Exponential Decay of Correlations in Multi-dimensional Dispersing Billiards

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 8, 2007

Abstract

We prove exponential decay of correlations for a "reasonable" class of multi-dimensional dispersing billiards: the scatterers are required to be C^3 smooth, the horizon is finite, there are no corner points and we assume subexponential complexity of the singularity set.

Introduction

In this paper we address statistical properties of multi-dimensional billiards. We restrict to dispersing billiards with finite horizon and no corner points. It is a long standing conjecture that correlations in these systems decay exponentially fast. This is what we are going to prove, modulo an extra assumption on the combinatorical structure of the singularity set – the so-called *sub-exponential complexity condition*, see sections 1 and 2.4.2 for a precise formulation.

Exponential decay of correlations for billiard maps were first obtained by Young in [Y] where she established the property for the two dimensional analogues of our dispersing billiards. At the same time this paper provided a powerful method, the so-called "Young tower construction" to study statistical properties for a large class of hyperbolic systems. Shortly Chernov in [Ch2] extended the result to further two dimensional dispersing billiards (allowing infinite horizon and corner points). In addition, Chernov's paper provided a list of assumptions that guarantee the framework of [Y] to work. Actually, Chernov's paper is our main reference: we prove exponential decay of correlations by verifying that the assumptions of [Ch2] all hold in the studied multi-dimensional dispersing billiard systems.

Witnessing the success in the two dimensional case, the billiard community was quite optimistic about extending these results to the multi-dimensional setting in the late nineties. However, the discovery of [BChSzT1] about the pathological behaviour of singularity manifolds in multi-dimensional billiards emerged as a serious obstacle in the proof of exponential decay of correlations. In addition, these phenomena called even for a reconsideration of earlier proofs of *ergodicity* in multi-dimensional (semi-)dispersing billiards. The papers [BChSzT2] and [B] handle the problem of (local) ergodicity for certain special cases: [BChSzT2] deals with the case of algebraic scatterers, while [B] treats strictly dispersing billiards with highly smooth scatterers (smoothness depending on the dimension of the billiard domain), and a strong condition (finite complexity) on the singularities.

It is important to mention the issue of ergodicity, as the billiards we consider are not contained in the above mentioned classes. We require only C^3 smoothness of the scatterers, and sub-exponential complexity (instead of the more restrictive finite complexity condition of [B]). Thus, a priori we do not even know if the studied systems are ergodic. However, ergodicity of the map (and its higher iterates) is, naturally, among the assumptions of [Ch2].

In the present paper we give a detailed local analysis of multi-dimensional dispersing billiards. This allows us to check all the assumptions of [Ch2] except for the ergodicity of the map. However, in [BBT], a separate paper joint with Bachurin, we show that the same class of billiards are ergodic. The relation of this paper to [BBT] is twofold. On the one hand, [BBT] uses the results of the detailed local analysis (in particular, the growth properties) presented here to prove ergodicity. On the other hand, the fact that our systems are ergodic – as established in [BBT] – completes the set of assumptions, and thus the proof of expontial decay of correlations à la Chernov and Young ([Ch2] and[Y]).

Before closing the introduction let us comment on the above mentioned "subexponential complexity" condition. First it is worth noting that conditions of this sort are standard in the studies of hyperbolic systems with singularities, see for example [Y], [Ch1] or [S]. In particular, subexponential complexity holds in *every two dimensional* dispersing billiard system with finite horizon and no corner points, cf. [CY]. Actually, this is a crucial fact in the proofs of [Y] and [Ch2] on the exponential mixing of the two dimensional case.

As to the multi-dimensional situation, 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.

1 Statement of result and structure of the proof

Let us consider a billiard domain in the d-dimensional flat torus $\mathbb{Q} \subset \mathbb{T}^d$, and a point particle that travels uniformly (follows straight lines with constant speed) within $\mathbb{Q}$, and bounces off the boundary (the scatterers) via elastic collisions (angle of incidence is equal to the angle of reflection). We will concentrate on the case of $d \geq 3$ and require some further properties.

Assumption 1.1. *The* ***boundary*** $\partial\mathbb{Q}$ *is assumed to be a finite collection of compact* $d-1$ *dimensional* $\mathbf{C^3}$***-smooth*** *submanifolds in* $\mathbb{T}^d$. *This implies, in particular, that it is possible to define the curvature operator, or second fundamental form* K *in any point of* $\partial\mathbb{Q}$. K *should be understood as the second fundamental form for the relevant one codimensional submanifold(s) with (unit) normal vectors pointing inward* $\mathbb{Q}$.

The billiard is ***stricly dispersing****. That is, the boundary components, as viewed from the exteriour, are stricly convex. In other words, the operator* K *is positive definite on* $\partial\mathbb{Q}$.

Note that, by compactness, smoothness and strict dispersivity, we have that the spectrum of the symmetric positive definite operator K is bounded away both from 0 and ∞ on $\partial\mathbb{Q}$. With a slightly sloppy notation this can be expressed as:

$$K \geq K_{min}, \tag{1.1}$$

and

$$K \leq K_{max}, \tag{1.2}$$

where $0 < K_{min} \leq K_{max} < \infty$ are constants depending only on the billiard domain.

To formulate our second main assumption, consider $q \in \partial\mathbb{Q}$ (a configuration point) and $v \in \mathbb{S}^{d-1}$ with $\langle n(q), v\rangle \geq 0$ (a velocity), where $n(q)$ is the unit normal vector of $\partial\mathbb{Q}$ and the condition on v means that we are considering an outgoing velocity. The pair $x = (q, v)$ is a phase point in our dynamical system (on further details see Section 2 below). Given x we may consider the *free flight function*: $\tau(x)$ measures the distance along the straight line that starts out of q in the direction of v untill it reaches $\partial\mathbb{Q}$ again.

Assumption 1.2. *We assume that* ***the horizon is finite****; there is a positive constant* $\tau_{max} < \infty$, *depending only on the billiard domain, such that for any phase point* $x = (q, v)$

$$\tau(x) \leq \tau_{max}. \tag{1.3}$$

We assume that ***there are no corner points****; that is, the smooth components of the boundary are disjoint (i.e. we require smoohtness, not only piecewise smoothness of* *the boundary). This implies, on the basis of compactness, that the free flight function is bounded from below: there exists a constant $\tau_{min} > 0$, depending only on the billiard domain, such that for any phase point $x = (q, v)$*

$$\tau(x) \geq \tau_{\min}. \tag{1.4}$$

The only additional assumption we need is more technical, thus we postpone the precise formulation to Section 2.4.2. Here we only give a preliminary description. It is well-known (see eg. Section 2.2.3 below) that the billiard map is discontinuous. We may consider the components of the phase space restricted onto which the map is continuous. The components for the higher iterates of the map can be defined similarly. The following quantity is of crucial importance: for any poistive integer n, let us denote by K_n the maximum number of the components of the nth iterate of the billiard map that can meet in a single phase point (for a precise formulation see Definiton 2.7).

Assumption 1.3. *We assume that the* ***complexity of the singularity set grows sub-exponentially*** *with n. That is, $K_n = \mathcal{O}(\lambda^n)$ holds for any $\lambda > 1$.*

Remark 1.4. *Actually, it is enough to require $K_n = \mathcal{O}(\lambda^n)$ for some $\lambda > 1$ smaller then the minimum expansion along unstable vectors, see Section 2.4.2.*

To formulate our main result, let us recall what is ment by "exponential decay of correlations".

Definition 1.5. *Consider a phase space M with a dynamics T and a T-invariant probability measure μ. We say that the dynamical system (M, T, μ) has exponential decay of correlations (EDC) , if for every $f, g : M \to \mathbb{R}$ Hölder-continuous pair of functions there exist constants $C < \infty$ and $a > 0$ such that for every $n \in \mathbb{N}$*

$$\left| \int_M f(x) g(T^n x) d\mu(x) - \int_M f(x) d\mu(x) \int_M g(T^n x) d\mu(x) \right| \leq C e^{-an}.$$

Theorem 1.6. *Consider a multidimensional dispersing billiard map that satisfies the Assumtions 1.1, 1.2 and 1.3. Then the dynamics enjoys exponential decay of correlations.*

Remark 1.7. *It is worth noting that, as a standard corollary of Theorem 1.6, the central limit theorem is also satisfied – see [Ch2] and [Y].*

As already mentioned in the introduction, we prove Theorem 1.6 by veryfing the assumptions of a Theorem from [Ch2] that guarantees exponential decay of correlations. To make our exposition more self-contained, we present Chernov's conditions separately in Appendix A.

The rest of the paper, which provides the verification of these assumptions, is organized as follows. In Section 2 we collect the most important prerequisites on multidimensional dispersing billiards. In particular, we recall that certain conditions from

Appendix A: Conditions A.1, A.2, A.4, A.5 and A.6 have already been proven for the studied billiards in [BChSzT1]. In addition, we formulate and prove some further properties of similar flavour, to be applied in the later sections, which – to our knowledge – have not been considered before.

The verification of Condition A.7 is the main novelty of our paper. To achieve this we are led to use a Riemannian structure different from the traditional "Euclidean" metric on the billiard phase space. However, the properties discussed in Section 2, in particular, conditions A.1, A.2, A.4, A.5 and A.6 are originally proved for the "Euclidean" metric. Thus it is to be verified that these properties remain valid with the use of the new Riemannian structure. This is the content of Section 3, which is of differential geometric nature.

Condition A.7 on the "growth properties of unstable manifolds" is proved in Section 4.

As already mentioned, Condition A.3, more precisely, the ergodicity of the map (and its iterates) is the only conditon that we do not prove in the present paper. For the proof of ergodicity we refer to [BBT]. Some explanation is given in Section 4.5. This completes the proof of Theorem 1.6.

In addition to Appendix A, we have also included Appendix B which contains some simple lemmas of geometric measure theory.

2 Preliminaries

In this section we repeat notions and statements from [BChSzT1] and [Ch2]. These will be referred to in several parts of the discussion. Although some of the referred statements are only found in the text of [BChSzT1] and are not highlighted as theorems, we do not repeat the proofs here. Instead, we give the precise references within the paper. Our aim with listing these statements is to collect all facts about billiards that are used from earlier works in one place, and keep the paper otherwise self-contained.

Notation 2.1. *Throughout the paper we will use the following conventions:*

Positive and finite global constants whose value is unimportant, will be denoted by c or C. So e.g. $f < C$ means that the function f is bounded from above The letters c and C may denote different values in different equations.

On the other hand, C_1, C_2, ... will denote global constants whose values are the same throughout the paper.

We say that two nonzero functions f and g on the phase space are equivalent, if $c \le \frac{f}{g} \le C$. In this case we will use the notation $f \sim g$.

2.1 Chernov's conditions

In his paper [Ch2] Chernov has proven a theorem that guarantees exponential decay of correlations, provided that we can check that the dynamical system satisfies certain conditions about hyperbolicity, regularity of unstable manifolds and the dynamics on them, and about the growth of unstable manifolds. A dynamical system that can be handled in this way should consist of

- A phase space M, which is the set of possible states of the system.
- A dynamics T which is an $M \to M$ map, or at least a map defined on a large subset of M.
- A measure μ on M which is T-invariant.
- *A Riemannian structure* on M, so that M (or $\bar{M}$) is a Riemannian manifold, possibly with boundary. The level of smoothness of maps and subsets of this Riemannian manifold is an important issue.

Instead of just referring to the work [Ch2], we will list the conditions and repeat the statement in the Appendix. This is done mainly to make our paper easier to read, but also to point out two minor details where we use modified versions of Chernov's conditions. Both modifications allow the proof in [Ch2] to remain unchanged.

2.2 The dynamical system

In accordance with Chernov's conditions, we will now describe our choice of phase space, dynamics and measure. However, we postpone our choice of the Riemannian structure until Section 3, since this is a key point of our proof, and certainly doesn't belong to the Preliminaries section.

2.2.1 the Poincaré section phase space and the dynamics

A billiard is the dynamical system describing the motion of a point particle in a domain $\mathbb{Q} \subset \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, which moves uniformly inside $\mathbb{Q}$, and bounces back elastically from the boundary of $\mathbb{Q}$. We require that $\mathbb{Q}$ be connected and closed, and that its boundary $\partial\mathbb{Q}$ consist of finitely many C^3-smooth 1-codimensional submanifolds of the flat torus $\mathbb{T}^d$.

In this paper we only consider dispersing billiards, that is, the boundary components are strictly convex from the outside. Actually, in the present work we even suppose that the C^3-smooth components of $\mathbb{Q}$ are disjoint: that is, there are no corner points of the billiard table. So our table $\mathbb{Q}$ could well be viewed as $\mathbb{Q} = \mathbb{T}^d \setminus \bigcup_{i=1}^s O_i$, where the $\bar{O}_i$ are disjoint obstacles with C^3-smooth boundary, e.g. spheres.

First we describe the flow phase space $\mathcal{M}$. This consists of all possible positions of the particle, equipped with the possible unit velocities:

$$\mathcal{M} = \left\{(q, v) \mid q \in \mathbb{Q}, v \in \mathbb{R}^d, \|v\| = 1\right\}.$$

We could identify phase points with opposite velocities and the same configuration point on $\partial\mathbb{Q}$, but this is not important for our purposes. We will only use the flow phase space at those occasions, when it is important to view the Poincaré section phase space as a submanifold of $\mathcal{M}$.

So now we describe the usual Poincaré section phase space $\tilde{M}$. Our phase space M will be a subset of this. In $\tilde{M}$, we only consider collision moments. Since kinetic energy is preserved in the system, we also fix the speed to be 1. We choose to describe the motion of the particle at a collision time by recording its velocity just *after* the collision (we use the 'outgoing' Poincaré section).

So a possible state of the particle is described by giving a boundary point $q \in \partial\mathbb{Q}$ and a unit velocity $v \in S^{d-1}$, which is often written roughly as $\tilde{M} = \partial\mathbb{Q} \times S_+^{d-1}$, where the $+$ indicates that only velocity vectors pointing inward $\mathbb{Q}$ are allowed. However, this is misleading, because $\tilde{M}$ really doesn't have a product structure. So we better write (still roughly)

$$\tilde{M} = \left\{(q, v) \mid q \in \partial\mathbb{Q}, v \in \mathbb{R}^d, \|v\| = 1, \langle v, n(q)\rangle \geq 0\right\}.$$

Notation 2.2. *$n(q)$ denotes the (unit) normal vector of $\partial\mathbb{Q}$ at q pointing inward $\mathbb{Q}$.*

$T : \tilde{M} \to \tilde{M}$ gives the state of the particle at the next collision as a function of the present state. When we apply the theorem of [Ch2], we will apply it to some higher iterate T^{n_0} of this dynamics. n_0 will be given later.

Notation 2.3. *If some quantity (e.g. q or n) is related to some phase point x, then we will often denote the the corresponding quantity related to Tx by the same letter and an index $_1$ (e.g. q_1, n_1).*

2.2.2 the invariant measure and the Euclidean metric

The natural Riemannian structure on $\tilde{M}$ is described in detail in many works including [BChSzT1], and we don't repeat those details here. We only mention that $\tilde{M}$ is viewed as a semisphere-bundle over $\partial\mathbb{Q}$, and the different velocity semispheres (fibres) at nearby configuration points are identified using the *parallel transport from the natural Riemannian structure of* $\partial\mathbb{Q}$ as a submanifold of $\mathbb{T}^d$. This results in the definition of a C^2 atlas on $\tilde{M}$ and a local product structure. Since we need to handle differential aspects of the dynamics, it is crucial to understand how M is embedded in $\mathcal{M}$.

Now the natural Riemannian structure is locally a product of the natural Riemannian structures on $\partial\mathbb{Q}$ and S^{d-1}. We will call this natural structure the "Euclidean structure". This is a little misleading, since this is really Euclidean only if $d = 2$, while it is not flat in higher dimensions. However, it still resembles a Euclidean structure in the sense that the induced norm (which is called the "Euclidean norm" in [BChSzT1]) has the form $\|(dq, dv)\|_e^2 = dq^2 + dv^2$, where $dq \in \mathcal{T}$ [1] is a tangent vector of the configuration space $\partial\mathbb{Q}$ and $dv \in \mathcal{J}$ is a tangent vector of the velocity hemisphere S_+^{d-1}. With this Riemannian structure, $\tilde{M}$ becomes a C^1 Riemannian manifold with boundary, and the boundary is

$$\partial\tilde{M} = \left\{(q, v) \mid q \in \partial\mathbb{Q}, v \in \mathbb{R}^d, \|v\| = 1, \langle v, n(q)\rangle = 0\right\},$$

[1] $\mathcal{T}$ and $\mathcal{J}$ will be introduced in Section 2.3.

the set of tangential collisions.

Notation 2.4. *For $x = (q, v) \in \tilde{M}$, $\varphi(x)$ will denote the angle of the velocity and the normal vector of the scatterer: $\varphi(x) = \sphericalangle(n(q), v)$, $\langle n(q), v\rangle = \cos\varphi(x)$. In two dimensions one often treats φ as a signed quantity, but for us, $0 \le \varphi \le \frac{\pi}{2}$.*

With this notation,

$$\partial\tilde{M} = \left\{(q, v) \in \tilde{M} \mid \varphi(x) = \frac{\pi}{2}\right\}.$$

The natural T-invariant measure μ on $\tilde{M}$ is defined in terms of the natural Riemannian structure: let μ be absolutely continuous with respect to the induced measure (which we will call the Lebesgue measure), and let the density be *const* $\cos\varphi(x)$ where *const* is a normalizing constant so that μ is a probability. This can vaguely be written as

$$d\mu = const\ \cos\varphi\ dq\ dv, \tag{2.1}$$

since the Lebesgue measure is locally a product of the natural (surface volume) measures on $\partial\mathbb{Q}$ and S^{d-1}.

Remark 2.5. *Note that the choosing to use "outgoing" velocities in the Poincaré map phase space (which is the usual choice) is just a matter of notation. We could as well identify every incoming velocity vector with the corresponding outgoing one, and view the outgoing velocity as a representative of the equivalence class. Accordingly, there is no asymmetry in the definition of the Euclidean metric: the Euclidean metric of the inverse dynamics is the same. As a consequence, replacing velocities with their opposites (up to identification of incoming/outgoing) leaves the measure invariant. This fact will be reflected by the formulas (2.6), (2.7), (2.16) and (2.17).*

2.2.3 singularities and the phase space we use

T is not continuous in the points of $\mathcal{S} = T^{-1}\mathcal{S}^0 = T^{-1}\partial\tilde{M}$, which we call the primary singularity set. So, to satisfy Condition A.1 we have to exclude $\mathcal{S}$ from the domain of T. Moreover, the derivative of T blows up near $\mathcal{S}$, which causes Condition A.5 to fail, unless we declare certain points of $\tilde{M}$ (which are close) to be separated by "artificial boundaries". This is done in the way usual in billiard theory. We partition the original phase space into infinitely many *homogeneity layers*:

$$\begin{aligned}
I_k &= \left\{x \in \tilde{M} \mid \frac{1}{k^2} < \frac{\pi}{2} - \varphi(x) < \frac{1}{(k-1)^2}\right\} \text{ for } k = k_0+1, k_0+2, \dots \text{ and} \\
I_{k_0} &= \left\{x \in \tilde{M} \mid \frac{1}{k_0^2} < \frac{\pi}{2} - \varphi(x)\right\}
\end{aligned} \tag{2.2}$$

Here the integer constant k_0 is arbitrary, and will be chosen later. The boundary of this partitioned phase space is

$$\Gamma^0 = \bigcup_{k=k_0}^{\infty} \Gamma_k^0$$

where

$$\Gamma_k^0 = \left\{ x \in \tilde{M} \mid \frac{\pi}{2} - \varphi(x) = \frac{1}{k^2} \right\}, \quad k = k_0, k_0+1, \ldots .$$

Correspondingly, the countably many manifolds in the set

$$\Gamma = T^{-1}\Gamma^0 \tag{2.3}$$

are the so called *secondary singularities.*

Now we can define the phase space we will use:

$$M = \tilde{M} \setminus (\mathcal{S}^0 \cup \Gamma^0) = I_{k_0} \cup \bigcup_{k=k_0+1}^{\infty} I_k,$$

where the components I_k are meant to have disjoint closures – as if they were moved apart from each other. This makes M non-compact, but many compactness arguments remain valid if we temporarily forget about artificial boundaries – that is, we look at $\tilde{M}$. On the other hand, regularity properties may be easier, specifically *distortion bounds* depend on secondary singularities.

