is an adaptation of Sinai's original fundamental theorem in [S], formulated for 2d dispersing billiards, see also [CM] Chapter 5. However, there are remarkable differences:*

- *in 2 dimensions, one can show the existence of long stable manifolds through a set of arbitrarily big relative measure in W. Here we only have a positive measure set of points y, but this is enough for the Hopf chain construction in Section 4.*
- *There is also an improtant improvement: we do not require $x \notin \Gamma^{(n)}$ for all n, just for some n_A fixed. This may make the proof of ergodicity simpler, as it can be seen in the proof of Theorem 4.4.*

This theorem has a time reversal counterpart.

Theorem 3.4. *Consider a dynamical system satisfying the assumptions A1, A3, A5, A6 and A8. Then, given any $A > 1$, there exists $n_A \in \mathbb{N}$ with the following property. For any $x \in M \setminus \Gamma^{(-n_A)}$ there exists a neighborhood $B_R(x)$ such that, for any LSM $W \subset B_R(x)$:*

$$m_W(\{ y \in W \,|\, r^u(y) > A \cdot r(W) \}) > 0.$$

Before the proof we state and prove a key lemma:

Lemma 3.5. *There exists a global constant C_g with the following property. Let W be a δ_0-LUM such that $\rho(W, \Gamma_1) > r(W)/C_g$, and that W lies in a connected component of $M \setminus \Gamma_1$. If $\varepsilon \le r(W)/C_g$, then $m_W(r^s > \varepsilon) > 0$.*

Proof. First we will prove the following claim: Let C_a be the constant in Assumption A5. Then given any $x \in W$ satisfying $r_k(x) > C_a\varepsilon/\Lambda^k$ for all $k \ge 1$, we have $r^s(x) > \varepsilon$.

To see this, take a point x that satisfies the assumption, and assume indirectly that $r^s(x) \le \varepsilon$. From the standard construction of stable manifolds in uniformly hyperbolic systems we know that the maximal stable manifold W^s of x can only terminate on Γ_n for some $n \ge 1$. So let $y \in \Gamma_n \cap \partial W^s$, $\rho_{W^s}(x, y) \le \varepsilon$. The assumption of the lemma implies that $n \ge 2$. The definition of Γ_n implies that $T^{n-1}y \in \Gamma_1$, but $T^{n-2}y \notin \Gamma_1$. Clearly $\rho_{T^{n-1}W^s}(T^n x, T^n y) \le \varepsilon/\Lambda^{n-1}$. Let γ be a curve of length $\le \varepsilon/\Lambda^{n-1}$ connecting $T^{n-1}x$ to $T^{n-1}y$ in $T^n W^s$ (as in Assumption A5). Clearly $T\gamma$ is a curve of length $\le \varepsilon/\Lambda^n$ connecting $T^n x$ to some $y^+ \in \Gamma_{-1}$. However, we have assumed $r_n(x) > C_a\varepsilon/\Lambda^n$, which implies that $r^u(T^n x) > C_a\varepsilon/\Lambda^n$. This means that in our alignment assumption (Assumption A5) applied to $T^n x$ and y^+, the first case is contradicted, so the second must hold, thus $T^{n-1}y \in \Gamma_{-1}$. But we have also assumed $r_{n-1}(x) > C_a\varepsilon/\Lambda^{n-1}$, thus applying, as above, the alignment assumption to $T^{n-1}x$ and $T^{n-1}y$, we now get $T^{n-2}y \in \Gamma_1$, a contradiction. So the claim is proven.

Lemma 2.4 allows us to estimate the measure of points which do not satisfy the condition in the claim. Specifically, the measure of these bad points can be compared to the measure of the ε-boundary of W. In order to compare this measure to the total measure of W, we apply the lemma to a *ball* in the unstable manifold.

In particular, if $W_a \subset W$ is a d_u dimensional unstable ball of radius $a = r(W) < \delta_0$, then

$$m_{W_a}(r_{W_a,n} < \varepsilon) \le (\alpha'\Lambda)^n C_1 a^{d-1}\varepsilon/\Lambda^n + C_2 a^d \varepsilon \le C_3 a^{d-1}\varepsilon \tag{4}$$

for all $n \ge 1$. Here and in the rest of the proof the global constants C_i may depend on $d(= d_u)$ via the volume and the area of the unit sphere in $\mathbb{R}^d$, as well as on the geometry of the cone fields in Assumption A3 and on δ_0, but this has no significance.

Substituting $\varepsilon \to C_a\varepsilon/\Lambda^n$ in (4), summing over $n \ge 1$ and applying the claim we get:

$$m_{W_a}(r^s < \varepsilon) \le C_4 \varepsilon a^{d-1}.$$

Since $m_{W_a}(W_a) \ge C_5 a^d$, we have $m_{W_a}(r^s > \varepsilon) > 0$ whenever $\varepsilon < a/C_g$. This observation fixes the global constant C_g. ∎

As we have assumed the time reversal counterparts for the properties A3, A5, A6 and A8, we may perform an analogous argument for LSMs, to have the time reversed analogue of Lemma 3.5:

Lemma 3.6. *There exists a global constant C_g with the following property. Let W be a δ_0-LSM such that $\rho(W, \Gamma_{-1}) > r(W)/C_g$, and that W lies in a connected component of $M \setminus \Gamma_{-1}$. If $\varepsilon \leq r(W)/C_g$, then $m_W(r^u > \varepsilon) > 0$.*

In the proof of the Fundamental Theorem we will use the following observation several times without mentioning:

Lemma 3.7. *There exists a global constant K such that for any LUM or LSM W if we have $W \subset B_R$ for some sphere B_R of some small radius R, then $r(W) \leq KR$.*

Proof. This follows from the uniform continuity of the cone fields in Assumption A3. $\square$

Proof of Theorem 3.1. The concept of constructing the large stable manifolds through a postitive measure set of points y is the following.

- Apply a high iterate T^n of the dynamics to W, so that it is stretched greatly
- Then use Lemma 3.5 to construct stable manifolds crossing T^nW, which will be comparable in radius to T^nW.
- Eventually, pull these stable manifolds back using T^{-n} to obtain long stable manifolds crossing W.

Now if n was chosen so that $\Lambda^{2n}/C_g > A$, then the stable manifolds constructed will have an inner radius greater than A times the inner radius of W.

To carry out this concept, we only need to make sure that the growth of our manifolds is not interrupted by singularities (including ∂M) along the n steps forward and back. This is ensured by chosing the set $B_R(x)$ so small, that its image under the n iterations still remains small, and keeps away from the singularities. The rest of the proof is dedicated to the technical details of this construction.

First we should tell how for a given $A > 1$ the integer n_A is chosen: we need $\Lambda^{2(n_A-1)}/C_g > A$ where $\Lambda > 1$ is the factor of minimal expansion from (Assumption A3), and C_g is the constant of Lemma 3.5. For brevity in the rest of the proof we omit the subscript A and use $n = n_A - 1$.

Now assume that we have some $x \in M \setminus \Gamma^{(n+1)}$: we should find an open ball $B_R(x)$ for which the statement of the Theorem holds. As $\Gamma^{(n+1)}$ is closed, there exists an open ball centered at x, $U = B(x)$ such that

$$U \cap \Gamma^{(n+1)} = \emptyset. \tag{5}$$

$B_R(x)$, the open ball of radius R should lie within U and should satisfy two further requirements, both of them ensuring that $B_R(x)$ lies "in the middle of" U in an appropriate sense. Later we will refer to these requirements as the first and the second requirement on $B = B_R(x)$.

(R1) On the one hand we need that $\rho(B_R(x), \partial U) > KAR$.

(R2) On the other hand we have that $\rho(T^n x, \Gamma_1) = R'' > 0$. Consider $B' = B_{R'}(T^n x)$, the open ball of radius $R' = \frac{R'' C_g}{K(C_g+1)}$ around $T^n x$. We require that $B \subset T^{-n} B'$.

The reason for the second requirement is as follows. If we have a LUM $W \subset B$, then $W' = T^n W$ is a LUM and $W' = T^n W \subset B'$. However given $W' \subset B'$, we have $r(W') < KR'$, thus, by the choice of R', $\rho(W', \Gamma_1) > r(W')/C_g$. On the other hand (5 implies that W' (as a subset ot $T^n U$) lies in one connected component of $M \setminus \Gamma_1$. These observations ensure that Lemma 3.5 applies to $W' = T^n W$ whenever $W \subset B$.