2.3 Hyperbolicity, cones and fronts

In billiard theory, several basic constructions and concepts are based on the notion of a local orthogonal manifold, which - for simplicity - we will call front. A front $\mathcal{W}$ is defined in the flow phase space $\mathcal{M}$ rather than in the Poincaré section. Take a smooth 1-codimemsional submanifold E of the flow configuration space $\mathbb{Q}$, and add the unit normal vector $v(r)$ of this submanifold at every point r as a velocity, continuously. Consequently, at every point the velocity points to the same side of the submanifold E. Then

$$\mathcal{W} = \{(r, v(r)) | r \in E\} \subset \mathcal{M},$$

where $v : E \to \mathbb{S}^{d-1}$ is continuous (smooth) and $v \perp E$ at every point of E. The derivative of this function v, called B plays a crucial role: $dv = Bdr$ for tangent vectors (dr, dv) *of the front.* B acts on the tangent plane $\mathcal{T}_r E$ of E, and takes its values in the tangent plane $\mathcal{J} = \mathcal{T}_{v(r)}\mathbb{S}^{d-1}$ of the velocity sphere. These are both naturally embedded in the configuration space $\mathbb{Q}$, and can be identified through this embedding. So we just write $B : \mathcal{J} \to \mathcal{J}$. B is nothing else than the curvature operator – or second fundamental form – of the submanifold E of $\mathbb{Q}$. Clearly, this is different from the curvature of $\mathcal{W}$ as a submanifold of $\mathcal{M}$. Obviously, B is symmetric.

Notice that fronts remain fronts during time evolution - at least locally, and apart from some singularity lines.

When we talk about a front, we sometimes think of it as the part of the (flow) phase space just described (for example, when we talk about time evolution under the flow), but sometimes just as the submanifold E (for example, when we talk about the tangent space or the curvature of the front). This should cause no confusion.

In the rest of this section we list technical details about the evolution of fronts and the construction of invariant cone fields required by the Hyperbolicity condition (Condition A.2). Later we will only use these details in two places:

- in Section 2.5.2 to understand the nature of anisotropic expansion of unstable vectors
- in sections 3.1 and 3.2 where we introduce a new Riemannian structure and show that T is uniformly hyperbolic w.r.t. this Riemannian structure in the sense of Condition A.2.

2.3.1 Evolution of fronts

During free propagation (that is, from one collision to the other) a tangent vector (dr^+, dv^+) of the post-collision front evolves into the tangent vector (dr_1^-, dv_1^-) of the pre-collisional front at the next collision given by the formulas

$$\begin{aligned} dr_1^- &= dr^+ + \tau dv^+, && (2.4)\\ dv_1^- &= dv^+ && (2.5)\end{aligned}$$

where τ is the length of the free run between the two collisions.

For this formula – and the next one – to make sense, we need to identify the tangent planes of the front at different moments of time. Let $\mathcal{T} = \mathcal{T}_r\partial\mathbb{Q}$ be the tangent plane of the scatterer at a collision point r. Just like $\mathcal{J}$, $\mathcal{T}$ is viewed together with its natural embedding into $\mathbb{Q}$. The identification of different $\mathcal{J}$'s is done in the usual way (cf. [SCh], [KSSz]):

- by translation parallel to v from one collision to the other,
- by reflection with respect to $\mathcal{T}$ (or, equivalently, by projection parallel to n) from pre-collision to post-collision moments.

At a moment of collision a tangent vector of a front changes non-continuously (the front is "scattered"): a tangent vector (dr^-, dv^-) of the pre-collision front evolves into the tangent vector (dr^+, dv^+) of the post-collision front given by

$$\begin{aligned} dr &:= dr^+ = dr^-, && (2.6)\\ dv^+ &= dv^- + 2\langle n, v\rangle V^*KVdr && (2.7)\end{aligned}$$

where

- $V : \mathcal{J} \to \mathcal{T}$ is the projection parallel to v: $Vdv = dv - \frac{\langle dv,n\rangle}{\langle v,n\rangle}v \in \mathcal{T}$ for $dv \in \mathcal{J}$,
- $V^* : \mathcal{T} \to \mathcal{J}$ (the adjoint of V) is the projection parallel to n: $V^*dq = dq - \frac{\langle dr,v\rangle}{\langle n,v\rangle}n \in \mathcal{J}$ for $dq \in \mathcal{T}$,

- $K : \mathcal{T} \to \mathcal{T}$ is the s.f.f. of the scatterer at the collision point,
- $\langle n, v \rangle = \cos\varphi$, where $\varphi \in [0, \frac{\pi}{2}]$ is the collision angle.

From these we can get the evolution of the second fundamental form:

$$B^+ = B^- + 2\cos\varphi V^* K V \tag{2.8}$$

(equation (2.3) from [BChSzT1]), where

- $B^- : \mathcal{J} \to \mathcal{J}$ is the s.f.f. just before collision, which also describes tangent vectors (dr^-, dv^-) of the pre-collision front through

$$dv^- = B^- dr^-. \tag{2.9}$$

- $B^+ : \mathcal{J} \to \mathcal{J}$ is the s.f.f. just after collision, which also describes tangent vectors (dr^+, dv^+) of the pre-collision front through

$$dv^+ = B^+ dr^+. \tag{2.10}$$

2.3.2 unstable and stable cone field

In Condition A.2, uniform hyperbolicity is formulated in terms of invariant cone fields. The construction of these cone fields is done in a standard way: cones consist of tangent vectors of appropriate fronts. More precisely: since fronts are subsets of the flow phase space, cones consist not of tangent vectors of fronts, but rather the traces of these vectors on $\mathcal{T}M$.

The fronts defining the unstable cone field satisfy

$$c < B^+ \quad , \tag{2.11}$$
$$c < B^- < C. \tag{2.12}$$

Similarly, the fronts defining the stable cone field satisfy

$$c < -B^- \quad , \tag{2.13}$$
$$c < -B^+ < C. \tag{2.14}$$

A vector (dq, dw) in the unstable cone satisfies

$$\|(dq, dw)\| \sim \|dq\| \sim \|dw\|. \tag{2.15}$$

Stable cones satisfy similar inequalities for the backward dynamics.

2.3.3 transition from Poincaré to orthogonal section

Let us consider a front directly after collision. It leaves a trace of velocities on the scatterer which can be viewed either as a (unit) vector field over $\partial\mathbb{Q}$ or as a $(d-1)$-dimensional submanifold in the Poincaré phase space. Direct calculations show that for a vector (dr, dv) tangent to the post-collisional front, the corresponding vector in the Poincaré phase space is $dx = (dq, dw)$ where:

$$\begin{aligned} dq &= V dr; && (2.16) \\ dw &= dv - \langle v, n\rangle V^* K dq && (2.17) \end{aligned}$$

(equation (4.4) from [BChSzT1]). Notice that this formula depends on the differentiable manifold structure of $\tilde{M}$ based on identification of nearby velocity hemispheres through the parallel transport of the scatterer.

2.3.4 Transversality of fronts and their traces on M

Consider two fronts with s.f.f.-s B_1^- and B_2^- (just before collision) that satisfy

$$B_2^- - B_1^- > c_1, \quad -C_1 < B_1^- < C_1.$$

This is a sufficient condition for the transversality of the (tangent spaces of the) fronts as $d-1$-dimensional subsets of $\mathcal{M}$. Then we know from [BChSzT1], Lemma **??**, that their traces in the Poincaré phase space are also transversal: their angle is at least some $c > 0$ depending only on c_1 and C_1 (and the geometry of the billiard table).

2.3.5 expansion along unstable manifolds

The rather technical formulas of this subsection will only be used in Section 2.5.2 to prove Lemma 2.10 and its corollary. The reader is encoureged to skip these details for the first reading.

Let W be a "u-manifold", that is, a $d-1 = \frac{\dim M}{2}$ dimensional submanifold of M, which has all of its tangent vectors in the unstable cone (at all of its points). Denote by $J^e_W(x)$ the Jacobian of T restricted to W at x. Then we know

$$J^e_W(x) \ \sim \ \det(\, V_1\,) \ \sim \ (\cos(\varphi_1))^{-1}, \tag{2.18}$$

which is equation (4.15) from [BChSzT1].

Let us now consider a further restriction of DT onto a subspace $R \subset \mathcal{T}_x W$ of the tangent plane of this u-manifold. For this we know that

$$\det(\, DT\,|_R) \ \sim \ \det(V_1\,|_{R'}) \tag{2.19}$$

where $R' = (V_1^{-1} \circ \pi_1 \circ DT)(R)$, and π is the natural projection of $\mathcal{T}_x\tilde{M}$ to $\mathcal{T} = \mathcal{T}_x\partial\mathbb{Q}$: $\pi((dq, dw)) = dq$. This is equation (4.16) from [BChSzT1].

2.4 known regularity properties

2.4.1 geometric regularity properties in the Euclidean structure

The first two regularity properties listed here are obvious, the others are proven in [BChSzT1]. The phase space is always equipped with the Euclidean structure.

1. T is piecevise Hölder continuous – i.e. it is Hölder continuous on the finitely many components of $\tilde{M} \setminus \mathcal{S}$, but of course also on the countably many components of $\tilde{M} \setminus (\mathcal{S} \cup \Gamma)$. Actually, the Hölder exponent is $\frac{1}{2}$, but for simplicity we will use that $\rho(Tx, Ty) \leq \sqrt[3]{\rho(x, y)}$ whenever $\rho(x, y)$ is small enough, and x and y are in the same component of continuity.

2. The expansion of T is bounded when not acting near singularities. In particular, there exists a $\delta > 0$ such that $\|DT_x\| \leq \frac{1}{\delta}$ whenever $\rho(x, \mathcal{S}) > \delta$.

3. Transversality: the angle between vectors of the stable and unstable cones C_x^s and C_x^u is uniformly bounded away from zero.

4. Alignment: the angle of any unstable manifold with any (one-step) singularity manifold in $\mathcal{S}$ or Γ is uniformly bounded away from zero.

5. Chernov's regularity conditions:

   (a) Uniform hyperbolicity: We know that Condition A.2 is satisfied by some iterate T^n of the dynamics. We will not make use of this fact, instead, we will prove the stronger statement of *one-step* uniform hyperbolicity for another Riemannian structure.

   (b) Curvature bounds: Condition A.4 is satisfied.

   (c) Distortion bounds: Condition A.5 is satisfied.

   (d) Absolute continuity: Condition A.6 is satisfied.

Remark 2.6. *To avoid confusion we mention that Condition A.6 is not explicitly stated in this form in [BChSzT1]. However, the relevant statement, Theorem 5.9 of [BChSzT1] is known to imply the absolute continuity property of Condition A.6, based on classical results by Anosov and Sinai from [AS]. Furthermore, we may allow for a little more flexibility: the manifolds W_1 and W_2 that appear in Condition A.6 may be arbitrary u-manifolds (manifolds with tangent planes in the unstable cone). Actually, it is this slightly generalized form of the absolute continuity property that we apply in [BBT]. See also [CM] on further details about different formulations of absolute continuity.*

2.4.2 Structure and complexity of the singularity set

In this section we discuss singularities of higher iterates of the dynamics T. We introduce the notation

$$\mathcal{S}^n = T^{-n}\mathcal{S}^0 \quad n = 1, 2, \ldots, -1, -2, \ldots$$

for the "n-step singularity set". So $\mathcal{S}^0 = \partial M$ is the "0-step singularity set", and the singularity set of T is $\mathcal{S} = \mathcal{S}^1$. The set of points where T^n is singular is

$$\mathcal{S}^{(n)} := \bigcup_{i=1}^{n} \mathcal{S}^n \text{ for } n \geq 1$$

$$\mathcal{S}^{(n)} := \bigcup_{i=n}^{-1} \mathcal{S}^n \text{ for } n \leq -1.$$

The n-step secondary singularity set Γ^n and the secondary singularity set of T^n, $\Gamma^{(n)}$ can be defined analogously. However, in this section we discuss the structure of the *primary* singularity set $\mathcal{S}^{(n)}$ of T^n. That is, the secondary singularities in (2.3) are not considered.

An important feature of the singularity set of billiards is the so-called "continuation property". This means that the primary singularities are one-codimensional submanifolds that can only terminate on each other, or on the boundary of M. More precisely, $\mathcal{S}$ is a finite union of one-codimensional compact submanifolds of M with boundary, and every boundary point is either an inner point of some other component, or it is on ∂M. See Figure 1.

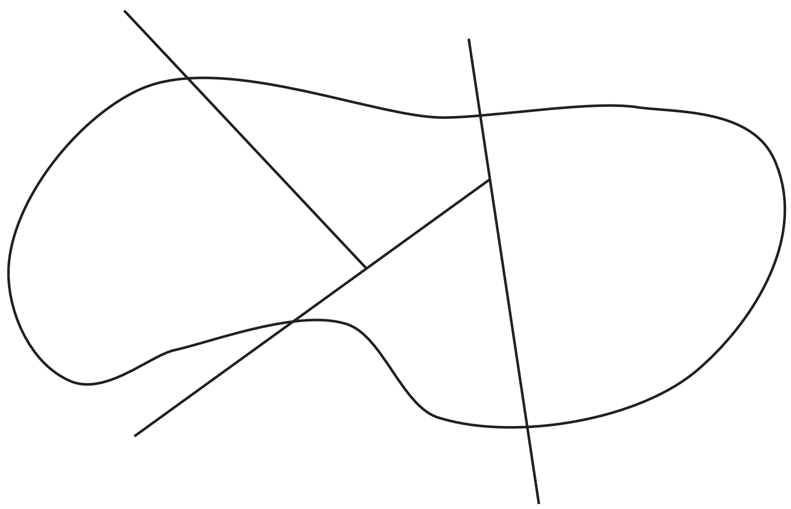

Figure 1: a piece of the phase space cut by singularities

The consequence of this continuation property is that singularities do cut the phase space: if a small open subset U of M is intersected by a singularity manifold, then it is indeed cut into two components. These are not necessarily connected components: U itself may be non-connected already, but even a connected set may well be cut into many pieces by a single 1-codimensional plane. So by "component" we mean those points of U which can be connected by a continuous curve in M which is disjoint from the entire (primary) singularity set $\mathcal{S}$.

Definition 2.7. *For every n, the set $\tilde{M} \setminus \mathcal{S}^{(n)}$ consists of finitely many (open) connected components, let us denote these by $\tilde{M}^{(n),i}$, where i is from some finite index set. Now for any set $U \subset \tilde{M}$, we denote by $K_{n,U}$ the number of $\overline{\tilde{M}^{(n),i}}$-s that are intersected by U. The quantity $K_{n,U}$ will be referred as the complexity of $\mathcal{S}^{(n)}$ on U.*

For $x \in \tilde{M}$ let us denote by $K_{n,x}$ the number of $\overline{\tilde{M}^{(n),i}}$ that contain x, which will be referred as the complexity of $\mathcal{S}^{(n)}$ at x. Finally, we define the complexity of $\mathcal{S}^{(n)}$ as

$$K_n := \sup_{x \in \tilde{M}} K_{n,x}.$$

Remark 2.8. *It follows from compactness of $\tilde{M}$ that for every n there exists an ε such that if the diameter of U is less than ε, then $K_{n,U} \le K_n$.*

A very common assumption in the theory of singular hyperbolic dynamical systems is that the complexity K_n is a subexponential function of n, or at most of $\mathcal{O}(\lambda^n)$ where λ is strictly less than the smallest expansion on the unstable cone. We will also have to assume this property, see Assumption 1.3.

2.5 Further regularity properties

In this section we discuss two further regularity properties, which are not new, and are in a sense known to Billiards experts, however, we could not locate a precise formulation and proof in the literature.

2.5.1 smoothness of one-step singularities

Much of the difficulty in the discussion of multi-dimensional dispersing billiards is related to a phenomenon discovered in [BChSzT1]: if we consider higher iterates of the dynamics, the singularity set will be non-smooth, i.e. its curvature blows up near certain "pathological" points. In the present work, however, it is important for us that such a pathology does *not* appear for the (non-iterated, or 1-step) dynamics itself. With the notations of Section 2.2.3: $T^{-n}\mathcal{S}^0$ may behave irregularly for $n \ge 2$, but $\mathcal{S} = T^{-1}\mathcal{S}^0$ is smooth. This is also true for the secondary singularity set Γ, and even for every submanifold in $\tilde{M}$ which is (similarly to a secondary singularity) the pre-image of an arbitrary $\{\varphi = const\}$ set. Moreover, there is a universal upper bound for the curvature of all of these manifolds. This fact is stated in the following proposition.

Proposition 2.9. *There is a global constant $K_{\mathcal{S}}$ such that for any $\varphi_0 \in [0, \frac{\pi}{2}]$, the curvature of the submanifold $T^{-1}(\{x \in \tilde{M} \mid \varphi(x) = \varphi_0\})$ at any of its points is at most $K_{\mathcal{S}}$.*

The proof of this proposition is postponed until Section 3.3.1, since the precise notion of "curvature" used in the statement is discussed in Section 3.3 only.

2.5.2 anisotropy near tangent collisions

A key feature of multi-dimensional dispersing billiard dynamics is the anisotropy of expansion in unstable directions. This means that near singularities the expansion is not only very strong, but also very direction-dependent. Indeed, certain unstable vectors are expanded enormously (of the order $1/\cos\varphi_1$), while others are only expanded moderately (of the order 1). There is in a sense only one strongly expanding direction, which is approximately the direction of perturbations within the plane of the trajectory (at the next collision). Here we need to make these statements precise, and draw the consequence that the strongly expanding direction is 'just orthogonal' to the secondary singularities – that is, the distance of nearby $\{\varphi_1 = const\}$ manifolds is increased by T by a factor of $1/\cos\varphi_1$.

Lemma 2.10. *For any $x \in M$ near a singularity, there is a tangent vector $dx \in \mathcal{T}_x M$ such that*

$$|d\varphi_1| \sim \|DTdx\| \sim \frac{1}{\cos\varphi(Tx)}\|dx\|.$$

Proof. We work in the tangent space of M at Tx, so for convenience we choose Tx as time zero and denote quantities at Tx without indices. We will say that a vector of $\mathcal{T}$ or $\mathcal{J}$ is in the "strongly scattering direction", if it is in the plane spanned by n and v (orthogonal to $\mathcal{T} \cap \mathcal{J}$). Such vectors have the property that they are greatly expanded by V or V^* (exactly by the factor $1/\cos\varphi$).

We will use that for a tangent vector (dr, dv) of a front just after collision, which has the vector (dq, dw) as its trace on $\mathcal{T}M$, we have $dq = Vdr$ (which is (2.16)) and

$$dw = B^- dr + V^* K \cos\varphi V dr,$$

which comes e.g. from the combination of (2.8) and (2.17).

Choose dx so that $DTdx = (dq, dw)$ has dq pointing in the strongly scattering direction. Then $\|dq\| = \|Vdr\| = \frac{1}{\cos\varphi}\|dr\|$, so the vector $\cos\varphi V dr$ has length $\|dr\|$ and is in the strongly scattering direction. Since K is positive definite and $c < K < C$, the vector $K\cos\varphi V dr$ also has a component in the strongly scattering direction which is at least $c\|dr\|$ long, and the other component is not longer than $C\|dr\|$ either. So $V^* K \cos\varphi V dr$ is at least $\frac{c}{\cos\varphi}\|dr\|$ long and points mainly in the strongly scattering direction. Since $B^- dr$ is of order $\|dr\|$ only (by (2.12)), the whole of dw points mainly in the strongly scattering direction, meaning that $\|dw\| \approx |d\varphi|$. Now (2.15) gives $\|DTdx\| \sim |d\varphi|$.

On the other hand, dx was chosen exactly so that (2.19) gives

$$\frac{\|DTdx\|}{\|dx\|} \sim \frac{1}{\cos\varphi},$$

so dx is really the vector we are looking for. □

Corollary 2.11. *Let γ_1 and γ_2 be two nearby $\{\varphi = const\}$ manifolds: $\gamma_1 = \{x \in M \mid \varphi(x) = \varphi_1\}$, $\gamma_2 = \{x \in M \mid \varphi(x) = \varphi_1 + d\varphi\}$. Then*

$$\mathrm{dist}(\gamma_1, \gamma_2) \sim \frac{1}{\cos\varphi_1}\mathrm{dist}(T^{-1}\gamma_1, T^{-1}\gamma_2).$$

Proof. (2.18) implies the upper bound: no vector can be expanded more than $\sim \frac{1}{\cos\varphi_1}$. The lower bound comes from Lemma 2.10:

$$\mathrm{dist}(\gamma_1, \gamma_2) = |d\varphi| \sim \frac{1}{\cos\varphi_1}\|dx\| \geq \frac{1}{\cos\varphi_1}\mathrm{dist}(T^{-1}\gamma_1, T^{-1}\gamma_2).$$

□

3 Riemannian structure and regularity

Our aim now is to introduce a Riemannian structure on the billiard phase space, which will be different from the natural Riemannian structure of the manifold.

The reason for this is the following. As mentioned in Section 2, the use of the Euclidean structure has the big advantage that the regularity properties A.4, A.5 and A.6 have already been proven in [BChSzT1]. On the other hand, uniform hyperbolicity (A.2) is only true for a higher iterate of the dynamics, e.g. the length of an unstable vector may decrease with one application of the (derivative of the) dynamics, which leads to difficulties when trying to prove the growth properties A.7. The key feature of the Riemannian structure we are about to introduce is that the induced metric exhibits *one step* expansion on unstable vectors.