That is exactly what we are going to use. Consider an arbitrary LUM $W \subset B$ and let $W' = T^n W$. Then, by uniform hyperbolicity (Assumption A3), $r(W') \geq \Lambda^n r(W)$. Furthermore, as discussed above, Lemma 3.5 applies to W' for suitable ε, in particular for $\varepsilon = \Lambda^n r(W)/C_g \leq r(W')/C_g$. We have:

$$m_{W'}(\{\, y' \in W' \,|\, r^s(y') > \Lambda^n r(W)/C_g \,\}) > 0. \tag{6}$$

Now we invoke Assumption A6: note that W lies within one connected component of $M \setminus \Gamma^{(n)}$, thus the distortions of T^n are bounded on W. Introduce

$$W_g \subset W, \qquad W_g = \{\, y \in W \,|\, r^s(T^n y) > \Lambda^n r(W)/C_g \,\}.$$

By the above mentioned distortion bounds (6) implies $m_W(W_g) > 0$.

Now for $y \in W_g$ let $y' = T^n y$ and consider the stable manifold of y'. We distinguish between two possibilities. If this stable manifold is not cut by the singularities of T^{-n}, then, by Assumption A3, its inner radius grows at least by a factor Λ^n when T^{-n} is applied. Thus in this case

$$r^s(y) \geq \Lambda^n r^s(y') > \Lambda^{2n} r(W)/C_g > A \cdot r(W)$$

by the definition of n. If, on the other hand, the stable manifold of y' is cut by the singularities of T^{-n}, then the stable manifold of y should extend to $\Gamma^{(n)}$. Thus, recalling the definition of U from (5), we have that the stable manifold of y reaches ∂U. Now by our first requirement on $B = B_R(x)$, for such a y:

$$r^s(y) \geq \rho(y, \partial U) \geq \rho(B, \partial U) > KAR \geq A \cdot r(W)$$

where we have also used $r(W) \leq KR$ which follows from $W \subset B$.

Summarizing, we have $r^s(y) > A \cdot r(W)$ whenever $y \in W_g$, which, along with $m_W(W_g) > 0$, completes the proof of the Theorem. □

The proof of Theorem 3.4 goes along the same lines, based on Lemma 3.6.

4 Fundamental theorem implies local ergodicity

Theorem 4.1. *Suppose that a dynamical system satisfies assumtions A1 ... A8. Then there is an n_0 such that any $x \in M \setminus (\Gamma^{(n_0)} \cup \Gamma^{(-n_0)})$ has an open neighborhood U_x, μ-a.e. point of which belongs to a single ergodic component.*

Before the proof we recall the Hopf construction of ergodic theory. This is the most common tool for proving (local) ergodicity in hyperbolic systems. On further details see eg. [BS] (Chapter 6), or [CM] (Chapter 6) in the billiard context.

Throughout the section "almost everywhere" (a.e.) is understood w.r.t μ or w.r.t Lebesgue measure. By equivalence (Assumption A2) these two notions coincide. Given a LUM (LSM) W, a.e. on W means a.e. w.r.t the induced Lebesgue measure on W.

Given a continuous function $f : M \to \mathbb{R}$ define $f^+(x) = \lim_{n\to\infty} \frac{f(x)+f(Tx)...+f(T^{n-1}x)}{n}$ and $f^-(x) = \lim_{n\to\infty} \frac{f(x)+f(T^{-1}x)...+f(T^{-n+1}x)}{n}$, the future and past time avarages. We will say that $x \in M$ is f-typical if $f^+(x)$ and $f^-(x)$ exist and coincide. By Birkhoff's ergodic theorem a.e. point on M is f-typical. Furthermore, by uniform continuity of f (M is assumed to be compact), f^+ is constant along LSMs and f^- is constant along LUMs.

Consider two f-typical points, y_1 and y_{K+1}. We will say that the alternating sequence of LUMs and LSMs W_i, $i = 1, ..., K$ is *a Hopf chain for f, connecting y_1 to y_{K+1}* if $y_1 \in W_1$ and $y_{K+1} \in W_K$, W_{2k-1} is a LUM and W_{2k} is a LSM, and for every i the intersection $W_{i-1} \cap W_i = \{y_i\}$ consists of exactly one point y_i which is f-typical. We will say that the Hopf chain starts from y_1 and connects y_1 to y_{K+1}. By the above facts the time averages exist, coincide and are the same for the interseciton points: $\bar{f} = f^+(y_i) = f^-(y_i)$, for $i = 1, ..., (K+1)$, and $\bar{f}$ is independent of i. In particular, the following Lemma can be proved by standard ergodic theory arguments (see again [BS] or [CM]):