3.1 Chernov-Dolgopyat metric in two dimensions

The main idea comes from [CD], where the authors use a metric which measures infinitesimal distances on the front, rather than in the Poincaré section. Since expansion of an unstable front is monotonous, this kind of length of unstable vectors clearly grows from collision to collision. For the sake of easier understanding, we first discuss the 2-dimensional construction, and give the multi-dimensional generalization thereafter.

In two dimensions, measuring distances "on the front" simply means using (dr, dv) instead of $(dq, dw) = (dq, d\varphi)$ [2] (with the notations of Section 2). It is important to note

[2]Notice that we are not ignoring the velocity component, so this metric is *not* the so-called *p-metric* of billiards literature, which only measures configurational distances on the front.

however, that (dr, dv) is the tangent vector of the front *after* collision, which results in an asymmetry in the behaviour of stable and unstable vectors.

So the metric, which we will call the Chernov-Dolgopyat metric or C-D metric, is defined as

$$\|(dq, d\varphi)\|_{C-D} := \|(dr, dv)\|_e = \|(dq\cos\varphi, d\varphi + Kdq)\|_e = \sqrt{(dq\cos\varphi)^2 + (d\varphi + Kdq)^2}.$$

Equivalently, the metric tensor has the form

$$g_{C-D}((dq_1, d\varphi_1), (dq_2, d\varphi_2)) = \begin{pmatrix} dq_1 & d\varphi_1 \end{pmatrix} \begin{pmatrix} \cos^2\varphi + K^2 & K \\ K & 1 \end{pmatrix} \begin{pmatrix} dq_2 \\ d\varphi_2 \end{pmatrix}.$$

To be absolutely precise: the matrix $\begin{pmatrix} \cos^2\varphi + K^2 & K \\ K & 1 \end{pmatrix}$ is the matrix of the metric tensor in the basis $\{e, f\}$ where $e \in \mathcal{T}$, $f \in \mathcal{J}$, $\|e\|_e = \|f\|_e = 1$. (This basis is orthonormal in the Euclidean metric.)

We will not rigorously prove hyperbolicity with respect to this metric here, since it will be done in Section 3.2 for the multi-dimensional case, and that is what we need. Instead, we discuss the relation of the C-D metric to the Euclidean.

It is easy to see using (2.15) and (2.11) that for vectors of the unstable cone, the C-D metric is equivalent to the Euclidean: $\|dx\|_{C-D} \sim \|dx\|_e$ for u-vectors. Also, $\|dx\|_{C-D} \leq C\|dx\|_e$ holds for every vector dx, since K is bounded. However, $\|dx\|_{C-D}$ can be much smaller than $\|dx\|_e$ for some vectors (in the stable cone) near the boundary of $\tilde{M}$. Indeed, the determinant of the matrix of the metric tensor is $\cos^2\varphi$, which vanishes on $\partial\tilde{M}$. This non-equivalence has the inconvenient consequence that g_{C-D} is not a Riemannian structure on $\tilde{M}$. More precisely, it is a Riemannian structure only on the inner part, and it cannot be extended to the boundary in a continuous non-vanishing manner. This is inconvenient for several reasons, e.g. no compactness arguments will work.

3.2 generalization to high dimension

The multi-dimensional generalization is quite straightforward. Consider a tangent vector (dq, dw) of the Poincaré section. The configurational part is $dq \in \mathcal{T}$, while the velocity part, $dw \in \mathcal{J}$ is in another space. So we have to be more careful than in 2D where these were numbers that one can add. The counterpart of (dq, dw) on the front is (dr, dv) where $dr \in \mathcal{J}$ and $dv \in \mathcal{J}$. The transition is via the operator $V : \mathcal{J} \to \mathcal{T}$. The formulas of the transition are listed in Section 2.3.3 as (2.16) and (2.17), and they are

$$dr = V^{-1}dq \tag{3.1}$$

$$dv = dw + \langle v, n\rangle V^* K dq.$$

The C-D metric is defined almost like in two dimensions. However, for some reason – to be explained in Remark 3.3 below – we need to insert a small scaling factor $\varepsilon_{C-D} > 0$ which ensures that the velocity component is taken into account with a sufficiently small weight:

Definition 3.1.

$$\|(dq, dw)\|_{C-D} := \|(dr, \varepsilon_{C-D} dv)\|_e = \|(V^{-1}dq, \varepsilon_{C-D}(dw + \langle v, n\rangle V^* K dq))\|_e. \tag{3.2}$$

Or, equivalently, the C-D metric tensor has the matrix (written in block form)

$$g = \begin{pmatrix} V^{-1*}V^{-1} + \varepsilon_{C-D}^2 \langle v, n\rangle^2 KVV^*K & \varepsilon_{C-D}^2 \langle v, n\rangle KV \\ \varepsilon_{C-D}^2 \langle v, n\rangle V^*K & \varepsilon_{C-D}^2 \end{pmatrix}$$

in the basis $\{e_1, \dots, e_{d-1}, f_1, \dots, f_{d-1}\}$ *where* $\{e_1, \dots, e_{d-1}\}$ *is a basis of* $\mathcal{T}$, $\{f_1, \dots, f_{d-1}\}$ *is a basis of* $\mathcal{J}$, *both are orthonormal with respect to the Euclidean metric,* K *is the matrix of the operator* $K : \mathcal{T} \to \mathcal{T}$ *w.r.t the basis* $\{e_1, \dots, e_{d-1}\}$, *and* V *is the matrix of the operator* $V : \mathcal{J} \to \mathcal{T}$ *w.r.t the bases* $\{e_1, \dots, e_{d-1}\}$ *and* $\{f_1, \dots, f_{d-1}\}$.

The factor ε_{C-D} is needed due to another typical multi-dimensional phenomenon.

In the following lemma we show that the billiard dynamics shows one-step uniform hyperbolicity w.r.t. the C-D metric, if ε_{C-D} is small enough. This is a very important advantage in comparison with the Euclidean metric, where we only have uniform hyperbolicity for some higher iterate of the dynamics.

Lemma 3.2. *If norms of tangent vectors are measured using the C-D metric and* ε_{C-D} *is small enough, then* T *exhibits uniform hyperbolicity in the sense of Condition A.2.*

Proof. The existence and properties of the invariant cone fields have already been established in [BChSzT1], and are listed in Section 2.3.2. So we only need to show that there exists some global constant $\Lambda_{C-D} > 1$ such that the $\|.\|_{C-D}$-norm for any tangent vector (dr^+, dv^+) of any post-collision front that corresponds to some u-vector is expanded at least by a factor Λ_{C-D} during a free flight and a collision – that is, until the front becomes a post-collision front again. So let (dr_1^-, dv_1^-) and (dr_1^+, dv_1^+) denote the time-evolved tangent vector just before and after the next collision, respectively.

The formulas of free propagation, (2.4) and (2.5) say that $(dr_1^-, dv_1^-) = (dr^+ + \tau dv^+, dv^+)$, so

$$\|(dr_1^-, \varepsilon_{C-D} dv_1^-)\|_e^2 = |dr^+|^2 + 2\tau\langle dr^+, dv^+\rangle + \tau^2 |dv^+|^2 + \varepsilon_{C-D}^2 |dv^+|^2. \tag{3.3}$$

Since $dv^+ = B^+ dr^+$ and $B^+ > c$ by (2.11), we have $|dv^+| \geq c|dr^+|$ and $\langle dr^+, dv^+\rangle \geq 0$, so (3.3) implies

$$\begin{aligned} \frac{\|(dr_1^-, \varepsilon_{C-D} dv_1^-)\|_e^2}{\|(dr^+, \varepsilon_{C-D} dv^+)\|_e^2} &\geq 1 + \tau^2 \frac{|dv^+|^2}{|dr^+|^2 + \varepsilon_{C-D}^2 |dv^+|^2} = \\ &= 1 + \tau^2 \frac{1}{\frac{|dr^+|^2}{|dv^+|^2} + \varepsilon_{C-D}^2} \geq 1 + \tau_{min}^2 \frac{1}{\frac{1}{c^2} + 1} \end{aligned} \tag{3.4}$$

whenever $\varepsilon_{C-D} \leq 1$. In words: the expansion of the tangent vector during free flight is considerable, since the velocity component is non-negligible, and the flight is not very short.

Similarly (2.6), the first of the collision formulas, together with $B^- < C$ from (2.12) imply that

$$\frac{\|(dr_1^+, \varepsilon_{C-D} dv_1^+)\|_e^2}{\|(dr_1^-, \varepsilon_{C-D} dv_1^-)\|_e^2} \geq \frac{\|dr_1^-\|^2}{\|dr_1^-\|^2 + \varepsilon_{C-D}^2 C^2 \|dr_1^-\|^2} = \frac{1}{1+\varepsilon_{C-D}^2 C^2}. \tag{3.5}$$

This and (3.4) give the statement when ε_{C-D} is small enough.

The argument for the uniform contraction of stable vectors would be completely analogous. Indeed, the minimum expansion along an unstable front from one *pre*-collision moment to the other can be obtained by multiplying the exact same expressions on the right hand sides of (3.4) and (3.5), now in opposite order. But that is exactly the inverse of stable contraction from one *post*-collision moment to the other.

□

Remark 3.3. *We note that in general* $\frac{\|(dr_1^+, dv_1^+)\|_e^2}{\|(dr_1^-, dv_1^-)\|_e^2} \geq 1$ *is* not *true for an unstable front: The tangent vector of the front may be contracted at collision. This does not happen if either the operator* B^- *or* K *is close to a scalar. However, in general it may happen that* $\langle B^- dr, V^* K V dr\rangle < 0$ *despite the fact that both* B^- *and* V^*KV *are positive definite. In such a case, (2.6) and (2.7) may give* $\frac{\|(dr_1^+, dv_1^+)\|_e^2}{\|(dr_1^-, dv_1^-)\|_e^2} < 1$. *This is another typical multi-dimensional phenomenon.*

The following two lemmas are about the relation of the C-D metric to the Euclidean. The statements are greatly different for unstable and stable vectors, which reflects the asymmetry in the definition of the metric.

Lemma 3.4. *The C-D and the Euclidean metric are equivalent for vectors of the unstable cone field. That is, there exists a global constant* $C < \infty$ *such that for any vector* dx *of any unstable cone* C_x^u,

$$\frac{1}{C}\|dx\|_e \leq \|dx\|_{C-D} \leq C\|dx\|_e.$$

Proof. First let us note that, after having chosen ε_{C-D} according to Lemma 3.2, we may keep it fixed so that it becomes a global constant. Let $dx = (dq, dw)$ be a vector of the unstable cone, and let (dr, dv) be its equivalent on the front. The transition is given by (2.16) and (2.17). Vectors of the unstable cone satisfy $B^+ > c$ (from (2.11)), which means that $\|(dq, dw)\|_{C-D} = \|(dr, \varepsilon_{C-D} dv)\|_e \sim \|dv\|$. But (2.15) says $\|(dq, dw)\|_e \sim \|dw\|$. Finally, a combination of (2.7), (2.16) and (2.17) give $\|dv\| \sim \|dw\|$. □

Lemma 3.5. *There exists a global constant* $C < \infty$ *such that for any vector* dx *of any* stable *cone* C_x^s,

$$\|DTdx\|_e \leq C\|dx\|_{C-D}.$$

Proof. Let us use the notation $dx = (dq, dw)$ and $DTdx = (dq_1, dw_1)$.

First, let the tangent vector of the front corresponding to $DTdx$ be (dr_1, dv_1). From (2.17) we have that $|dw_1| \leq |dv_1| + |\cos\varphi_1 V_1^* K_1 dq_1|$. Since $\|V^{-1}\| = 1$, (2.16), (2.10) and

(2.14) give $|dv_1| \le C|dq_1|$. Besides, since $\|\cos\varphi V^*\| = 1$, (1.2) gives $|\cos\varphi_1 V_1^* K_1 dq_1| \le K_{max}|dq_1|$. These together imply that

$$\|DT dx\|_e \le C|dq_1|. \tag{3.6}$$

Second, let the tangent vector of the front corresponding to dx be (dr, dv). The definition of the C-D metric implies that

$$\|dx\|_{C-D} \ge |dr|. \tag{3.7}$$

Due to (3.6) and (3.7), it is enough to show that

$$|dq_1| \le C|dr|, \tag{3.8}$$

and this is what we will do.

In order to prove (3.8) we envoke some notation from [BChSzT1]. Given an invertible linear operator O (that may depend on the phase point we are considering), the relation $c \prec O \prec C$ means that there exist global constants $C_1, C_2 > 0$ such that $\|O\| \le C_1$ and $\|O^{-1}\| \le C_2$, uniformly on the phase space. Furthermore, we envoke the key technical Lemma 4.3 from [BChSzT1]: given two symmetric, positive definite operators $K' : \mathcal{T} \to \mathcal{T}$ and $B' : \mathcal{J} \to \mathcal{J}$ with $c \prec K' \prec C$ and $c \prec B' \prec C$, we also have

$$c \prec B'V^{-1} + \langle n, v\rangle V^* K' \prec C. \tag{3.9}$$

Now let us rewrite (2.4), (2.5), (2.9) and (3.1) as

$$dr = dr_1 - \tau dv_1^- = dr_1 - \tau B_1^- dr_1 = (I - \tau B_1^-)V_1^{-1} dq_1$$

where I is the identity operator. We use (2.8) to express B_1^- in terms of B_1^+ and obtain:

$$dr = \left((I - \tau B_1^+)V_1^{-1} + 2\tau\langle n_1, v_1\rangle V_1^* K_1\right) dq_1.$$

Now we may invoke (3.9) with $B' = I - \tau B_1^+$ and $K' = 2\tau K_1$ to prove (3.8). To see that these operators are indeed bounded from above and below we refer to (2.14) on the one hand, and to our assumptions (1.1), (1.2), (1.3) and (1.4) on the other hand. $\square$

Although the use of the C-D metric ensures uniform hyperbolicity, it has the disadvantage that it is not a true Riemannian structure. This can be seen exactly as in 2 dimensions: the determinant of the metric tensor (w.r.t. the Euclidean) is $\cos^2\varphi$, which vanishes on the boundary of $\tilde{M}$, where the metric is degenerate. This has many unpleasant consequences. First of all, Chernov's Condition A.1 about the dynamical system formally demands a true Riemannian manifold, with $\bar{M}$ compact. However, the problems with the non-Riemannian nature of the $C-D$ metric are deeper, and not just formal. Some details will be explained in Section 3.4, Remark 3.19. For these reasons, we will use a regularized version of the C-D metric structure, which will be truely Riemannian, and will not exhibit the unpleasant features of the original C-D metric.

Definition 3.6. *The Riemannian structure we use on the Poincaré phase space $\tilde{M}$ is the "Regularized Chernov-Dolgopyat" metric tensor field defined by*

$$g := g_{C-D} + \varepsilon_g g_e.$$

Here g_e is the Euclidean metric tensor field (the natural Riemannian structure on $\tilde{M}$), and $\varepsilon_g > 0$ is an arbitrary constant.

The choice of ε_g will be based on the following Proposition:

Proposition 3.7. *The "Regularized Chernov-Dolgopyat" metric tensor field g is a C^1 Riemannian structure on $\tilde{M}$. If ε_g is small enough, then T is uniformly hyperbolic w.r.t. g. That is, Condition A.2 holds.*

Proof. The fact that g_{C-D} is a C^1 field of symmetric tensors of type $(0, 2)$ is clear from the definition: it is built up of K, $\langle n, v\rangle V$, $\langle n, v\rangle V^*$, V^{-1} and V^{-1*}, which are all bounded and continuously differentiable up to the boundary of $\tilde{M}$. g_{C-D} is also positive semi-definite, which is clear from (3.2). Since g_e is a C^1 Riemannian structure, $g = g_{C-D} + \varepsilon_g g_e$ is also a C^1 field of symmetric tensors of type $(0, 2)$, which is positive definite (everywhere) if $\varepsilon_g > 0$. One can check by direct calculation that the determinant of g (w.r. to the Euclidean) is uniformly bounded away from 0. This altogether means that g is truely a Riemannian structure.

To prove uniform hyperbolicity (i.e. that Condition A.2 holds), we still use the invariant cone field already introduced in [BChSzT1] and described in Section 2.3.2, used in Lemma 3.2 as well. So we only need to see that vectors of the unstable cone are expanded, and vectors of the stable cone are contracted at least by a factor $\Lambda > 1$.

For unstable vectors this is easy, since the C-D and the Euclidean metric are equivalent on unstable vectors (by Lemma 3.4), so the term $\varepsilon_g g_e$ in g is negligable if ε_g is small enough. So expansion is inherited from the C-D metric (Lemma 3.2). Indeed,

$$\|dx\|_e \leq C\|dx\|_{C-D}$$

for every vector dx of the unstable cone, thus

$$\|dx\| \leq \sqrt{1 + \varepsilon_g^2 C^2}\|dx\|_{C-D},$$

so

$$\|DTdx\| \geq \|DTdx\|_{C-D} \geq \Lambda_{C-D}\|dx\|_{C-D} \geq \Lambda_{C-D}\frac{1}{\sqrt{1 + \varepsilon_g C}}\|dx\|,$$

which proves the statement for unstable vectors if ε_g is so small that

$$\frac{\Lambda_{C-D}}{\sqrt{1 + \varepsilon_g^2 C^2}} > 1.$$

The case of stable vectors dx is also easy, once we have the difficult Lemma 3.5 about the dynamical comparison of the metrics on stable vectors. Using that lemma (and Lemma 3.2 about the hyperbolicity of the C-D metric), we can simply write

$$\|DTdx\|^2 = \|DTdx\|^2_{C-D} + \varepsilon_g^2 \|DTdx\|^2_e \leq \frac{\|dx\|^2_{C-D}}{\Lambda^2_{C-D}} + \varepsilon_g^2 C^2 \|dx\|^2_{C-D} = \tag{3.10}$$

$$= \left(\frac{1}{\Lambda^2_{C-D}} + \varepsilon_g^2 C^2 \right) \|dx\|^2_{C-D} \tag{3.11}$$

which proves the statement for stable vectors if ε_g is so small that

$$\frac{1}{\Lambda^2_{C-D}} + \varepsilon_g^2 C^2 < 1.$$

□

3.3 Curvature bounds and Riemannian structure

"Bounded curvature" is a commonly used regularity property in Dynamical Systems theory. In the literature there are many statements which claim that the curvature of certain submanifolds of the phase space or the configuration space is bounded. There is a variety of notions of curvature used in these statements. The essence of curvature bounds is always the fact that "if two points are near, then their tangent spaces are also near", so one needs to compare vectors of different tangent spaces. This can be done without any special care if the containing manifold is Euclidean, but in general one would need to identify nearby tangent spaces through the parallel transport of the manifold.

Since we are going to use a Riemannian structure on $\tilde{M}$ which is not Euclidean, and even different from the natural Riemannian structure, we will now formulate precisely what we mean by curvature bounds. We use notation which is standard in differential geometry, see e.g. [KN].

The proper notion for the curvature of an unstable manifold is the curvature as of a *submanifold*, so it's not a quantity of inner geometry. It should describe how fast the submanifold "bends away" from the geodesics tangent to it, e.g. a cylinder (surface) as a subset of $\mathbb{R}^3$ with a small radius should be considered heavily curved, although its inner geometry is Euclidean. Only this way can bounded curvature mean that the submanifold can be viewed (at the cost of an arbitrarily small error) as a plane, if the scale is small enough. The quantity which measures curvature in this sense is the second fundamental form. We have already used this quantity in describing fronts as subsets of $\mathbb{Q}$. Since in $\mathbb{Q}$ fronts are one-codimensional, the notion of second fundamental form was easier there, but the generalization to higher codimensions is also known in differential geometry.

Let M be a Riemannian manifold with Riemannian metric tensor field g, and let ∇ be the connection defined by g.

Definition 3.8. *The second fundamental form of a C^2 submanifold W at the point x is $II : T_x \times T_x \to N_x$, where T_x is the tangent space and N_x is the normal space of W at x, defined by*

$$II(v, w) = \nabla_v^{\perp} w.$$

Here $\perp$ means "component orthogonal to W". For this definition to make sense, at least w has to be a vector field, but the value of II(v,w) will only depend on the value of w at x, as long as w is a tangent vector field of W (at least in every point of W near x). II is bilinear.

Remark 3.9. *To clarify why this quantity is indeed the proper notion of curvature of W as a submanifold, i.e. the amount of non-flatness of W in M, here is a small picture about its meaning with coordinates. We will not use this picture later, all our proofs will be based on the definition.*

Let us choose $\{e_1, \ldots, e_k\}$ to be an orthonormal basis of T_x and $\{n_1, \ldots, n_l\}$ to be an orthonormal basis of N_x, and choose, as a coordinate chart, normal coordinates built from the basis $\{e_1, \ldots, e_k, n_1, \ldots, n_l\}$. Denote the coordinates as $(x^1, \ldots, x^k, y^1, \ldots, y^l)$. Then in this coordinate chart the submanifold W (near the origin) will be the graph of a function $f : \mathbb{R}^k \to \mathbb{R}^l$. In the Taylor polinomials of f the constant and the linear term are zero by the choice of the coordinate system, and the quadratic term is exactly the quadratic transformation from $\mathbb{R}^k$ to $\mathbb{R}^l$ defined by the components of II in the bases $\{e_1, \ldots, e_k\}$, $\{n_1, \ldots, n_l\}$.

That is, the second degree Taylor polinomial of f at 0 is

$$y^a = T_2^a(x^1, \ldots, x^k) = II^a_{bc} x^b x^c$$

where

$$II(e_b, e_c) = II(\frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c}) = II^a_{bc} \frac{\partial}{\partial y^a} = II^a_{bc} n_a.$$

Now we can make Condition A.4 about curvature bounds precise:

Condition 3.10. *There should exist a constant $K_W < \infty$ such that at any point x of any unstable manifold W, the second fundamental form of W at x (as a bilinear operator) is bounded by K_W. That is, for any $v, w \in \mathcal{T}_x W$, $\|II(v, w)\| \leq K_W \|v\| \|w\|$.*

The reader may check that – although differenty formulated – this is exactly what was proven for the Euclidean metric of $\tilde{M}$ in Theorem 5.5 of [BChSzT1], and – although differently said – this is exactly what is used in [Ch2] as the "bounded curvature" assumption.

3.3.1 Proof of Proposition 2.9 about bounded curvature of one-step singularities

The proof of Proposition 2.9 will be based on the following lemma. Let V and W be two smooth 1-codimensional submanifolds of a Riemannian manifold $\mathcal{M}$, and let them be transversal. Then $V \cap W$ is also a 1-codimensional submanifold of V.

Lemma 3.11. *For every $C < \infty$ and $\alpha > 0$ there exists a $C' < \infty$ such that if at a point $x \in V \cap W$ the s.f.f. of V and the s.f.f. of W are both bounded by C and the angle of V and W is at least α, then the s.f.f. of $V \cap W$ as a submanifold of V is bounded by C' (at x).*

Proof. Let us denote the covariant derivation within V by ∇^V. Let v and w both be tangent vectors (vector fields) of $V \cap W$. The quantity we wish to estimate is the component of $\nabla^V_v w$ orthogonal to $V \cap W$. But $\nabla^V_v w$ is just the component of $\nabla_v w$ parallel to V, that is,

$$\nabla^V v_w = \nabla_v w - II_V(v, w),$$

where II_V is the s.f.f. of V. Due to our assumption, the length of $II_V(v, w)$ is at most $C|v||w|$, and of course its component orthogonal to $V \cap W$ cannot be longer either. So it is enough to find an estimate for the other term, i.e. the component of $\nabla_v w$ orthogonal to $V \cap W$. Now denote the (unit) normal vectors of V and W by e and f. Our assumptions imply that the components of $\nabla_v w$ in the direction of e and f are both bounded by $C|v||w|$. The statement is that the component of $\nabla_v w$ within the plane of e and f is also bounded by some $C'|v||w|$. But this is clear since the angle of e and f is at least α, so any vector within their plane which is long, must have a long component is at least one of their directions. ☐

To prove Proposition 2.9 we apply the lemma with $\mathcal{M}$ the flow phase space of the billiard, and $V = M$, the Poincaré section phase space. W is chosen to be the 1-step singularity manifold of the flow dynamics. It is easy to see that the s.f.f. of $V = M$ within $\mathcal{M}$ is bounded, since M is a compact smooth submanifold of $\mathcal{M}$. We now only need to see that the s.f.f. of W is bounded in the points of M, and that M and W are uniformly transversal. Both of these can be seen easily, since the minimum free flight τ_{min} was supposed to be nonzero - i.e. there are no corner points.

3.4 Equivalence of Riemannian structures and inherited regularity properties

In this section we consider the problem of having two different Riemannian structures on the same differentiable manifold. The essence of the statements is that under the mildest possible regularity conditions (adequate differentiability of the metric tensor fields) the regularity properties of submanifolds and maps (curvature bounds and distortion bounds) are inherited from one Riemannian manifold (or rather: one Riemannian structure) to the other. That is, these notions are independent of the choice of the Riemannian structure.

In the statements to come, M will always denote a C^2 differentiable manifold. g, $\tilde{g}$ will be C^1 Riemannian metric tensor fields on M, and W a C^2 smooth submanifold of M. Since the notions of 'covariant differentiation', 'second fundamental form', 'orthogonality' and 'norm' depend on the Riemannian structure, we will use ∇, II, $\perp$, $\|.\|$ to denote them when g is used, and $\tilde{\nabla}$, $\tilde{II}$, $\tilde{\perp}$, $\|.\tilde{\|}$ to denote them w.r.t. $\tilde{g}$. The phrase 'second fundamental form' will be abbreviated as s.f.f.

Definition 3.12. *Let M be a C^2 differentiable manifold, possibly with boundary. Two C^1 Riemannian metric tensor fields g and $\tilde{g}$ on M are said to be C^1 equivalent with constant K if both $\tilde{g}$ and $\nabla\tilde{g}$ (as tensors) are bounded by K when g is used for the definition of norm and covariant derivation, and vice versa. In detail: for any $x \in M$, $u, v, w \in \mathcal{T}_xM$*

$$\begin{aligned} |\tilde{g}(v,w)| &\leq K\|v\|\|w\|, \\ |(\nabla_u\tilde{g})(v,w)| &\leq K\|u\|\|v\|\|w\|, \\ |g(v,w)| &\leq K\|v\|^{\sim}\|w\|^{\sim}, \\ |(\tilde{\nabla}_u g)(v,w)| &\leq K\|u\|^{\sim}\|v\|^{\sim}\|w\|^{\sim}. \end{aligned}$$

If translated to norms of vectors, the first and third inequality of the definition say that $\|v\|^{\sim} \leq \sqrt{K}\|v\|$ and $\|v\| \leq \sqrt{K}\|v\|^{\sim}$, but for conveniance we will omit the square root and use $\|v\|^{\sim} \leq K\|v\|$, $\|v\| \leq K\|v\|^{\sim}$. This is fine since $K \geq 1$.

Lemma 3.13. *On a compact C^2 differentiable manifold M (possibly with boundary) any two C^1 Riemannian metric tensor fields g and $\tilde{g}$ are C^1 equivalent with some constant K (depending of course on the tensor fields).*

Proof. Since $\tilde{g}$ and g are continuous, the norm of $\tilde{g}$ w.r.t. g, $\|\tilde{g}\| = \sup\{|\tilde{g}(v,w)| \mid \|v\| = \|w\| = 1\}$ is a continuous function on M, so it has a finite maximum, since M is compact. Similarly, $\nabla\tilde{g}$ is continuous, so its norm w.r.t. g, $\|\nabla\tilde{g}\| = \sup\{|(\nabla_u\tilde{g})(v,w)| \mid \|u\| = \|v\| = \|w\| = 1\}$ is a continuous function on M, so it has a finite maximum. The same is true with g and $\tilde{g}$ interchanged. The greatest of these four maxima can be chosen as K. □

Lemma 3.14. *For any $K_1, K_2 < \infty$ there is a $\tilde{K} < \infty$ (depending only on K_1, K_2 and $\dim(M)$), such that for any g, $\tilde{g}$ and W, if g and $\tilde{g}$ are C^1 equivalent with constant K_1 and the s.f.f. of W w.r.t g is bounded by K_2, then the s.f.f. of W w.r.t $\tilde{g}$ is bounded by $\tilde{K}$.*

Before we can start the proof, we state and prove a sublemma:

Sublemma 3.15. *The norm of the vector-valued tensor S defined by $S(v,w) = \tilde{\nabla}_v w - \nabla_v w$ is bounded by a constant $\hat{K}$ depending only on K_1 and $\dim(M)$. That is, $\|S(v,w)\| \leq \hat{K}\|v\|\|w\|$.*

Proof. The fact that S is indeed a vector-valued tensor is known, see e.g. [KN].

At any point $x \in M$ we can take normal coordinates w.r.t. g. In this coordinate chart, the Christoffel-symbols of ∇ are zero (*at the single point* x). This has two consequences. First, the components of S are exactly the Christoffel-symbols of $\tilde{\nabla}$. Secondly, in this coordinate chart, at x, the partial derivatives of the components of $\tilde{g}$ are exactly the components of $\nabla\tilde{g}$, and are thus bounded by K_1 because of the C^1 equivalence we assumed. So are the components of g^{-1}, again by the equivalence of g and $\tilde{g}$. Since the Christoffel-symbols of $\tilde{\nabla}$ can be expressed in terms of the above two, they can clearly be estimated using K_1 and $\dim(M)$. This implies a similar estimate for the norm of S. The estimate is clearly independent of the choice of the point x. □

Now we can turn to the proof of Lemma 3.14.

Proof. of Lemma 3.14. We want to estimate $\tilde{II}$, the s.f.f. of W w.r.t. $\tilde{g}$. We will use the definition of II, the definition of S and the fact that

$$\nabla_u v = \nabla_u^W v + \nabla_u^{\perp} v = \nabla_u^W v + II(u,v),$$

where $\nabla_u^W v$ is parallel to W. We write

$$\begin{aligned}
\tilde{II}(u,v) &= (\tilde{\nabla}_u v)^{\tilde{\perp}} = (\nabla_u v + S(u,v))^{\tilde{\perp}} \\
&= (\nabla_u^W v + II(u,v) + S(u,v))^{\tilde{\perp}} \\
&= (\nabla_u^W v)^{\tilde{\perp}} + (II(u,v))^{\tilde{\perp}} + (S(u,v))^{\tilde{\perp}}.
\end{aligned}$$

The first term is zero, and the other two can be overestimated if we omit the $\tilde{\perp}$. So we get – using all the assumptions of the lemma and the statement of the sublemma

$$\begin{aligned}
\|\tilde{II}(u,v)\|\tilde{} \leq \|II(u,v)\|\tilde{} + \|S(u,v)\|\tilde{} \leq \\
\leq K_1\|II(u,v)\| + K_1\|S(u,v)\| \leq K_1K_2\|u\|\|v\| + K_1\hat{K}\|u\|\|v\| \leq \\
\leq K_1K_2K_1\|u\|\tilde{}K_1\|v\|\tilde{} + K_1\hat{K}K_1\|u\|\tilde{}K_1\|v\|\tilde{} = K_1^3(K_2+\hat{K})\|u\|\tilde{}\|v\|\tilde{}.
\end{aligned}$$

$\square$

Lemma 3.16. *For any $K < \infty$ there is a $K' < \infty$ (depending only on K), such that for any g, $\tilde{g}$ and W if g and $\tilde{g}$ are C^1 equivalent with constant K and the s.f.f. of W w.r.t. g is bounded by K, then the restrictions of g and $\tilde{g}$ to W, $g|_W$ and $\tilde{g}|_W$, are C^1 equivalent with constant K'.*

Proof. To have the equivalence, we first need to see that the norm w.r.t. $g|_W$ of $\tilde{g}|_W$ is bounded, and vice versa. This is obviously inherited from g and $\tilde{g}$, so $K' = K$ would do.

Second, we need that the derivatives are bounded. Let us denote, for a moment, the covariant derivative w.r.t. $g|_W$ by $\nabla^{g|_W}$, and similarly for $\tilde{g}$. We need to see that the norm w.r.t. $g|_W$ of $\nabla^{g|_W}\tilde{g}|_W$ is bounded (and vice versa), but this is of course the same as the norm w.r.t. g, so we introduce no new notation.

To understand $\nabla^{g|_W}\tilde{g}|_W$, we first describe how $\nabla^{g|_W}$ acts on vectors. It is known (see e.g. [KN] again) that when we split the covariant derivative $\nabla_u v$ into tangential and orthogonal components using $\nabla_u v = \nabla_u^W v + II(u,v)$, the tangential component is nothing else than the covariant derivative w.r.t. the metric tensor restricted to W, so $\nabla_u^W v = \nabla_u^{g|_W} v$.

We will express $\nabla^{g|_W}\tilde{g}|_W$ by using the definition of covariant differentiation for a tensor, and the above fact. Let u, v, w be tangent vectors of W (at the same point).

$$\begin{aligned}
(\nabla_u^{g|_W}\tilde{g}|_W)(v,w) = u(\tilde{g}(v,w)) - \tilde{g}(\nabla_u^{g|_W}v, w) - \tilde{g}(v, \nabla_u^{g|_W}w) = \\
u(\tilde{g}(v,w)) - \tilde{g}(\nabla_u v, w) + \tilde{g}(II(u,v), w) - \tilde{g}(v, \nabla_u w) + \tilde{g}(v, II(u,w)) = \\
(\nabla_u\tilde{g})(v,w) + \tilde{g}(II(u,v), w) + \tilde{g}(v, II(u,w)).
\end{aligned}$$

These three terms can be readily estimated using the bounds on $\|\nabla\tilde{g}\|$, $\|\tilde{g}\|$ and $\|II\|$ that we have assumed. We get

$$|(\nabla_u^{g|_W}\tilde{g}|_W)(v,w)| \le K\|u\|\|v\|\|w\| + K\|II(u,v)\|\|w\| + K\|v\|\|II(u,w)\| \le$$
$$K\|u\|\|v\|\|w\| + KK\|u\|\|v\|\|w\| + K\|v\|K\|u\|\|w\| = (K+2K^2)\|u\|\|v\|\|w\|.$$

We got that $\|\nabla^{g|_W}\tilde{g}|_W\| \le K' = K + 2K^2$. The bound for $\|\nabla^{\tilde{g}|_W} g|_W\|$ is exactly the same. ∎

Now we investigate how sensitive distortions are to the choice of the Riemannian structure. The following lemma states that if a map of one manifold to the other satisfies certain distortion bounds, then modifying the Riemannian structures on both manifolds up to C^1 equivalence, distortion bounds remain valid. Later we will apply the result to the restriction of the dynamics to an unstable manifold, which maps to another unstable manifold.

In the lemma let M and M' be two C^2 differentiable manifolds. Let M carry two C^1 Riemannian metric tensor fields, g and $\tilde{g}$. Similarly, let M' carry two C^1 Riemannian metric tensor fields, g' and $\tilde{g}'$. Let $T : M \to M'$ be a C^1 map and let $x, y \in M$. The two metrics on M defined by g and $\tilde{g}$ are denoted by d and $\tilde{d}$, the two metrics on M' defined by g' and $\tilde{g}'$ are denoted by d' and $\tilde{d}'$. The Jacobian of T w.r.t. (g, g') is denoted by J, and the Jacobian of T w.r.t. $(\tilde{g}, \tilde{g}')$ is denoted by $\tilde{J}$.

Lemma 3.17. *For any $k > 0$, $K < \infty$ and any $h \in o(1)$ there exist $\tilde{k} > 0$ and $\tilde{h} \in o(1)$ (depending only on h, K, k and $\dim(M)$), such that*
if

- *g and $\tilde{g}$ are C^1 equivalent (on M) with constant K and*
- *g' and $\tilde{g}'$ are C^1 equivalent (on M') with constant K and*
- *$d'(Tx, Ty) \ge k\ d(x, y)$ and*
- *$|\log J_y - \log J_x| \le h(d'(Tx, Ty))$*

then

- *$\tilde{d}'(Tx, Ty) \ge \tilde{k}\ \tilde{d}(x, y)$ and*
- *$|\log \tilde{J}_y - \log \tilde{J}_x| \le \tilde{h}(\tilde{d}'(Tx, Ty))$.*

Proof. The first statement follows from the third assumption and the equivalence of the metrics:

$$\tilde{d}'(Tx, Ty) \ge \frac{1}{K} d'(Tx, Ty) \ge \frac{1}{K} k d(x, y) \ge \frac{1}{K} k \frac{1}{K} \tilde{d}(x, y),$$

so $\tilde{k} = \frac{k}{K^2}$ will do.

For the second statement, let A be a parallelepiped (ordered $\dim(M)$-tuple of tangent vectors) at x, and B a parallelepiped at y (let them be nondegenerate). With slight abuse of notation, we denote their images under the derivative of T by TA and TB. Denote the volume element (canonical $\dim(M)$-form) associated to g, $\tilde{g}$ g' and $\tilde{g}'$ by V, $\tilde{V}$, V' and $\tilde{V}'$, respectively. Then we have

$$J_x = \frac{V'(TA)}{V(A)}; \quad J_y = \frac{V'(TB)}{V(B)}; \quad \tilde{J}_x = \frac{\tilde{V}'(TA)}{\tilde{V}(A)}; \quad \tilde{J}_y = \frac{\tilde{V}'(TB)}{\tilde{V}(B)}, \tag{3.12}$$

which are of course independent of the choice of A and B.

Also, the ratios $\frac{\tilde{V}(.)}{V(.)}$ and $\frac{\tilde{V}'(.)}{V'(.)}$ are independent of the argument – they are actually the square root of the appropriate determinant:

$$\frac{\tilde{V}(.)}{V(.)} = \sqrt{\det_g \tilde{g}} \quad ; \quad \frac{\tilde{V}'(.)}{V'(.)} = \sqrt{\det_{g'} \tilde{g}'}$$

which of course, depend on the base point. The (covariant) derivative of $\det_g \tilde{g}$ can be expressed (say, coordinate-vise) in terms of $\tilde{g}$ and $\nabla\tilde{g}$, so it is bounded by some constant $K' < \infty$ depending only on K and $\dim(M)$. $\det_g \tilde{g}$ is also separated from zero, so the same is true for its logarithm with some $K'' < \infty$. This implies that

$$\left|\log\frac{\tilde{V}(B)}{V(B)} - \log\frac{\tilde{V}(A)}{V(A)}\right| = \left|\frac{1}{2}\log\det_g \tilde{g}(y) - \tfrac{1}{2}\log\det_g \tilde{g}(x)\right| \leq K''d(x,y). \tag{3.13}$$

Similarly,

$$\left|\log\frac{\tilde{V}'(TB)}{V'(TB)} - \log\frac{\tilde{V}'(TA)}{V'(TA)}\right| \leq K''d'(Tx,Ty). \tag{3.14}$$

Knowing this, we can force these quantities to show up in the expression for $\log\tilde{J}_y - \log\tilde{J}_x$. First write

$$\frac{\tilde{J}_y}{\tilde{J}_x} = \frac{\tilde{V}'(TB)}{\tilde{V}(B)}\frac{\tilde{V}(A)}{\tilde{V}'(TA)} = \frac{\tilde{V}'(TB)}{V'(TB)}\frac{V'(TB)}{V(B)}\frac{V(B)}{\tilde{V}(B)}\frac{\tilde{V}(A)}{V(A)}\frac{V(A)}{V'(TA)}\frac{V'(TA)}{\tilde{V}'(TA)},$$

than take the logarithm to get

$$\log\tilde{J}_x - \log\tilde{J}_y =$$
$$\left(\log\tfrac{\tilde{V}'(TB)}{V'(TB)} - \log\tfrac{\tilde{V}'(TA)}{V'(TA)}\right) + \left(\log\tfrac{\tilde{V}(A)}{V(A)} - \log\tfrac{\tilde{V}(B)}{V(B)}\right) + \left(\log\tfrac{V'(TB)}{V(B)} - \log\tfrac{V'(TA)}{V(A)}\right).$$

The first and second term can be estimated using (3.14) and (3.13), while the third term is $\log J_y - \log J_x$ by (3.12) and can be estimated using the last assumption of the lemma. We get

$$|\log\tilde{J}_y - \log\tilde{J}_x| \leq K''d'(Tx,Ty) + K''d(x,y) + h(d'(Tx,Ty)).$$

Now we use the equivalence of metrics and the third assumption to replace all distances by $\tilde{d}'(Tx, Ty)$, and get

$$|\log \tilde{J}_y - \log \tilde{J}_x| \leq K''K\tilde{d}'(Tx, Ty) + K''\frac{1}{k}K\tilde{d}'(Tx, Ty) + \sup\{h(s) \mid s \leq K\tilde{d}'(Tx, Ty)\}.$$

So $\tilde{h}(t) = K''K(1 + \frac{1}{k})t + \sup\{h(s) \mid s \leq Kt\} \in o(1)$ will do. $\square$

Corollary 3.18. *All the regularity properties listed in sections 2.4 and 2.5 are inherited from g_e to g. Specifically:*

- *piecewise Hölder continuity of the dynamics*
- *bounded expansion away from the singularities*
- *transversality of stable and unstable cone fields*
- *alignment*
- *curvature bounds for stable and unstable manifolds*
- *distortion bounds*
- *absolute continuity*
- *Proposition 2.9 about the smoothness of one-step singularities, and*
- *Corollary 2.11 about the anisotropy near tangent collisions*

remain valid when the phase space is equipped with the regularized Chernov-Dolgopyat Riemannian structure instead of the Euclidean.