Lemma 4.2. *Consider a set $\tilde{M} \subset M$ of positive μ-measure. If for any continuous function f there exists a set $\tilde{M}_f \subset \tilde{M}$ of full μ-measure, such that every pair of points $y_a, y_b \in \tilde{M}_f$ can be connected by a Hopf chain (for f), then $\tilde{M}$ belongs (mod zero μ-measure) to one ergodic component.*

Keeping the continuous function f fixed we recall some further definitions from [BS] (or [CM]). We will say that a point $x \in M$ is f-good if (i) it is f-typical, (ii) $r^u(x) > 0$ and $r^s(x) > 0$, (iii) a.e. point $y \in W^u(x)$ is f-typical and satisfies $r^s(y) > 0$, and (iv) a.e. point $y \in W^s(x)$ is f-typical and satisfies $r^u(y) > 0$.

Absolute continuity (Assumption A7), along with the a.e. existence of LUMs and LSMs (Assumption A4) imply the following lemma (see e.g. [CM]):

Lemma 4.3. *A.e. $x \in M$ is f-good. Furthermore, for an f-good $x \in M$, a.e. point of its LUM $W^u(x)$ (and of its LSM $W^s(x)$) is, in addition to being f-typical, also f-good.*

In the proof of local ergodicity we will construct Hopf chains through intersection points which are all f-good.

Proof of Theorem 4.1. Let us fix $A > \max(1000, 100/C_{\delta_{us}})$. Choose $n_0 = n_{A^3}$. For $x \in M \setminus (\Gamma^{(n_0)} \cup \Gamma^{(-n_0)})$ we define below the "big neighborhood" B_x and the "small neighborhood" U_x. Let $B_x = B_R(x)$ be the open ball for which both forms of the fundamental theorem, Theorem 3.1 and Theorem 3.4 hold (with $A \to A^3$, and thus with $A \to A'$ whenever $1 < A' \le A^3$). Take $R_0 = R/A^5$ and fix $U_x = B_{R_0}(x)$ which becomes this way a tiny ball in the middle of B_x.

We will show that the assumptions of Lemma 4.2 hold for $\tilde{M} = U_x$, using the fundamental theorem several times with A, and once with A^3, always in B_x.

Fix a continuous function f and two f-good points $y_a, y_b \in U_x$.

Let us fix $y_1 = y_a$. There exists $0 < R_1$ so small that $r^u(x) \ge R_1$, and thus $W_1 = W^u_{R_1}(y_1)$ exists. We may also assume, for simplicity, that this manifold is contained in U_x. As y_1 is f-good, almost every point of W_1 is f-good. Fixing $R_2 = AR_1$, by the Fundamental Theorem 3.1 (applied in B_x with A), $r^s(y) > R_2$ for a positive measure set of points y on W_1. These two observations imply that there is an f-good point $y_2 \in W_1$ such that $W_2 = W^s_{R_2}(y_2)$ exists and, by the choice of B_x and U_x, is necessarily contained in B_x. We may continue this procedure: applying consecutively either Theorem 3.1 or Theorem 3.4, along with the fact that a.e. point on the consecutive manifolds is f-good, we construct a Hopf chain starting from y_a with (un)stable manifolds of increasing size: $W_i = W^{\S}_{R_i}(y_i)$, where $R_i = A^{i-1}R_1$, $W_{i-1} \cap W_i = \{y_i\}$ is a good point, and $\S = u$ or $\S = s$ depending on the parity of i. We only need to ensure that $W_i \subset B_x$, which remains true as long as $\rho(y_i, U_x) < A^4 R_0$ and $R_i < A^4 R_0$. This definitely holds throughout the process if we stop at the first occasion when $R_i > 100R_0$. Let us denote this final manifold in our Hopf chain as $W' = W^u_{R_a}(z_a)$.[1]

We can repeat the same Hopf chain construction starting from y_b. We arrive at two Hopf chains: one starting from y_a and terminating in $W' = W^u_{R_a}(z_a)$, the other starting form y_b and terminating in $W'' = W^u_{R_b}(z_b)$. If we manage to connect the LUMs W' and W'' via a stable manifold in such a way that the intersection points with both LUMs are f-good, the proof is complete.

Note we have $W', W'' \subset B_x$ and $R_a, R_b > 100R_0$. We may assume $R_a > R_b$. By the choice of A we also have $\rho(z_a, z_b) < R_a C_{\delta_{us}}$.

Fix $W = W^u_{R_0}(z_b)$. Then Lemma 2.3 implies that W' s-shadows W.