Proof. In the case of

- piecewise Hölder continuity of the dynamics
- bounded expansion away from the singularities
- transversality of stable and unstable cone fields
- alignment
- absolute continuity, and
- Corollary 2.11.

the statement is easy to see without detailed analysis: these depend only on the C^0 equivalence of metrics, which is the comparability of length.

The statement about curvature bounds and smoothness of the one-step singularities follows from Proposition 3.7 (which claims that the C-D structure is indeed C^1), and lemmas 3.13 and 3.14.

The fact that distortion bounds are inherited follows from the curvature bounds, plus lemmas 3.16 and 3.17. $\square$

Remark 3.19. *We note that the reason for using the regularized structure instead of the C-D metric tensor is deeper than the easyness and robustness of the proof we gave for the regularity properties. Actually, we have checked by explicit calculation, that the unstable manifolds* do not *satisfy the bounded curvature assumption, if the unregularized C-D metric is used: the curvature blows up near the boundary of $\tilde{M}$, even in the simplest 3-dimensional configurations. Surprisingly, this does* not *happen in 2D (despite the degeneracy of the metric), which is why Chernov and Dolgopyat could use this tool with such success in [CD]. Blow-up of curvatures w.r.t. the C-D metric is – beside the anisotropy of the unstable expansion and the pathological behaviour of higher-order singularities – another typical multi-dimensional phenomenon.*

4 Growth properties

In this section we prove that the studied multidimensional dispersing billiard systems satisfy Chernov's growth properties. More precisely, we show that there is some fixed integer n_0 such that the n_0th iterate of the billiard map T satisfies Condition A.7.

Throughout the section we will use the notation $A^{[\delta]}$ to denote the (closed) δ-neighbourhood of a subset A of the phase space, or of an unstable manifold.

In accordance with the exposition of Appendix A, it is worth introducing the following notations.

Let $\hat{T}$ be a n_0th iterate of the original billiard map, i.e. $\hat{T} = T^{n_0}$, where $n_0 \in \mathbb{N}$ is to be specified later. Thus the singularity set for $\hat{T}$ is $\hat{\Gamma} = \Gamma^{(n_0)}$, and $\hat{T}$ expands unstable vectors (and contracts stable vectors) at least by a factor $\hat{\Lambda} = \Lambda^{n_0}$.

For $\delta_0 > 0$, we call W a **δ_0-LUM** if it is a LUM and $\operatorname{diam} W \le \delta_0$. For an open subset $V \subset W$ and $x \in V$ denote by $V(x)$ the connected component of V containing the point x.

Let $n \ge 0$. We call an open subset $V \subset W$ a (δ_0, n)-subset if $V \cap \hat{\Gamma}^{(n)} = \emptyset$ (i.e., the map $\hat{T}^n$ is smoothly defined on V) and $\operatorname{diam} \hat{T}^n V(x) \le \delta_0$ for every $x \in V$. Note that $\hat{T}^n V$ is then a union of δ_0-LUM-s.

Proposition 4.1. *There is a fixed $\delta_0 > 0$, furthermore, there exist constants $\alpha \in (0,1)$ and $\beta, D, \kappa, \sigma > 0$ with the following property. For any $0 \le \delta < 1$ and any δ_0-LUM W there is an open $(\delta_0, 0)$-subset V_δ^0 and an open $(\delta_0, 1)$-subset $V_\delta^1 \subset W \setminus \hat{\Gamma}^{[\delta]}$ (one of these may be empty) such that the two sets are disjoint, $m_W(W \setminus (V_\delta^0 \cup V_\delta^1)) = 0$ and $\forall \varepsilon > 0$*

First Growth Property:

$$m_W(\{ x \in V_\delta^1 \mid \rho(\hat{T}x, \partial\hat{T}V_\delta^1(x)) < \varepsilon \}) \leq$$
$$\leq \alpha\hat{\Lambda} \cdot m_W(\{ x \in W \mid \rho(x, \partial W) < \varepsilon/\hat{\Lambda} \}) + \varepsilon\beta\delta_0^{-1} m_W(W); \tag{4.1}$$

Second Growth Property:

$$m_W(\{ x \in V_\delta^0 \mid \rho(x, \partial V_\delta^0(x)) < \varepsilon \}) \leq D\delta^{-\kappa} m_W(\{ x \in W \mid \rho(x, \partial W) < \varepsilon \}); \tag{4.2}$$

and **Third Growth Property***:*

$$m_W(V_\delta^0) \leq D\, m_W(\{ x \in W \mid \rho(x, \partial W) < \delta^\sigma \}). \tag{4.3}$$

Remark 4.2. *Note that Propostion 4.1 is slightly stronger than Condition A.7. Most importantly, here we allow for arbitrary* $0 \leq \delta < 1$*, while the Condition requires only* δ *sufficiently small. Allowing for* $\delta = 0$ *in the first growth property (note in such a case the second and the third growth properties are trivial) provides useful estimates, see also Remarks 4.4, 4.7 and Corollary 4.13.*

4.1 Outline

The first growth property is much more difficult than the other two. Reason for this is that in the second and the third growth properties we have a large amount of freedom, due to the fact that an arbitrary power of δ may appear (δ^κ and δ^σ, respectively). This allows for the use of quite crude measure estimates in their proof, see the exposition in sections 4.2.2 and 4.3.2.

The case of the first growth property is completely different. Here there is no δ, the inequality is sharp and thus there is very limited freedom in the measure estimates.

The two terms appearing on the RHS of (4.1) are responsible for two different effects. The first term estimates the measure of points that get close to the singularities. The second term corresponds to the fact that some components may grow large when applying $\hat{T}$ and may fail to have diameter less than δ_0. Thus one needs to partition these components. The second term estimates the measure of points that get close to these artificial boundaries. Handling the effect of this further chopping is rather standard, (see [Ch1] and Section 4.4 below). Thus what is to be understood is how LUMs are expanded and, simultaneously, partitioned by singularities when iterates of T are applied. This will be the content of our Lemmas 4.3 and 4.5, which will be referred to as 1-step and n-step Lemmas, respectively.

Throughout the rest of the section we will consider the original billiard map T. Recall the concept of δ_0-LUM and (δ_0, n)-subset from above. We also introduce another notation:

Given a (δ_0, n)-subset V, define a function $\mathbf{r_{V,n}}$ on V by

$$r_{V,n}(x) = \rho_{T^nV(x)}(T^n x, \partial T^n V(x)) \tag{4.4}$$

Note that $r_{V,n}(x)$ is the radius of the largest open ball in $T^n V(x)$ centered at $T^n x$. In particular, $r_{W,0}(x) = \rho_W(x, \partial W)$.

Note that at the formulation of Condition A.7 an analogous quantity for $\hat{T}$, the function $\hat{r}$ has been used. However, throughout untill the end of Section 4.3.2, we may forget $\hat{T}$ and $\hat{r}$ and cosider the quantities for the original billiard map T.

First we describe how the above mentioned growth-fractioning process acts when the first iterate of T is applied. Given a δ_1-LUM (the constant δ_1 will be chosen later in Section 4.1.1) and some $\delta \geq 0$, we construct a subset $G_\delta(W) \subset W$, the (δ-)**gap** of W, that contains points that are δ-close to the first step singularity Γ in an appropriate sense. The complement of $G_\delta(W)$ will be denoted by $F_\delta(W)$ and will be referred to as the **remaining part** of W. The subscripts δ and/or the dependence on W will be sometimes omitted for brevity if no confusion arises. Then we show that this construction does not create too much new boundary: the sizes and shapes of the components of F and G can be controlled as expressed in our 1-step Lemma. To formulate it, recall that $K_{W,1}$ is the first step complexity from Definition 2.7, Λ is the factor of uniform expansion from Proposition 3.7, and introduce $\lambda = \Lambda^{1/100}$.

Lemma 4.3. *There are some global constants $D_1, \kappa_1, \sigma_1 > 0$ with the following property. Consider a δ_1-LUM W with the corresponding gap $G = G_\delta(W)$ and remaining part $F = F_\delta(W)$ constructed for some $0 \leq \delta < 1$. Then:*

(G0) $F_\delta \subset W \setminus (\Gamma^{[\delta]})$,

(G1) $m_W(r_{F_\delta,1} < \varepsilon) \leq \lambda^2 K_{W,1} \cdot m_W(r_{W,0} < \varepsilon/\Lambda)$,

(G2) $m_W(r_{G_\delta,0} < \varepsilon) \leq D_1 \delta^{-\kappa_1}\, m_W(r_{W,0} < \varepsilon)$,

(G3) $m_W(G_\delta) \leq D_1\, m_W(r_{W,0} < \delta^{\sigma_1})$.

Remark 4.4. *If $\delta = 0$, we have $G_\delta = W \cap \Gamma$, which is of zero Lebesgue measure. Thus (G2) and (G3) are trivial in this case, however, (G1) is important.*

Now if we knew $K_{W,1} < \Lambda^{1/2}$ say, Lemma 4.3 would essentially imply the three Growth Properties of Propostion 4.1 for $\hat{T} = T$. However, there is no reason for such a relation.

We would like to emphasize that the necessity of using a higher iterate of T is a special feature of *multidimensional* dispersing billiards. In the two dimensional case recent important observations, see [CD], made it possible to prove the growth properties for T itself, regardless of complexity. However, as the geometry is more complicated, it is not possible to adapt directly the exposition of [CD] to higher dimensions. If $d \geq 3$, complexity issues seem unavoidable, thus to gain enough expansion, it is essential to switch to some higher iterate of T. It seems, however, very difficult to consider higher iterates directly, as the higher order singularity manifolds do not possess uniform curvature bounds (see [BChSzT1] and [B]). Thus we perform an inductive argument: given a sufficiently small LUM W and $\delta \geq 0$, we construct the n-gap $G^n_\delta(W)$ and its complement, the n-remaining set $F^n_\delta(W)$ in an inductive manner. Then, with an inductive application of Lemma 4.3, we obtain our n-step Lemma:

Lemma 4.5. *Let $\delta_n = \delta_1^{3^n}$. For any fixed integer $n \in \mathbb{N}$, there exist global constants $\sigma_n, \kappa_n > 0$ and $D_n > 0$ with the following property. Consider an arbitrary δ_n-LUM W with the corresponding n-gap $G_\delta^n = G_\delta^n(W)$ and n-remaining part $F_\delta^n = F_\delta^n(W)$ constructed for some $0 \le \delta < 1$. Then we have:*

(Gn0) $F_\delta^n \subset W \setminus (\Gamma^{(n)})^{[\delta]}$,

(Gn1) $m_W(r_{F_\delta^n,n} < \varepsilon) \le \lambda^{4n} K_{W,n} \cdot m_W(r_{W,0} < \varepsilon/\Lambda^n)$,

(Gn2) $m_W(r_{G_\delta^n,0} < \varepsilon) \le D_n \delta^{-\kappa_n} m_W(r_{W,0} < \varepsilon)$,

(Gn3) $m_W(G_\delta^n) \le D_n m_W(r_{W,0} < \delta^{\sigma_n})$.

Remark 4.6. *It is worth noting that $D_n = D_1(K_1\lambda^3)^n$, $\sigma_n = \sigma_1/3^n$ and $\kappa_n = \kappa_{n-1}/3 + d - 1$ (thus, in paticular $\kappa_n \le \kappa_1 + 3(d-1)/2$ for all $n \in \mathbb{N}$).*

Remark 4.7. *If $\delta = 0$, we have $G_\delta^n = W \cap \Gamma^{(n)}$, which is of zero Lebesgue measure. Thus (Gn2) and (Gn3) are trivial in this case, however, (Gn1) is important. (See also Remark 4.4.)*

We will apply this n-step lemma for a fixed $n = n_0$, to be chosen below in Section 4.1.1. Considering $\hat{T} = T^{n_0}$, our Assumption 1.3 ensures that, for n_0 chosen appropriately, statement (Gn1) implies the first Growth Property of Proposition 4.1. (Gn2) and (Gn3) will imply the second and third Growth Properties, respectively, with the choice of $\delta_0 = \delta_{n_0}$, $D = D_{n_0}$, $\kappa = \kappa_{n_0}$ and $\sigma = \sigma_{n_0}$.

4.1.1 Further remarks and how the constants are chosen

Before turning to the constructions and the proofs in detail, we close this subsection with some further remarks on the exposition in general.

In the proof of Lemma 4.3 it is crucial that we can apply a locally flat picture. This is possible as LUM-s and the *first step* sigularity manifolds (i.e., the components of $\mathcal{S}$ and Γ) possess uniform curvature bounds. Thus, on sufficiently short distance scales, we may regard the intersections of LUMs and first step singularity manifolds as the intersection of $d-1$ dimensional flat disks with $2d-3$ dimensional hyperplanes in $\mathbb{R}^{2d-2}$. Moreover, by the alignment property, this intersection is transversal.

Recall that $\lambda = \Lambda^{1/100}$ (where $\Lambda > 1$ is the factor of uniform expansion). We choose δ_1 in Lemma 4.3 in such a way that, for $\delta_1^{1/3}$-LUMs, measure estimates based on the locally flat picture are accurate up to λ-precision. The reason for $1/3$ is that, given a δ_1-LUM W, Hölder continuity ensures that all the connected components of TW are $\delta_1^{1/3}$-LUMs, and thus can be regarded as locally flat pieces in the above sense. Two further requirements on the smallness of δ_1 is that distortions on this scale should not exceed λ, cf. Corollary 3.18, and that $K_{V,k} \le K_k$ for $k \le n_0$ and for any $\delta_1^{1/3}$-LUM V, see Remark 2.8.

Now in Lemma 4.5 $\delta_n = \delta_1^{3^n}$. By virtue of Hölder continuity this implies that, for a δ_n-LUM W, the components of $T^k W$ for all $k \leq n$ are δ_1-LUMs, and thus satisfy the hypotheses of Lemma 4.3.

Now when constructing G_δ we essentially cut out the δ-neighborhood of Γ from W (what exactly is happening is explained in Subsection 4.2 below). In the inductive construction of G_δ^n we apply this to some component of $T^k W$ ($k \leq n$), and pull back to W. More precisely, we apply the one step construction and, correspondingly, Lemma 4.3 to these components with $\delta \mapsto \delta'$, where δ' is some suitable positive power of δ. Then Hölder continuity ensures that we cut out some neighborhood of $T^{-k}\Gamma$ from W. The neighborhood of the singularity from which F_δ^n refrains will be smaller as we proceed with the induction. Thus what we can ensure in the end is that we are a certain positive distance away from the singularity – this is the δ that appears in the statement of Lemma 4.5. In particular, the measure estimates of Lemma 4.5 can be ensured in terms of this δ, despite of the fact that the distance to certain singularity components, cut out at earlier steps of the induction, will be much bigger than δ.

Finally, as there are many global constants appearing in different arguments, some of which depend on some others, here we summarize how these constants are chosen to make the expostion more transparent.

1. We have seen that some ε_g can be chosen in Definition 3.6, which ensures that the good metric inherits uniform expansion/contraction of unstable/stable vectors from the C-D metric, with some factor $\Lambda > 1$. The new metric satisfies transversality, alignment and the curvatures of LUMs and first step singularity manifolds are uniformly bounded, see Corollary 3.18 – by the choice of ε_g the constants appearing in these statments are fixed for the rest of the argument.

2. The next constant to fix is n_0, the integer power of the one-step dynamics we use as the map $\hat{T} = T^{n_0}$. The expansion factor for $\hat{T}$ is $\hat{\Lambda} = \Lambda^{n_0}$. By Assumption 1.3 we may ensure that "expansion prevails fractioning", that is, $K_{n_0} \leq \sqrt{\Lambda}^{n_0}$, which guarantees $K_{n_0}\lambda^{4n_0} = \alpha\Lambda^{n_0} = \alpha\hat{\Lambda}$ for some $\alpha < 1$. (cf. (Gn1) from Proposition 4.5 and the First Growth Property from Proposition 4.1).

3. The next constant we choose is k_0, the integer we start the labelling of homogeneity strips (and, correspondingly, secondary singularities) with (cf. Formulas (2.2)). We want k_0 to be so big that

   - k_0^2 is big enough,
   - $\sum_{k \geq k_0} k^{-2} \approx \frac{1}{k_0}$ is small enough

   in comparison to some other global constants.

   Once k_0 is fixed, the distortion bounds and the absolute continuity in Corollary3.18 get a precise formulation.

4. Finally, we fix the constant δ_1 which specifies the δ_1-LUM-s for which Lemma 4.3 is to be proven. We choose δ_1 small enough such that:
   - For any fixed $k = 1, ...n_0$, given any $\delta_1^{3^k}$-LUM W, the set $W \setminus \mathcal{S}^{(k)}$ has at most K_k components. This is possible by complexity and continuation.
   - $\delta_1^{1/3}$-LUMs are locally flat up to λ precision and distortions of T are at most λ on them (see the discussion in Section 4.1.1 above).

4.2 The one-step construction and its properties

4.2.1 Construction of the gap G_δ

Since we will now need to introduce a sequence of global constants depending on each other, please recall Notation 2.1 concerning the convention on C-s.

It is time to describe how the gap $G_\delta(W)$ (and correspondingly, the remaining part $F_\delta(W)$) are constructed. Pulling back the components of TW we get $W \setminus \mathcal{S} = \cup_{i=1}^{K_{W,1}} W_i$.

We will construct $G_\delta(W)$ as $\cup_{i=1}^{K_{W,1}} G_\delta(W_i)$. For fixed i the set $G_\delta(W_i)$ will cover all points of W_i that are in the δ-neighborhood of (either primary, or secondary) singularities intersecting W_i.

To treat the effect of secondary singularities consider TW_i, which lies, by definition, on a fixed scatterer. However, it may be partitioned by the homogeneity layers into countably many components to be denoted by $TW_{i,k} = TW_i \cap I_k$ ($k \geq k_0$). Here TW_{i,k_0} lies in the "middle of the phase space" while the further components lie in the vicinity of the boundary, cf. (2.2). Note, furthermore, that TW_i can be foliated by φ =const. hypersurfaces, corresponding to phase points where φ, the angle of incidence (cf. Notation 2.4) is fixed. These foliae will be denoted as γ_φ. In particular, the hypersurfaces that separate the neighboring $TW_{i,k}$-s from each other are elements of this foliation. In other words, $TW_{i,k}$ ($k > k_0$) consists of foliae γ_φ with $\frac{\pi}{2} - \frac{1}{k^2} < \varphi < \frac{\pi}{2} - \frac{1}{(k+1)^2}$.

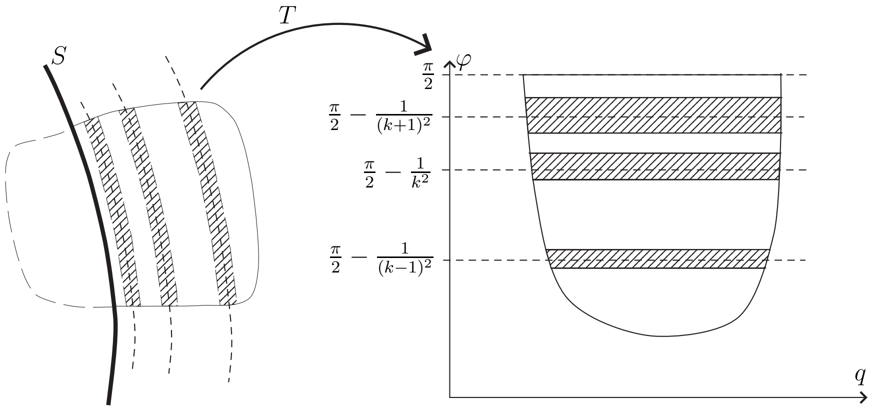

Figure 2: construction of the gap G_δ

Now we define:

$$G_\delta(W_i) = \left(W_i \cap \mathcal{S}^{[\delta]}\right) \bigcup \cup_{i,k} G_\delta(W_{i,k}), \tag{4.5}$$

where the construction of $G_\delta(W_{i,k})$ is described below. Let us, however, first note that the effect of the primary singularities is already taken care of in the first set appearing in the above expression.

In $G_\delta(W_{i,k})$ we will consider the effect of secondary singularities as it appears in $W_{i,k}$. Fix a global constant C_1, the value of which will be chosen below (it is determined by the alignment property and the constants of Corollary 2.11). Consider $TW_{i,k}$ which, by the above description, can be visualized as a narrow strip consisiting of γ_φ-s which satisfy the required bounds. Now $TG_\delta(W_{i,k})$ will consist of the two exterior substrips of $TW_{i,k}$, i.e. the sets made up of the foliae γ_φ with $\frac{\pi}{2} - \frac{1}{k^2} < \varphi \leq \frac{\pi}{2} - \frac{1}{k^2} + C_1\delta k^2$ on the one hand, and with $\frac{\pi}{2} - \frac{1}{(k+1)^2} - C_1\delta k^2 \leq \varphi < \frac{\pi}{2} - \frac{1}{(k+1)^2}$ on the other hand. See Figure 2.

Now when pulling back this set to W by T, $G_\delta(W_{i,k})$ will consist of two regions around two neighboring elements of the (secondary) singularity set Γ. Corollary 2.11 ensures that width of these regions is $\sim \delta$. Exploiting this fact along with the alignment property we may choose C_1 above in such a way that G_δ covers $W \cap (\Gamma^{[\delta]})$.

Remark 4.8. *It is worth noting that, for k big enough (more precisely, for $k > C\delta^{-\frac{1}{5}}$), the two exterior strips overlap. In such a case we cut out the full $W_{i,k}$: G_δ will consist of at most $K_{W,1}$ thicker strips, corresponding to overlaps coming from big k, and narrower strips, coming from smaller k where there is no overlap. In particular, by straightforward calculations:*

(i) the number of boundary components of G_δ does not exceed $C\delta^{-\frac{1}{5}}$.

(ii) the width, and consequently, the measure of the thicker strips does not exceed $C\delta^{\frac{4}{5}}$.

This construction may seem too complicated at first sight. However, it has several advantages that will help us to prove Lemma 4.3 with relatively simple arguments. Most importantly, (most of) the boundary components of G_δ (and thus of F_δ), defined this way, are pre-images of certain foliae γ_φ as well. This ensures that Corollary 2.11 applies to them, which is very useful when proving (G1), i.e. when estimating the measure of points that *will lie* in the ε neighborhood of these boundary components.

Note that for a simpler choice of G_δ – setting simply $G_\delta = \Gamma^{(\delta)} \cap W$, say – it would be much more difficult to check (G1) in lack of direct applicability of Corollary 2.11.

Remark 4.9. *It is useful to note that the construction of the gap does only depend on the singularity set. In particular, given δ_1-LUMs W and $\tilde{W} \subset W$ we have $G_\delta(\tilde{W}) = G_\delta(W) \cap \tilde{W}$ for any δ.*

4.2.2 Proof of Lemma 4.3

Statement (G0) follows by the construction of G_δ and its properties described above.

To prove statements (G1)-(G3), first we state two simple geometrical sublemmas. For both of them consider any nonempty bounded measurable set $W \subset \mathbb{R}^m$, and a 1-codimensional plane $E \subset \mathbb{R}^m$. E cuts $\mathbb{R}^m$ into two half-spaces, which we will call 'left' and 'right'. Accordingly, W is cut into a 'left' and 'right' part, W_l and W_r (one of these may be empty). Our sublemmas will compare sets of points in W near different parts of the boundary. We will apply them with $m = d - 1$, i.e. on u-manifolds. For the proof see Appendix B.

Sublemma 4.10. *For any* $\varepsilon \geq 0$

$$\mathrm{Leb}(\{x \in W_l \,|\, \rho(x, \partial W_l) < \varepsilon\}) \leq \mathrm{Leb}(\{x \in W \,|\, \rho(x, \partial W) < \varepsilon\}) \tag{4.6}$$

and the same holds for W_r. *See Figure 3.*

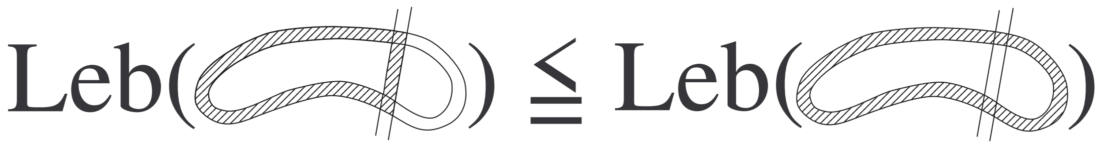

Figure 3: the statement of Sublemma 4.10

Sublemma 4.11. *For any* $\varepsilon \geq 0$ *and* $0 \leq \xi \leq 1$,

$$\mathrm{Leb}\left(\{x \in W \,|\, \rho(x, E) \leq \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x, \partial W) \leq \varepsilon\}\right) \;\leq\; \xi \mathrm{Leb}(\{x \in W \,|\, \rho(x, \partial W) \leq \varepsilon\}). \tag{4.7}$$

See Figure 4.

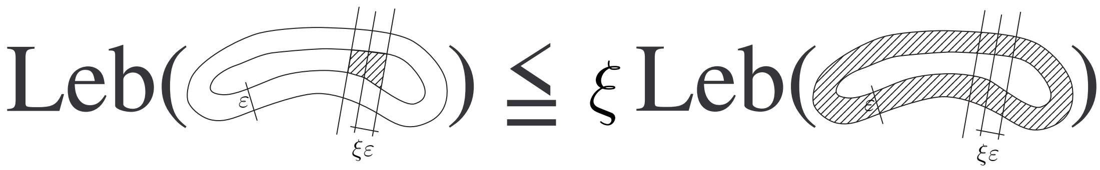

Figure 4: The statement of Sublemma 4.11

Let us prove (G2) first. What appears in addtion to $\{r_{W,0} < \varepsilon\}$ on the LHS is, by Remark 4.8, the ε neighborhood of finitely many hyperplanes within W. Moreover, the number of these hyperplanes does not exceed $C\delta^{-\frac{1}{5}}$. On the other hand, applying Sublemma 4.11 (with $\xi = 1$), we have an upper bound $m_W(r_{W,0} < \varepsilon)$ on the contribution of any such hyperplane. This proves (G2) with $\kappa_1 = \frac{1}{5}$.

To prove (G3) we first give an upper bound on its LHS, i.e., on the measure of G_δ. We again envoke Remark 4.8. G_δ consists, on the one hand, of finitely many "thick" strips (of measure estimated by Remark 4.8/(ii)), the number of which is uniformly bounded above. Thus their contribution to $m_W(G_\delta)$ is not more than $C\delta^{\frac{4}{5}}$. On the other hand, we have at most $C\delta^{-\frac{1}{5}}$ many components of width (and thus measure) δ. Their overall contribution to $m_W(G_\delta)$ does not exceed $C\delta^{\frac{4}{5}}$. Altogether we have the upper bound $C\delta^{\frac{4}{5}}$ on the LHS of (G3).

We complete the proof of (G3) with $\sigma_1 = \frac{4}{5(d-1)}$. The case when the whole of W is contained in the set

$$W_c := \{r_{W,0} < \delta^{\sigma_1}\}$$

is trivial. If, on the other hand, $W \setminus W_c \neq \emptyset$, then it is easy to see that W_c does necessarily contain two disjoint hemispheres of radius δ^{σ_1}. Thus, for the measure that appears on the right hand side of (G3):

$$m_W(W_c) \geq \Gamma_{d-1}(\delta^{\sigma_1})^{d-1} = C\delta^{\frac{4}{5}}$$

where Γ_{d-1} is the volume of the $(d-1)$-dimensional unit ball. This means that (G3) holds in this case as well if D_1 is chosen appropriately.

To prove (G1) recall the notation for W_i from Section 4.2.1 for $i = 1, ..., K_{W,1}$. Introduce furthermore

$$\begin{aligned} F_i &:= W_i \setminus \mathcal{S}^{[\delta]}, \\ F_i^{\varepsilon/\Lambda} &:= \{\, x \in F_i \,|\, \rho_W(x, \partial F_i) < \varepsilon/\Lambda \,\}, \\ F_{i,+}^{\varepsilon} &:= \{\, x \in F_i \,|\, \rho_{TW}(Tx, T\partial F_i) < \varepsilon \,\}. \end{aligned} \tag{4.8}$$

Our first observation is that

$$F_{i,+}^{\varepsilon} \subset F_i^{\varepsilon/\Lambda} \tag{4.9}$$

which follows from the fact that T expands distances on u-manifolds uniformly by a factor Λ. This is a trivial observation, nonetheless, it is important to emphasize how hard we worked for it (this is the reason why we had to introduce a new metric) and what an important role it plays. It is (4.9) that enables us to reduce the proof of the growth lemmas to estimates on the one step dynamics (i.e. to Lemma 4.3).

Before proceeding we note that if we had only primary singularities, $F_\delta \cap W_i$ would coincide with F_i, and the set of points in W for which $r_{F_\delta,1} < \varepsilon$ – that is, the set which appears on the LHS of (G1) – would coincide with $\cup_i F_{i,+}^{\varepsilon}$. The second key observation is that "the contribution of secondary singularities is negligible", that is, for any i:

$$m_W\,(\, x \in F_i \,|\, r_{F_\delta,1}(x) < \varepsilon \,) \leq \lambda m_W(F_i^{\varepsilon/\Lambda}). \tag{4.10}$$

To prove (4.10) consider the connected components of $T(F_i \cap F_\delta)$. These connected components have boundaries of two different types. On the one hand there is $T\partial F_i$, arising form primary singularities. On the other hand, recalling the details of the construction from Section 4.2.1, we see that the secondary singularities give rise to further boundary components, namely countably many foliae γ_{φ_k} for $\varphi_k \approx \frac{\pi}{2} - \frac{1}{k^2}$ with $k \geq k_0$ [3]. We may denote these foliae as $TE_{i,k}$, where $E_{i,k} \subset F_i$ is a hyperplane by our convention of local flatness. It is worth noting that it is crucial that we may apply a locally flat approximation as only first step singularity manifolds appear, cf. Section 4.1.1.

Corresponding to the above characterization of boundary components:

$$\{ x \in F_i \,|\, r_{F_\delta,1}(x) < \varepsilon \} \subset F^{\varepsilon}_{i,+} \cup \left(\cup_{k=k_0}^{\infty} \{ x \in F_i \,|\, \rho_{TW}(Tx, TE_{i,k}) < \varepsilon \} \right). \tag{4.11}$$

Let us concentrate on the contribution of secondary singularities. Now we may reveal why it was so important to construct G_δ as explained in Section 4.2.1: we have that $TE_{i,k}$ are themselves constant φ foliae. Thus, by consecutive applications of Corollary 2.11 and the alignment property:

$$\{ x \in F_i \,|\, \rho_{TW}(Tx, TE_{i,k}) < \varepsilon \} \subset \{ x \in F_i \,|\, \rho_W(x, E_{i,k}) < \frac{C\varepsilon}{k^2} \}.$$

Applying Sublemma 4.11 (with $\varepsilon \to \varepsilon/\Lambda$ $\xi \to \frac{C\Lambda}{k^2}$) plugged into (4.11), along with (4.9) implies

$$m_W \left(x \in F_i \,|\, r_{F_\delta,1}(x) < \varepsilon \right) \leq \left(1 + \sum_{k=k_0}^{\infty} \frac{C}{k^2} \right) m_W(F_i^{\varepsilon/\Lambda}).$$

We may choose k_0 so big that (4.10) holds.

To complete the proof, note that by the continuation property and the convention on local flatness F_i can be considered as the result of the following process: start with W, cut it along a hyperplane, keep one of the two pieces, cut it again along a hyperplane and repeat the above for finitely many (at most $K_{1,W}$) times. (See Section 2.4.2, especially Figure 1.) Thus, by consecutive applications of Lemma 4.10 we see that

$$m_W(F_i^{\varepsilon/\Lambda}) \leq m_W(r_{W,0} < \varepsilon/\Lambda). \tag{4.12}$$

Plugging this into (4.10) and summing over i completes the proof of (G1).

To terminate, we admit that we have cheated a little bit by using a locally flat picture, which is true only up to λ-precision. Thus, further λ factors appear on the RHS-s of the obtained estimates. As for statements (G2) and (G3), this can be swallowed in a suitably chosen D_1, while we have a prefactor $\lambda^2 K_{1,W}$ in (G1), which corresponds exactly to the claim.

Remark 4.12. *In the Corollary to follow, and throughout the rest of the section, we will use the following notation. Whenever ∂TW appears, it is understood in terms of the modified phase space of Section 2.2.3, i.e., TW is cut by secondary singularities.*

[3] Having a closer look at the exposition of Section 4.2.1, we see that (i) only finitely many such foliae contribute, nonetheless, their number is unbounded, more precisely, is bounded only in terms of δ, cf. Remark 4.8; (ii) for each k there are two such foliae, corresponding to the two exterior strips within the kth homogeneity layer. These details, however, do not modify the exposition, thus we disregard them, to avoid overcomplified notation.

Corollary 4.13. *The statement (G1) from Lemma 4.3 has a special significance. In particular, it can be formulated for $\delta = 0$ (cf. Remark 4.4). As it contains no additional information on gap construction, this is the version that indeed estimates how much new boundary is created by the singularity manifolds when T is applied. We formulate it for future record. Given a δ_1-LUM W, for any $\varepsilon > 0$ we have:*

$$m_W(\{x \in W | \rho(Tx, \partial TW) < \varepsilon\}) \leq \lambda^2 K_{1,W} \cdot m_W(r_{W,0} < \varepsilon/\Lambda). \tag{4.13}$$

4.3 The n-step construction and its properties

4.3.1 Construction of the n-gap G_δ^n

We construct $G_\delta^n(W)$ by induction. We will use the construction of Section 4.2 repeatedly. In particular, the first step of the induction is exactly the first step construction described there.

Recall that at the nth step of the induction we need to treat δ_n-LUMs with $\delta_n = \delta_1^{3^n}$.

Now assume inductively that, given an arbitrary δ_{n-1}-LUM, we already know how to construct the relevant $n-1$-gap for $0 \leq \delta < 1$.

To proceed consider a δ_n-LUM W. Below we describe how for a given $0 \leq \delta < 1$ the n-gap G_δ^n (and correspondingly, the n-remaining part F_δ^n) in W is to be constructed. As W is a δ_n-LUM, it is, in particular, a δ_1-LUM, thus the whole exposition of subsection 4.2 applies to it. This means we may consider its 1-gap $G_\delta(W)$, its primary components W_i and its secondary components $W_{i,k}$.

Furthermore, the $TW_{i,k}$-s are δ_{n-1}-LUMs by the Hölder continuity of the dynamics (note $\delta_{n-1} = \delta_n^{1/3}$). By our inductive assumption the $(n-1)$-gaps for any $TW_{i,k}$ can be constructed for $0 \leq \delta < 1$. In particular, let us construct the $(n-1)$-gap for any such $TW_{i,k}$ with $\delta \to \delta^{1/3}$. We will denote these $(n-1)$-gaps lying in some fixed $TW_{i,k}$ by

$$G_{i,k} := G_{\delta^{1/3}}^{n-1}(TW_{i,k}) \subset TW_{i,k}. \tag{4.14}$$

To construct the n-gap G_δ^n for W, we need to identify all points that get close to some singularity manifold within the first n iterates. Points that are close to some singularity right now are contained in the 1-gap $G_\delta(W)$. We need to add those points x for which $T^i x$ is close to some singularity manifold for some $i = 1, ..., n$. These are exactly the preimages of points that start out from some $TW_{i,k}$ and get close to some singularity within the first $n-1$ iterates, in other words, the preimages of $G_{i,k}$. This is the reason for constructing the n-gap as:

$$\begin{aligned} G_\delta^n(W) &= G_\delta(W) \cup {G_\delta}^F(W), \\ {G_\delta}^F(W) &= \cup_{i,k} T^{-1}(G_{i,k}) \setminus G_\delta(W), \end{aligned} \tag{4.15}$$

where the superscript F stands for future.

Remark 4.14. *As a consequence of our observation in Remark 4.9, we have that the construction of the n-gap depends only on the (n-step) singularity set as well. In particular, given δ_n-LUMs W and $\tilde{W} \subset W$ we have $G_\delta^n(\tilde{W}) = G_\delta^n(W) \cap \tilde{W}$ for any δ (and any n).*

4.3.2 Proof of Lemma 4.5

We will prove Lemma 4.5 for constants κ_n, σ_n and D_n satisfying the following recursive relations:

$$\sigma_n = \sigma_{n-1}/3, \qquad \kappa_n = \kappa_{n-1}/3 + (d-1), \qquad D_n = D_{n-1} K_1 \lambda^3. \tag{4.16}$$

This is the reason for the relations mentioned in Remark 4.6.

All the statements (Gn0)-(Gn3) from Lemma 4.5 are proved by induction on n, nonetheless, these inductions are independent. Before performing the four inductions, let us mention that the general strategy for all of them is essentially the same. This strategy relies on the decomposition (4.15). In rough terms, to prove (Gn1)-(Gn3) we are to establish upper bounds on the measure of certain sets related to $G_\delta^n(W)$. The contribution of $G_\delta(W)$ is simply estimated by Lemma 4.3 (i.e., Formulas (G0)-(G3)). To treat ${G_\delta}^F(W)$ we assume inductively that we already have the relevant bound on $G_{i,k}$, the $(n-1)$-gap of the δ_{n-1}-LUM $TW_{i,k}$, for any i and k. "Pulling back" the relevant estimates to W relies on two key observations. On the one hand, the $TW_{i,k}$ are homogeneous u-manifolds, thus we have bounds on the distortions of T, which help us express estimates in terms of measures in W. On the other hand, by means of Corollary 4.13, we can reformulate our estimates in terms of distances on W.

To prove (Gn0) we make the following observations. Any point in W that lies in the δ neighborhood of Γ belongs to $G_\delta(W)$ by (G0). On the other hand, by the inductive assumption, if $x' \in TW_{i,k}$ (for some i, k fixed) lies in the $\delta^{1/3}$ neighborhood of $\Gamma^{(n-1)}$, then $x' \in G_{i,k}$. Thus, by Hölder continuity of T, if $x (= T^{-1}x') \in W$ lies in the δ neighborhood of $T^{-1}\Gamma^{(n-1)}$, it should belong to ${G_\delta}^F$. The fact that $\Gamma^{(n)} = \Gamma \cup T^{-1}\Gamma^{(n-1)}$ completes the proof of (Gn0).

To prove (Gn3) we need to provide an upper bound on the measure of $G_\delta^n(W)$. On the one hand, by (G3):

$$m_W(G_\delta) \le D_1 m_W(r_{W,0} < \delta^{\sigma_1}). \tag{4.17}$$

To estimate the measure of ${G_\delta}^F$, we use our inductive assumption on $G_{i,k}$ for any i, k fixed. Recall that $G_{i,k}$ is the $n-1$ gap for $\delta^{1/3}$ in $TW_{i,k}$. Since by (4.16) we have $(\delta^{1/3})^{\sigma_{n-1}} = \delta^{\sigma_n}$, this means that the inductive assumption – i.e., (Gn3) formulated for $n \to n-1$ and $\delta \to \delta^{1/3}$ in $TW_{i,k}$ – reads as

$$m_{TW_{i,k}}(G_{i,k}) \le D_{n-1} m_{TW_{i,k}}(r_{TW_{i,k},0} < \delta^{\sigma_n}). \tag{4.18}$$

Now we are going to use that $TW_{i,k}$ is, on the one hand, a δ_1-LUM, and, on the other hand, it is homogeneous. Thus we have that T distorts measures on it at most by a factor

λ (recall how the constants are chosen from subsection 4.1.1). Thus (4.17) implies:

$$m_{W_{i,k}}(T^{-1}G_{i,k}) \le D_{n-1}\lambda m_{W_{i,k}}(r_{W_{i,k},1} < \delta^{\sigma_n}).$$

We may sum first over k and then over i to obtain

$$m_W({G_\delta}^F) \le D_{n-1}\lambda\, m_W(\{x \in W | \rho(Tx, \partial TW) < \delta^{\sigma_n}\}). \tag{4.19}$$

where the RHS is understood according to Remark 4.12.

Now we apply Corollary 4.13 with $\varepsilon \to \delta^{\sigma_n}$:

$$m_W(\{x \in W | \rho(Tx, \partial TW) < \delta^{\sigma_n}\}) \le K_{W,1}\lambda^2\, m_W(r_{W,0} < \delta^{\sigma_n}/\Lambda) \le K_1\lambda^2\, m_W(r_{W,0} < \delta^{\sigma_n}). \tag{4.20}$$

Note that $D_n = D_{n-1}\lambda^3 K_1$ (4.16). Furthermore, as $D_1 \le D_n$ and $\sigma_n \le \sigma_1$, the decomposition (4.15) along with the three inequalities (4.17), (4.19) and (4.20) altogether imply (Gn3).

As a preparation for the inductive proof of (Gn2) we state the following geometric sublemma. For the proof see Appendix B.

Sublemma 4.15. *Let $W \in \mathbb{R}^m$ be any nonempty bounded measurable set, $\varepsilon \ge 0$ and $k > 1$. Then*

$$\mathrm{Leb}(\{x \in W \mid \rho(x, \partial W) \le k\varepsilon\}) \le k^m \mathrm{Leb}(\{x \in W \mid \rho(x, \partial W) \le \varepsilon\}).$$

See Figure 5.