Note that $r(W) = R_0$ and $r(W') < A100R_0$, thus $A^3 r(W) > 2r(W')$. Applying Theorem 3.1 (with $A \to A^3$) we have $r^s(z) > 2r(W')$ for a positive measure set of points $z \in W$. As W' s-shadows W, these stable manifolds necessarily intersect W'. Furthermore, as a.e. point (both on W and on W') is f-good, recalling also the absolute continuity property (Assumption A7), we can choose a LSM such that the intersection points are f-good. This guarantees we have connected y_a to y_b with a Hopf chain, which completes the proof of Theorem 4.1. $\square$

As a corollary, we can prove ergodicity of those multi-dimensional dispersing billiard

[1] If we happen to exceed $100R_0$ at a stable step, apply the fundamental theorem with $A \to 1{,}0001$ once.

maps, for which the growth property of Assumption A8 could be checked in [BT]. For details of the conditions of this theorem, please consult the above paper.

Theorem 4.4. *Let (M, T, μ) be the billiard map for a multi-dimensional billiard. Assume the surface of every scatterer is C^3-smooth, and its second fundamental form (curvature) is positively denifed everywhere (the billiard is dispersing). Assume that the horizon is finite (the free flight is bounded), and that the scatterers are disjoint (there are no corner points). Assume furthermore that the singularities of the map satisfy the* subexponential complexity condition – *that is,*

$$K(n) = \sup_{x \in M} \left(\text{ number of connected components } M_i \text{ of } M \setminus \Gamma^{(n)} \text{ with } x \in \overline{M_i} \right)$$

grows subexponentially in n. Then (M, T, μ) is ergodic. Moreover, any higher iterate of the map, (T^n, μ) is also ergodic.

Remark 4.5. *Note that as multi-dimensional dispersing billiards are hyperbolic systems (actually, uniformly hyperbolic, that is, Assumption A3 is satisfied) ergodicity of all the iterates automatically implies K-mixing ([SCh],[KS]) and the Bernoulli property ([ChH], [OW]). Ergodicity of every higher iterate of the map is the property required in Assumption A.3 of [BT].*

Proof. The proof is based on [BT], where it is shown that the billiard maps under consideration satisfy assumptions A1 ... A8. Thus the statement of Theorem 4.1 applies.

Let M^0 be the set of those phase points $x \in M \setminus \partial M$, which are at most once singular along their entire trajetory. Points of $M \setminus \partial M \setminus M^0$ are the phase points which are at least twice singular. These are contained in a countable union of two-codimensional submanifolds, so they cannot "cut" the phase space. That is, the connected components of M^0 coincide with the connected components of M, modulo a set of zero measure. We claim that *every* point of M^0 has an entire open neighbourhood which belongs to a single ergodic component. This implies that every connected component of M can intersect only one ergodic component.

First, consider an $x \in M^0$ which is never singular – that is, $x \notin \Gamma_n$ for any $n \in \mathbb{Z}$. Then Theorem 4.1 directly applies, to give the neighbourhood of x we are looking for.

Second, consider an $x \in M^0$ which is singular exactly once, in the future – that is, $x \in \Gamma_n$ for a single $n \geq 1$. Then we can apply Theorem 4.1 to $T^{-k}x$ with $k + n > n_0$ to get an open neighbourhood of $T^{-k}x$ which belongs to a single ergodic component. Applying T^k to that neighbourhood gives the desired neighbourhood of x, since T^{-k} is continuous (with a continuous inverse) in a (sufficiently small) neighbourhood of x.

Third, consider an $x \in M^0$ which is singular exactly once, in the past – that is, $x \in \Gamma_{-n}$ for a single $n \geq 1$. Then we can apply Theorem 4.1 to $T^k x$ with $k + n > n_0$ to get a neighbourhood of $T^k x$, then apply T^{-k} to get the desired neighbourhood of x.

Since every point of M^0 is covered in one of the three cases, we have proven our claim, and conclude that the ergodic components of M consist of entire connected components of M.

Now to get ergodicity, we only need to see that different connected components of M belong to the same ergodic component. But connected components of the billiard map correspond to scatterers of the billiard, and these scatterers obviously cannot be partitioned into two sets so that no travel from one set to the other is possible. So if there would exist several ergodic components, two of them would certainly contain scatterers which are "neighbours" in the sense that a positive measure set of points on one of them is mapped to the other by T, which means that these sets cannot be invariant – a contradiction.