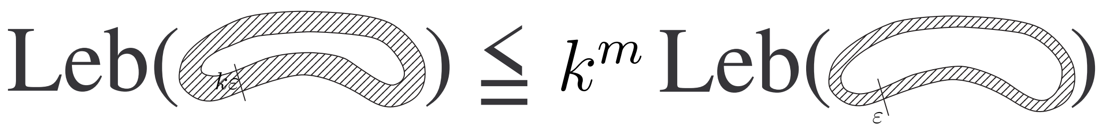

Figure 5: the statement of Sublemma 4.15

Now we can prove (Gn2). By (4.15):

$$\{x \in W | r_{G_\delta^n,0}(x) < \varepsilon\} \subset \{x \in G_\delta | r_{G_\delta^n,0}(x) < \varepsilon\} \cup \{x \in {G_\delta}^F | r_{G_\delta^n,0}(x) < \varepsilon\}.$$

For the first component we apply (G2):

$$m_W(r_{G_\delta,0} < \varepsilon) \le D_1\delta^{-\kappa_1} m_W(r_{W,0} < \varepsilon), \tag{4.21}$$

while the second component, as a subset of ${G_\delta}^F$, is at least δ away from the first-step singularity set Γ. This implies that T expands distances *at most* by a factor δ on this set (due to 2 in Section 2.5.1) and Corollary 3.18).

As a consequence, recalling also (4.15):

$$\{x \in {G_\delta}^F | r_{G_\delta^n,0}(x) < \varepsilon\} \subset T^{-1}\{x \in TW_{i,k} | r_{G_{i,k},0}(x) < \varepsilon/\delta\}. \tag{4.22}$$

We estimate the contribution for fixed i and k. In particular, we assume inductively that $G_{i,k}$, as the $(n-1)$-gap for $\delta \to \delta^{1/3}$ within $TW_{i,k}$ satisfies (Gn2) in the relevant form:

$$m_{TW_{i,k}}(r_{G_{i,k},0} < \varepsilon/\delta) \le D_{n-1}\delta^{-\kappa_{n-1}/3} m_{TW_{i,k}}(r_{TW_{i,k},0} < \varepsilon/\delta).$$

We apply distortion bounds on T restricted to $W_{i,k}$ the same way as in the proof of (Gn3):

$$m_{W_{i,k}}(r_{T^{-1}G_{i,k},1} < \varepsilon/\delta) \le D_{n-1}\lambda\delta^{-\kappa_{n-1}/3} m_{W_{i,k}}(r_{W_{i,k},1} < \varepsilon/\delta),$$

and sum over k and i (recall Remark 4.12):

$$m_W(\{x \in {G_\delta}^F | r_{G_\delta^n,0}(x) < \varepsilon\}) \le D_{n-1}\lambda\delta^{-\kappa_{n-1}/3} m_W(\{\, x \in W \mid \rho(Tx, \partial TW) < \varepsilon/\delta\}). \tag{4.23}$$

Applying Corollary 4.13 to the RHS of (4.23) – the same way as in (4.20) – implies:

$$m_W(\{x \in {G_\delta}^F | r_{G_\delta^n,0}(x) < \varepsilon\}) \le D_n\delta^{-\kappa_{n-1}/3} m_W(r_{W,0} < \varepsilon/\delta).$$

Finally let us invoke Sublemma 4.15 with $k = 1/\delta$ to bound the RHS from above:

$$m_W(r_{W,0} < \varepsilon/\delta) \le \delta^{-(d-1)} m_W(r_{W,0} < \varepsilon),$$

which completes the inductive proof of (Gn2) as $\delta^{-\kappa_n} = \delta^{-\kappa_{n-1}/3}\delta^{-(d-1)}$ by (4.16).

To prove (Gn1) we need to introduce some more notation. By definition, $F_\delta = W \setminus G_\delta$ and $F_\delta^n = W \setminus G_\delta^n$. Furthermore let

$$\begin{aligned} \tilde{W}_{i,k} &:= W_{i,k} \cap F_\delta \ (\subset W_{i,k} \subset W), \\ F_{i,k} &:= T\tilde{W}_{i,k} \setminus G_{i,k} \ (\subset TW_{i,k} \subset TW). \end{aligned} \tag{4.24}$$

Note that

$$F_\delta = \cup_{i,k}\tilde{W}_{i,k} \quad \text{and} \quad F_\delta^n = \cup_{i,k} T^{-1} F_{i,k}$$

where the unions are disjoint. This implies

$$\{x \in F_\delta | r_{F_\delta,1}(x) < \varepsilon\} = \cup_{i,k}\{x \in \tilde{W}_{i,k} | r_{\tilde{W}_{i,k},1}(x) < \varepsilon\}, \tag{4.25}$$

and

$$\{x \in F_\delta^n | r_{F_\delta^n,n}(x) < \varepsilon\} = \cup_{i,k}\{x \in T^{-1}F_{i,k} | r_{T^{-1}F_{i,k},n}(x) < \varepsilon\}. \tag{4.26}$$

On the other hand – recalling also Remarks 4.9 and 4.14 – by (4.24) $F_{i,k}$ is the $(n-1)$ remaining part in $T\tilde{W}_{i,k}$. Thus we may assume inductively that (Gn1) holds for $n-1$:

$$m_{T\tilde{W}_{i,k}}(r_{F_{i,k},n-1} < \varepsilon) \le K_{T\tilde{W}_{i,k},n-1}\lambda^{4(n-1)} m_{T\tilde{W}_{i,k}}(r_{T\tilde{W}_{i,k},0} < \varepsilon/\Lambda^{n-1}).$$

As the $T\tilde{W}_{i,k}$ are homogeneous δ_1-LUM-s, the distortions of T are suitably bounded and we have

$$m_{\tilde{W}_{i,k}}(r_{T^{-1}F_{i,k},n} < \varepsilon) \le K_{T\tilde{W}_{i,k},n-1}\lambda^{4(n-1)}\lambda m_{\tilde{W}_{i,k}}(r_{\tilde{W}_{i,k},1} < \varepsilon/\Lambda^{n-1}). \tag{4.27}$$

Note that $K_{.,.}$ describes the effect of primary singularities, thus it can be defined for any (not necessarily homogeneous) unstable manifold, in particular, we may consider the quantity $K_{TW_i,n-1}$. We also have

$$K_{T\tilde{W}_{i,k},n-1} \le K_{TW_i,n-1}$$

by means of which, keeping i *fixed*, we sum (4.27) over k:

$$m_{W_i}(\{x \in W_i | r_{F_\delta^n,n}(x) < \varepsilon\}) \le K_{TW_i,n-1}\lambda^{4(n-1)}\lambda m_{W_i}(\{x \in W_i | r_{F_\delta,1}(x) < \varepsilon/\Lambda^{n-1}\}). \tag{4.28}$$

Here we have also used the characterizations (4.25) and (4.26).

To bound the RHS from the above, first we apply (G1) to W_i for i fixed (with $\varepsilon \to \varepsilon/\Lambda^{n-1}$):

$$m_{W_i}(r_{F_\delta,1} < \varepsilon/\Lambda^{n-1}) \le K_{W_i,1}\lambda^2 m_{W_i}(r_{W_i,0} < \varepsilon/\Lambda^n).$$

To proceed recall that, as long as only primary singularities are concerned, the W_i are the smooth components of W. Thus $K_{W_i,1} = 1$ and

$$m_{W_i}(r_{F_\delta,1} < \varepsilon/\Lambda^{n-1}) \le \lambda^2 m_{W_i}(r_{W_i,0} < \varepsilon/\Lambda^n). \tag{4.29}$$

By means of the continuation property, we apply, as we did in the proof of (4.12) in subsection 4.2, Sublemma 4.10:

$$m_{W_i}(r_{W_i,0} < \varepsilon/\Lambda^n) \le \lambda m_W(r_{W,0} < \varepsilon/\Lambda^n).$$

Here the additional λ factor appears as the error term of the locally flat estimate. This last formula gives, along with (4.29) and (4.28) and summation on i:

$$m_W(r_{F_\delta^n,n} < \varepsilon) \le \left(\sum_{i=1}^{K_{W,1}} K_{TW_i,n-1} \right) \lambda^{4n} m_W(r_{W,0} < \varepsilon/\Lambda^n).$$

Finally, as a consequence of the continuation property we have

$$\left(\sum_{i=1}^{K_{W,1}} K_{TW_i,n-1} \right) = K_{W,n}$$

which completes the inductive proof of (Gn1).

4.4 Proof of Proposition 4.1

To complete the proof of the Growth Properties for $\hat{T} = T^{n_0}$, choose n_0 according to the exposition of Section 4.1.1. In particular, by Assumption 1.3, $K_{n_0}\lambda^{4n_0} = \alpha\hat{\Lambda}$ for some $\alpha < 1$ (here $\hat{\Lambda} = \Lambda^{n_0}$ is the factor of expansion for $\hat{T}$).

The constants for which we prove Propostion 4.1 are the above mentioned α along with $\delta_0 = \delta_{n_0}$, $\sigma = \sigma_{n_0}$, $D = D_{n_0}$ and $\kappa = \kappa_{n_0}$ chosen according to Lemma 4.5 (see also Remark 4.6).

Let us consider an arbitrary $\delta_0(=\delta_{n_0})$-LUM W, and an arbitrary $0 \leq \delta < 1$. To prove the Proposition, first we should tell what the sets V_δ^0 and V_δ^1 are. As W is a δ_{n_0}-LUM, we may apply the $n(=n_0)$ step construction of Section 4.3.1 to it and construct $G_\delta^{n_0}(W)$ and $F_\delta^{n_0}(W)$. Now define:

$$V_\delta^0 = G_\delta^{n_0}(W), \qquad W_\delta^1 = F_\delta^{n_0}(W).$$

Note that as $G_\delta^{n_0}$ and $F_\delta^{n_0}$ make a partition of W, we have

$$m_W(W \setminus (V_\delta^0 \cup W_\delta^1)) = 0, \tag{4.30}$$

and by (Gn0)

$$W_\delta^1 \subset W \setminus \hat{\Gamma}^{[\delta]}. \tag{4.31}$$

However, we cannot use W_δ^1 as V_δ^1. The reason is that, in terms of $\hat{T}$, W_δ^1 is not a $(\delta_0, 1)$-subset. In paticular, we should have that the components of $\hat{T}V_\delta^1$ have diameter less than δ_0. The construction of Section 4.3.1, on the other hand, ensures only that the components of $\hat{T}W_\delta^1$ have diameter less than $\delta_1^{1/3}$, which is much greater than δ_0.

To obtain smaller components, we construct below V_δ^1 by removing sets of zero m_W-measure from W_δ^1. This, of course, does not spoil the validity of (4.30) and (4.31).

To proceed we may reformulate statements (Gn1)-(Gn3) from Lemma 4.5 in terms of $\hat{T} = T^{n_0}$. Keeping also in mind how $r_{W,n}$ is defined, (Gn1) reads as

$$m_W(\{\, x \in W_\delta^1 \mid \rho(\hat{T}x, \partial\hat{T}W_\delta^1(x)) < \varepsilon \,\}) \leq \alpha\hat{\Lambda} \cdot m_W(\{\, x \in W \mid \rho(x, \partial W) < \varepsilon/\hat{\Lambda} \,\}) \tag{4.32}$$

(Gn2) as

$$m_W(\{\, x \in V_\delta^0 \mid \rho(x, \partial V_\delta^0(x)) < \varepsilon \,\}) \leq D\delta^{-\kappa}\, m_W(\{\, x \in W \mid \rho(x, \partial W) < \varepsilon \,\}); \tag{4.33}$$

and (Gn3) as

$$m_W(V_\delta^0) \leq D\, m_W(\{\, x \in W \mid \rho(x, \partial W) < \delta^\sigma \,\}). \tag{4.34}$$

Note that (4.33) and (4.34) are exactly the Second and the Third Growth Properties, respectively.

Furthermore, (4.32) is almost the First Growth Property, actually, it is an even better upper bound on the set of points that get close to the boundaries of $\hat{T}W_\delta^1$. Recall that we need to partition the components of W_δ^1 into smaller pieces to arrive at V_δ^1. We will see that the contribution of these additional boundary components can be estimated by the second term that appears on the RHS of (4.1).

By the exposition of Section 4.3.1, the set W^1_δ has finitely many components (the number of which, actually, depends on δ). Consider any such component and denote it by Δ^1: we would like to chop Δ^1 into pieces which do not grow larger then δ_0 in diameter when $\hat{T}$ is applied. Our argument will roughly follow [Ch1].

What we know is that $\hat{T}\Delta^1$ is a $\delta_1^{1/3}$-LUM, thus by our convention on local flatness, in measure related calculations it can be considered as a piece of $\mathbb{R}^{d-1}$. Furthermore, it is a homogeneous LUM, thus the distortions of $\hat{T}$ restricted to Δ^1 are suitable bounded.

We shall work in $\mathbb{R}^{d-1}$. Let us fix $\delta' = \frac{\delta_0}{\sqrt{2(d-1)}}$. For given numbers $a_1, ..., a_{d-1}$ such that $0 \le a_i < \delta'$; $i = 1, ..., d-1$ consider the $d-1$ families of hyperplanes:

$$L_{a_i} := \{(x_1, ..., x_{i-1}, a_i + n_i\delta', x_{i+1}, ..., x_{d-1}) \,|\, n_i \in \mathbb{Z}\} \qquad i = 1, ..., d-1.$$

For example, the intersection $\hat{T}\Delta^1 \cap L_{a_1}$ consists of parallel hyperplanar pieces inside $\hat{T}\Delta^1$. Let us denote their altogether $(d-2)$-dimensional area by A_{a_1}. By Fubini theorem

$$\int_0^{\delta'} A_{a_1} da_1 = m_{\hat{T}W}(\hat{T}\Delta^1).$$

Thus there is definitely one particular a_1' for which

$$A_{a_1'} \le \frac{m_{\hat{T}W}(\hat{T}\Delta^1)}{\delta'}. \tag{4.35}$$

We may repeat the above argument for all the other coordinates to get the numbers a_i' for which an inequality analogous to (4.35) holds. For brevity we introduce the notation:

$$L := \bigcup_{i=1}^{d-1} L_{a_i'}$$

and, furthermore, we denote the $d-2$-dimensional area of $L \cap \hat{T}\Delta^1$ by A. We shall add, inside Δ^1, to W^1_δ the pre-image of L to get V^1_δ. To estimate the new boundary term one further notation is introduced:

$$\Omega := \{x \in \Delta^1 \,|\, \rho(\hat{T}x, \hat{T}L) < \varepsilon \text{ and } \rho(\hat{T}x, \partial\hat{T}\Delta^1) \ge \varepsilon\}.$$

Note that, for $x \in \Delta^1 \subset W^1_\delta$, $W^1_\delta(x) = \Delta^1$, thus our aim is to estimate $m_W(\Omega)$. By the above formulas:

$$m_{\hat{T}W}(\hat{T}\Omega) \le 2\varepsilon A \le 2\varepsilon(d-1)\frac{m_{\hat{T}W}(\hat{T}\Delta^1)}{\delta'}.$$

Now as the distortions of $\hat{T}$ restricted onto Δ^1 are bounded, we have:

$$m_W(\Omega) < \varepsilon\delta_0^{-1}\beta m_W(\Delta^1)$$

where the global constant $\beta > 0$ depends, in addition to distortion bounds, on the accuracy of the locally flat approximation as well. We chop all the Δ^1-s (the connected components of W^1_δ) according to this machinery to get V^1_δ. After summation we get:

$$m_W(\{x \in V^1_\delta \mid \rho(\hat{T}x, \partial \hat{T} V^1_\delta(x)) < \varepsilon\}) - m_W(\{x \in W^1_\delta \mid \rho(\hat{T}x, \partial \hat{T} W^1_\delta(x)) < \varepsilon\}) \leq \leq \varepsilon \delta_0^{-1} \beta m_W(W). \tag{4.36}$$

Now (4.36) along with (4.32) implies (4.1). This completes the proof of the First Growth Property, and thus of Propostion 4.1.

4.5 The conditions of ergodicity

Following the program outlined in the Introduction, the arguments presented in the preceding sections verify Chernov's axioms from Appendix A with the only exception of Condition A.3, that is, the ergodicity of the billiard map T (and its higher iterates). As we already mentioned, to complete the proof of Theorem 1.6, we refer to [BBT] where Condition A.3 is verified.

Now [BBT] proves (local) ergodicity based on a set of assumptions that are different and, literally, independent of the conditions in Appendix A. However, the assumptions (apart for some well-known regularity properties) in [BBT] follow from those of Appendix A with two exceptions. First, the setting of [BBT] is symmetric with respect to the roles of the stable and unstable direction, while Chernov's axioms are formulated only in terms of the unstable direction. Second, [BBT] requires some kind of proper alignment of unstable manifolds and *negative* time singularities. This does not coincide with the notion of Alignment mentioned in Section 2.4 and used in Section 4.

4.5.1 regularity properties and growth lemma for stable manifolds

The standard argument for checking that certain statements, already proven for the unstable direction and the forward dynamics, can be reformulated for the stable direction and the backward dynamics is to refer to the time reflection symmetry of the billiard map (see e.g.[CM]). This reasoning is correct if one measures distances with respect to the Euclidean metric, which is, indeed, symmetric with respect to time reflection – see Remark 2.5. Note, however, that the metric we use – the regularized Chernov-Dolgopyat metric of Definition 3.6 – no longer has this time reflection symmetry. Fortunately, the regularity properties of unstable manifolds – in particular conditions A.4 and A.5 from Appendix A – have all been verified for the Euclidean metric (see Section 2). Condition A.7 on the growth properties of unstable manifolds is the only property which is stated and proven in terms of the regularized Chernov-Dolgopyat metric and is not verified in terms of the Euclidean one.

The version of the growth lemma that appears among the assumptions of [BBT] (both in a stable and in an unstable form) is slightly weaker than Condition A.7; it is exactly the $\delta = 0$ version of (A.2). Recalling the definitons for the function $r_{V,k}$ and that actually $\hat{T} = T^{n_0}$, this statement reads as (see also remarks 4.4, 4.7 and Corollary 4.13):

$$m_W(\{x \in W | \rho(T^{n_0}x, \partial T^{n_0}W) < \varepsilon\}) \leq \alpha_0 \Lambda^{n_0} \cdot m_W(\rho(x, \partial W) < \varepsilon/\Lambda^{n_0}) + \varepsilon\beta_0\delta_0^{-1} m_W(W), \tag{4.37}$$

where $\alpha_0 \in (0, 1)$ and $\beta_0 > 0$ are some global constants.

To distinguish between the two metrics we will use ρ_E for the distances measured in the Euclidean metric and ρ_{CD} for the regularized Chernov-Dolgopyat metric. We claim that (4.37) with $\rho = \rho_{CD}$ implies the same statement with $\rho = \rho_E$, with slightly worse constants. To see this, first recall from Section 3 that the two metrics are equivalent. This implies

$$m_W(\{x \in W | \rho_E(T^n x, \partial T^n W) < \varepsilon\}) \leq m_W(\{x \in W | \rho_{CD}(T^n x, \partial T^n W) < C\varepsilon\}) \tag{4.38}$$

and

$$m_W(\rho_{CD}(x, \partial W) < C\varepsilon/\Lambda^n) \leq m_W(\rho_E(x, \partial W) < C^2\varepsilon/\Lambda^n). \tag{4.39}$$

But clearly $C^2\varepsilon/\Lambda^n \leq \varepsilon/\tilde{\Lambda}^n$ if $\tilde{\Lambda}$ is slightly less than Λ, and n is large enough. Thus (4.38), (4.39) and (4.37) for the regularized C-D metric directly imply (4.37) for the Euclidean metric with $\Lambda \to \tilde{\Lambda}$ and a (possibly) larger n_0. As mentioned above, this statement is directly transferable to the inverse map and the stable direction.

4.5.2 Alignment for negative time singularities

Let us introduce the set $\Gamma^- = \Gamma^0 \cap T\Gamma^0$. (Remember from Section 2.2.3 that Γ^0 is the boundary of our phase space after introducing homogeneity layers.) We think of Γ^- as the singularity set of the invesre dynamics T^{-1}, although it is important that it also contains the boundary of the phase space.

The notion of alignment required in [BBT] says roughly that unstable manifolds intersect Γ^- transversally, and their angle at any intersection point is at least some global constant c. However, this rough form of the assumption is not satisfied by our systems: it is known that even in 2 dimensions unstable manifolds may be tangent to $T\partial\tilde{M}$. Actually, for every component S of $T\partial\tilde{M}$, this happens exactly on one side of S (remember that S is a one-codimensional submanifold that cuts the phase-space into two). Indeed, this is the side of S that contains images (under T) of points near $\partial\tilde{M}$. The other side typically consists of images of points which were not near $\partial\tilde{M}$, but near $T^{-1}\partial\tilde{M}$, on the side which eventually avoided the nearly-tangent collision. See Figure 6 for an explanation: Ta is a phase point on $\partial\tilde{M}$. b is on the side of $T^{-1}\partial\tilde{M}$ which avoids the nearly tangent collision and travels directly to the neighbourhood of $T^2a \in T\partial\tilde{M}$, while c is first mapped near Ta, and only thereafter near to T^2a. The violation of the rough form of Alignment happens on the side of $T\partial\tilde{M}$ containing T^2c.