All that is left is to prove ergodicity for every power T^n of the dynamics. Since (un)stable manifolds of T are also (un)stable manifolds of T^n, the Hopf chains are the same, so we still know that every connected component of the phase space (corresponding to every scatterer) belongs to a single ergodic component. So, to see ergodicity of every T^n, we only need to check that the scatterers cannot be grouped into two or more classes so that transition is only possible from every class to the next one (and from the last one to the first). This is true for geometrical reasons: one can always find three scatterers A, B and C so that transition (of a positive measure set of points) is possible from A to B, form B to C, form C to A and from B to A as well – so the lenght of any cycle must be a divisor of both 2 and 3. For a more precise formulation and a detailed proof of this statement see [CM], Lemma 6.23. The setting there is two-dimensional, but the statement and the proof remain unchanged in the multi-dimensional case. □

Remark 4.6. *As a closing remark, let us comment on the above mentioned "subexponential complexity" condition. There is no doubt in the billiard community that such a condition should be generic in the set of all finite horizon billiard systems, in any reasonable sense of genericity. There is a sketch of proof for such a statement in [B], however, the issue is definitely subject to further investigation.*

Acknowledgements

We are grateful to the anonymus referee for many useful comments which helped us to clarify certain details, especially related to the alignment property (our Assumption A5). The financial support of Hungarian National Foundation for Scientific Research (OTKA), grants TS040719, T046187, TS049835 and F60206 is thankfully acknowledged. P. Bálint is grateful for the support of the Bolyai scholarship of the Hungarian Academy of Sciences. Part of the work was done during the visit of P. Bachurin to Budapest in December 2005, supported by Princeton University and the Technical University of Budapest.

References

[BChSzT1] P. Bálint, N. I. Chernov, D. Szász and I. P. Tóth, *Geometry of Multi-dimensional Dispersing Billiards*, Astérisque **286** (2003) 119–150.

[BChSzT2] P. Bálint, N. I. Chernov, D. Szász and I. P. Tóth, *Multidimensional Semi-dispersing Billiards: Singularities and the Fundamental Theorem*, Annales Henri Poincaré **3** 451 – 482 (2002)

[BT] P. Bálint and I.P. Tóth, *Exponential Decay of Correlations in Multidimensional Dispersing Billiards*, preprint, available at `http://www.renyi.hu/~bp/pub.html`

[B] P. Bachurin; *On the structure of the singularity manifolds of dispersing billiards*, preprint, available at http://www.arxiv.org, math.DS/0505620

[BS] M. Brin and G. Stuck; *Inroduction to Dynamical Systems*, Cambridge University Press, Cambridge, 2002.

[Ch1] N. Chernov; *Local ergodicity of hyperbolic systems with singularities*, Funct. Anal. Appl. **27**, 51–54 (1993)

[Ch2] N. Chernov; *Statistical Properties of Piecewise Smooth Hyperbolic Systems in High Dimensions*, Discrete and Continuous Dynamical Systems, **5**, 425–448, (1999).

[Ch3] N. Chernov, *Decay of correlations and dispersing billiards*, J. Statist. Phys. **94** (1999), 513–556.

[ChH] N. Chernov and N. Haskell, *Non-uniformly Hyperbolic K-systems are Bernoulli*, Ergod. Th. Dyn. Sys. **16** (1996), 19–44

[CM] N. Chernov and R. Markarian, *Chaotic billiards*, to be published in the Mathematical Surveys and Monographs by the AMS.

[KS] A. Katok and J.-M. Strelcyn; *Invariant manifolds, entropy and billiards; smooth maps with singularities*, Lecture Notes in Mathematics, Springer, **1222**, 1986

[KSSz] A. Krámli, N. Simányi, D. Szász *A "transversal" fundamental theorem for semi-dispersing billiards.* Comm. Math. Phys. **129** (1990), 535–560.

[LW] C. Liverani and M. Wojtkowski, *Ergodicity in Hamiltonian Systems*, Dynamics Reported (New Series) **4** (1995), 130–202.

[OW] D. Ornstein and B. Weiss; *On the Bernoulli nature of systems with some hyperbolic structure* **18** (1998), 441–456

[S] Ya. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russian Mathematical Surveys **25** (1970), 137–189.

[SCh] Ya. G. Sinai and N. Chernov, *Ergodic Properties of Certain Systems of 2–D Discs and 3–D Balls.*, Russian Mathematical Surveys, **(3) 42** (1987), 181–201.