Now we can explain the refined form of the alignment assumption in [BBT]: we only expect an unstable manifold (developing in time) to be transversal to Γ^- at the time of their *first encounter*. The following proposition states this in a precise form. It implies

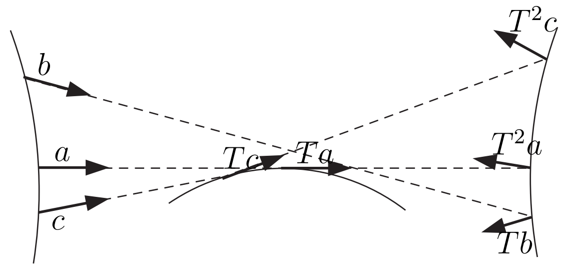

Figure 6: possible trajectories of points near a singularity

the Alignment assumption of [BBT]. (The only difference is that the assumption in [BBT] is formulated in terms of distances instead of angles, since no smoothness of these subsets is formally assumed there.)

Proposition 4.16. *There exists a global constant $c > 0$ with the following property: Let W be an unstable manifold contained entirely in a connected component of $M \setminus \Gamma^-$, and let $x \in \bar{W} \cap \Gamma^-$. If the inverse image of x under T as a one-sided limit,*

$$x^- := \lim_{y \in W,\ y \to x} T^{-1} y \tag{4.40}$$

is not *in Γ^-, then the angle of W and any smooth component S of Γ^- at x is at least c.*

Proof. If $S \subset \partial M = \Gamma^0 \cap \partial\tilde{M}$, then this transversality is known from [BChSzT1] and stated as (part of) the alignment property in Section 2.4.

If $S \subset T\Gamma^0_k$ for some $k \in \{k_0, k_1, \dots\}$ (that is, S is a secondary singularity of T^{-1}), then the extention of T^{-1} to S as a one-sided limit is the same for both sides, and the only possible inverse image x^- of x is in $\Gamma^0_k \subset \Gamma^-$, so there's nothing to prove (the conditions of the proposition cannot hold).

So the only interesting case is when $S \subset T\partial\tilde{M}$, and $x^- \notin \partial\tilde{M}$. With the notation of Figure 6 this corresponds to $x = T^2 a$, $x^- = a$, and the limit in (4.40) is through points $y = Tb$. Let us look at this case now.

In high dimensions (when $d \geq 3$) we have $\dim(W) + \dim(S) = (d-1) + (2d-3) > 2d - 2 = dim(M)$, so transversality of W and S can only mean that there is a $d-1$-dimensional subspace $U \subset \mathcal{T}_x S$ which is transversal to W in the sense that the angle between any $(dq_1, dw_1) \in U$ and any $(dq_2, dw_2) \in \mathcal{T}_x W$ is at least c. Now $d-1$-dimensional submanifolds can typically be considered as traces of fronts on the Poincaré phase space. (Remember from Section 2.3 that fronts are subsets of the flow phase space.) The (tangent space of the) unstable manifold W is known from Section 2.3.2 to correspond to a well-understood convex front described by (2.11) and (2.12), so it is enough to show an appropriate front whose trace is in S.

Please recall from Section 2.3.4 that if two fronts with s.f.f.-s B_1^- and B_2^- (just before collision) satisfy

$$B_2^- - B_1^- > c_1, \quad -C_1 < B_1^- < C_1, \tag{4.41}$$

then the angle of their traces in the Poincaré phase space is at least some $c > 0$ depending only on c_1 and C_1 (and the geometry of the billiard table).

Consider the point $T^{-1}x \in \partial\tilde{M}$, which is Ta in Figure 6. Let τ_1 denote the flight time from x^- to $T^{-1}x$, and τ_2 the flight time from $T^{-1}x$ to x.

The tangent space of $T^{-1}W$ at x^- (as a one-sided limit) is described by a convex front according to (2.11), so the pre-collision s.f.f. of its image at x satisfies

$$0 < B_1^- < \frac{1}{\tau_1 + \tau_2}. \tag{4.42}$$

On the other hand, consider the post-collision front with $B^+ = \infty$ at $T^{-1}x$ – or with other words, the vectors (dr, dv) with $dr = 0$. One can see that the (backward) traces of all these vectors are tangent to $\partial\tilde{M}$ at $T^{-1}x$, so we have found a front whose image at x defines a $d-1$-dimensional subspace of S. But the (pre-collision) image of this front at x is a sphere with

$$B_2^- = \frac{1}{\tau_2}. \tag{4.43}$$

Putting (4.42) and (4.43) together, we get

$$B_2^- - B_1^- \geq \frac{1}{\tau_1} - \frac{1}{\tau_2} \geq \frac{\tau_{min}}{2\tau_{max}^2},$$

so the conditions in (4.41) are satisfied with $c_1 = \frac{\tau_{min}}{2\tau_{max}^2}$, $C_1 = \frac{1}{2\tau_{min}}$. □

With these considerations, all the assumptions used in [BBT] are checked, and that paper gives the ergodicity of our billiards. So Condition A.3 is fulfilled, and the proof of Theorem 1.6 is complete.

Remark 4.17. *In an analogous way one could check Condition A.7 in full generality for the Euclidean metric. This means that the set of conditions from Appendix A can be verified for the Euclidean metric ρ_E. However, in our proof of the growth properties the use of ρ_{CD} is crucial (this is the only way we could reduce the statement with an inductive argument to the one step Lemma 4.3). Furthermore, our arguments throughout Section 4 use heavily the curvature and distortion bounds for the metric ρ_{CD}, which is the reason for the necessity of the differential geometric analysis of Section 3.*

Acknowledgements

We are grateful to Balázs Csikós for his help concerning the Riemannian geometry section. Many details of this proof were worked out in the Ervin Schrödinger Institute in Vienna during the authors' visit in January 2006. The stimulating atmosphere and the hospitality of the institute is thankfully acknowledged along with the support of the Austrian-Hungarian bilateral program. The research work of the authors is partially supported by the following OTKA (Hungarian National Research Fund) grants: F 60206 (for P.B.), T 46187 (for P.T.) and TS 49835 (for P.B. and P.T.); and by the Bolyai scholarship of the Hungarian Academy of Sciences (for P.B.).

Appendix

A Chernov axioms

Here we provide, for the reader's convenience, a very short, yet mainly self-contained formulation of Theorem 2.1 from [Ch2]. For self-containedness, many notions and notations are repeatedly introduced. First we give the conditions A.1 ... A.7 which are required, and then the statement of the theorem. We note that the theorem is applied with the substitutions $T \to \hat{T}$, $\Lambda \to \hat{\Lambda}$.

Condition A.1. ***The dynamical system*** *is a map $T : M \setminus \Gamma \to M$, where M is an open subset in a C^1 Riemannian manifold, $\bar{M}$ is compact. Γ is a closed subset in $\bar{M}$, and T is a C^2 diffeomorphism of its range onto its image. Γ is called the singularity set.*

We note that this condition is slightly different from that formulated in [Ch2]. There the Riemannian manifold was assumed to be C^∞. However, the proof of the theorem goes through without modification for a C^2 differentiable manifold with a C^1 Riemannian structure.

Condition A.2. ***Uniform hyperbolicity.*** *We assume there are two families of cone fields C^u_x and C^s_x in the tangent planes $\mathcal{T}_x M$, $x \in \bar{M}$ 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 $\bar{M}$, their axes have the same dimensions across the entire $\bar{M}$ which we denote by d_u and d_s, respectively;*
- *$d_u + d_s = \dim M$;*
- *the angles between C^u_x and C^s_x are uniformly bounded away from zero:*

$$\exists\ \alpha > 0 \text{ such that } \forall x \in M \text{ and for any } dw_1 \in C^u_x \text{ and } dw_2 \in C^s_x \text{ one has}$$

$$\sphericalangle(dw_1, dw_2) \geq \alpha$$

The C^u_x are called the unstable cones whereas C^s_x are called the stable ones.

The property that the angle between stable and unstable cones is uniformly bounded away from zero is called **transversality**.

Some notation and definitions. For any $\delta > 0$ denote by $\mathcal{U}_\delta$ the δ-neighbourhood of the closed set $\Gamma \cup \partial M$. We denote by ρ the Riemannian metric in M and by m the Lebesgue

measure (volume) in M. For any submanifold $W \subset M$ we denote by ρ_W the metric on W induced by the Riemannian metric in M, by m_W the Lebesgue measure on W generated by ρ_W and by $\mathrm{diam} W$, the diameter of W in the ρ_W metric.
LUM-s. To be able to formulate the further properties to be checked the reader is kindly reminded of the notion of local unstable manifolds. We call a ball-like submanifold $W^u \subset M$ a local unstable manifold (LUM) if (i) $\dim W^u = d_u$, (ii) T^{-n} is defined and smooth on W^u for all $n \geq 0$, (iii) $\forall x, y \in W^u$ we have $\rho(T^{-n}x, T^{-n}y) \to 0$ exponentially fast as $n \to \infty$.
We denote by $W^u(x)$ (or just $W(x)$) a local unstable manifold containing x. Similarly, local stable manifolds (LSM) are defined.

Condition A.3. ***SRB measure.*** *The dynamics T has to have an invariant ergodic Sinai-Ruelle-Bowen (SRB) measure μ. That is, there should be an ergodic probability measure μ on M such that for μ-a.e. $x \in M$ a LUM $W(x)$ exists, and the conditional measure on $W(x)$ induced by μ is absolutely continuous with respect to $m_{W(x)}$.*
Furthermore, the SRB-measure should have nice mixing properties: the system (T^n, μ) is ergodic for all finite $n \geq 0$.

In our case the SRB measure is simply the Liouville measure defined by (2.1). Absolute continuity and invariance of μ are straightforward, while ergodicity is proved in [BBT], based on the conditions mentioned in Section 4.5.

Condition A.4. ***Bounded curvature.*** *There should exist a global constant $K_W < \infty$ such that the curvature of any unstable manifold at any of its points is at most K_W.*

The meaning of the word "curvature" for the purpose of this condition is made precise in Section 3.3. Accordingly, the condition is formulated more precisely as Condition 3.10.
Some notation. Denote by $J^u(x) = |\det(DT|E^u_x)|$ the Jacobian of the map T restricted to $W(x)$ at x, i.e. the factor of the volume expansion on the LUM $W(x)$ at the point x.

Condition A.5. ***Distortion bounds****. Let x, y be in one connected component of $W \setminus \Gamma^{(n-1)}$, which we denote by V. Then*

$$\log \prod_{i=0}^{n-1} \frac{J^u(T^i x)}{J^u(T^i y)} \leq \varphi\left(\rho_{T^n V}(T^n x, T^n y)\right)$$

where $\varphi(\cdot)$ is some function, independent of W, such that $\varphi(s) \to 0$ as $s \to 0$.

Condition A.6. ***Absolute continuity****. Let W_1, W_2 be two sufficiently small LUM-s, such that any LSM W^s intersects each of W_1 and W_2 in at most one point. Let $W_1' = \{x \in W_1 : W^s(x) \cap W_2 \neq \emptyset\}$. Then we define a map $h : W_1' \to W_2$ by sliding along stable manifolds. This map is often called a holonomy map. This has to be absolutely* *continuous with respect to the Lebesgue measures m_{W_1} and m_{W_2}, and its Jacobian (at any density point of W_1') should be bounded, i.e.*

$$1/C' \leq \frac{m_{W_2}(h(W_1'))}{m_{W_1}(W_1')} \leq C'$$

with some $C' = C'(T) > 0$.

Some further notation. Let $\delta_0 > 0$. We call W a **δ_0-LUM** if it is a LUM and $\operatorname{diam} W \leq \delta_0$. For an open subset $V \subset W$ and $x \in V$ denote by $V(x)$ the connected component of V containing the point x.

Let $n \geq 0$. We call an open subset $V \subset W$ a (δ_0, n)-subset if $V \cap \hat{\Gamma}^{(n)} = \emptyset$ (i.e., the map $\hat{T}^n$ is smoothly defined on V) and $\operatorname{diam} \hat{T}^n V(x) \leq \delta_0$ for every $x \in V$. Note that $\hat{T}^n V$ is then a union of δ_0-LUM-s. Define a function $\mathbf{\hat{r}_{V,n}}$ on V by

$$\hat{r}_{V,n}(x) = \rho_{\hat{T}^n V(x)}(\hat{T}^n x, \partial \hat{T}^n V(x)) \tag{A.1}$$

Note that $\hat{r}_{V,n}(x)$ is the radius of the largest open ball in $\hat{T}^n V(x)$ centered at $\hat{T}^n x$. In particular, $\hat{r}_{W,0}(x) = \rho_W(x, \partial W)$.

Now we formulate Chernov's Growth Properties, in essentially (cf. Remark A.8) the same form as they appeared in [Ch1] and [Ch2].

Condition A.7. ***Growth of unstable manifolds.*** *Let us assume there is a fixed $\delta_0 > 0$, furthermore, there exist constants $\alpha \in (0,1)$ and $\beta, D, \kappa, \sigma, \zeta > 0$ with the following property. For any sufficiently small $\delta > 0$ and any δ_0-LUM W there is an open $(\delta_0, 0)$-subset V_δ^0 and an open $(\delta_0, 1)$-subset $V_\delta^1 \subset W \setminus \hat{\Gamma}^{[\delta]}$ (one of these may be empty) such that the two sets are disjoint, $m_W(W \setminus (V_\delta^0 \cup V_\delta^1)) = 0$ and $\forall \varepsilon > 0$*

First Growth Property:

$$m_W(\hat{r}_{V_\delta^1,1} < \varepsilon) \leq \alpha_0 \Lambda \cdot m_W(\hat{r}_{W,0} < \varepsilon/\Lambda) + \varepsilon \beta_0 \delta_0^{-1} m_W(W) \tag{A.2}$$

Second Growth Property:

$$m_W(\hat{r}_{V_\delta^0,0} < \varepsilon) \leq D_0 \delta^{-\kappa}\, m_W(r_{W,0} < \varepsilon) \tag{A.3}$$

and **Third Growth Property***:*

$$m_W(V_\delta^0) \leq D_0\, m_W(\hat{r}_{W,0} < \zeta \delta^\sigma) \tag{A.4}$$

Remark A.8. *Note that these growth properties are slightly different from those assumed in [Ch2], namely, there it was assumed that $V_\delta^0 \subset \hat{\Gamma}^{[\delta]}$. However, it is easy to see, that the whole strategy of [Ch2] works without this assumption. What is indeed important that the set V_δ^1 is disjoint both from V_δ^0 and $\hat{\Gamma}^{[\delta]}$, and that the measure of V_δ^0 can be estimated by the third growth lemma.*

How V_δ^0 and $\hat{\Gamma}^{[\delta]}$ are related is discussed in Section 4.2.1.

Now we can formulate Theorem 2.1 from [Ch2].

Theorem A.9. ***(Chernov, 1999)*** *Under the conditions A.1,...,A.7, the dynamical system enjoys exponential decay of correlations and the central limit theorem for Hölder-continuous functions.*

The properties stated in the theorem are defined in Definition 1.5 and Remark 1.7 (see also [Ch2]) .

B Geometric Lemmas

Here we prove the geometric sublemmas 4.10, 4.11 and 4.15. Please recall the relevant notation and the statement of the these sublemmas from sections 4.2.2 and 4.3.2. Sublemmas 4.10 and 4.11 will be easy corollaries of Sublemma B.1 below. We denote the distance in $\mathbb{R}^m$ by ρ.

Sublemma B.1. *For any $\varepsilon \geq 0$ and $0 \leq \xi \leq 1$,*

$$\mathrm{Leb}\left(\{x \in W_l \,|\, \rho(x,E) \leq \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \leq \varepsilon\}\right) \leq \xi \mathrm{Leb}(\{x \in W_r \,|\, \rho(x,\partial W) \leq \varepsilon\}).$$

The same is true with W_l and W_r interchanged.

Proof. Denote the two sets to compare by A and B, so

$$A = \{x \in W_l \,|\, \rho(x,E) \leq \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \leq \varepsilon\},$$

$$B = \{x \in W_r \,|\, \rho(x,\partial W) \leq \varepsilon\}.$$

At any point $z \in E$, denote the line orthogonal to E by e_z. Denote the Lebesgue measure on e_z by Leb_{e_z}, and the Lebesgue measure on E by Leb_E. Finally, let $A_z = A \cap e_z$, $B_z = B \cap e_z$. See Figure 7.

We can calculate the measure of A and B as

$$\mathrm{Leb}(A) = \int_E \mathrm{Leb}_{e_z}(A_z)\,\mathrm{dLeb}_E(z) \quad , \quad \mathrm{Leb}(B) = \int_E \mathrm{Leb}_{e_z}(B_z)\,\mathrm{dLeb}_E(z).$$

To get the statement of the lemma, it is clearly enough to see that

$$\mathrm{Leb}_{e_z}(A_z) \leq \xi \mathrm{Leb}_{e_z}(B_z) \text{ for any } z \in E. \tag{B.1}$$

To see this, let C_z be the interval of length $\xi\varepsilon$ in e_z which is just left of E. Clearly, $A_z \subset C_z$ and thus $\mathrm{Leb}_{e_z}(A_z) \leq \xi\varepsilon$.

If C_z is not entirely a subset of W, then either it is entirely outside W, or it contains a point of ∂W. In both cases, A_z is empty and (B.1) is trivial.

So, suppose $C_z \subset W$. Let v be the nearest point of $\partial W \cap e_z$ on the right of E, and let D_z be the interval of length ε in e_z just left of v. If $d := \rho(v,z) \geq \varepsilon$, then $D_z \subset B_z$,

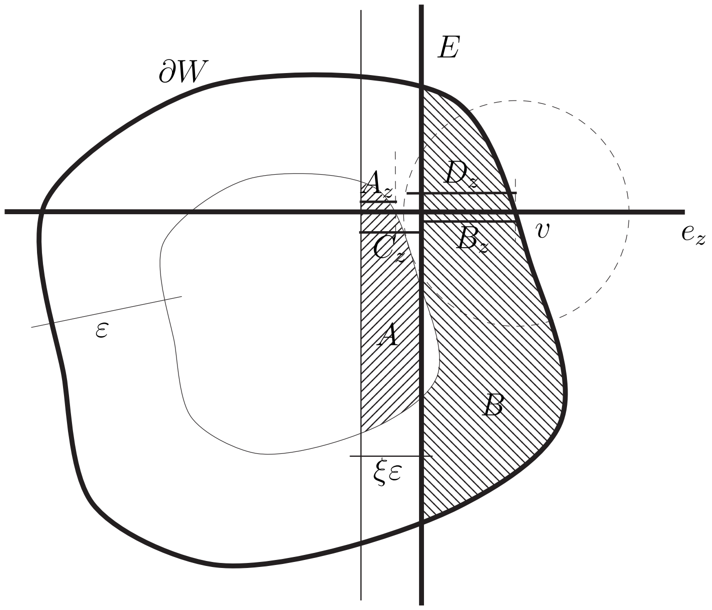

Figure 7: notation in the proof of Sublemma B.1

so $\mathrm{Leb}_{e_z}(B_z) \geq \varepsilon$ and (B.1) is trivial again. If not, then the intersection of C_z and D_z belongs to neither A_z nor B_z, so the estimate still holds:

$$\mathrm{Leb}_{e_z}(A_z) \leq \mathrm{Leb}_{e_z}(C_z \setminus D_z) \leq \max\{\xi\varepsilon - (\varepsilon - d), 0\} = \max\{\xi d - (1-\xi)(\varepsilon - d), 0\} \leq \xi d$$

and

$$\mathrm{Leb}_{e_z}(B_z) \geq \rho(v, z) = d$$

imply (B.1).

The statement with W_l and W_r interchanged is the same (with other notation). □

Proof. of Sublemma 4.10

The set on the left hand side can be decomposed as

$$\{x \in W_l \,|\, \rho(x, \partial W_l) \leq \varepsilon\} =$$
$$= \{x \in W_l \,|\, \rho(x, \partial W) \leq \varepsilon\} \;\cup (\{x \in W_l \,|\, \rho(x, E) \leq \varepsilon\} \setminus \{x \in W \,|\, \rho(x, \partial W) \leq \varepsilon\})\,.$$

The measure of the second term can be estimated using Sublemma B.1 with $\xi = 1$. We get

$$\mathrm{Leb}\{x \in W_l \,|\, \rho(x, \partial W_l) \leq \varepsilon\} \leq \mathrm{Leb}\{x \in W_l \,|\, \rho(x, \partial W) \leq \varepsilon\} + \mathrm{Leb}\{x \in W_r \,|\, \rho(x, \partial W) \leq \varepsilon\},$$

which is exactly what we need. The statement for W_r is the same. □

Proof. of Sublemma 4.11

The set on the left hand side can be decomposed into the parts to the left and right of E as

$$\begin{aligned} &\{x \in W \,|\, \rho(x,E) \le \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \le \varepsilon\} = \\ = &\left(\{x \in W_l \,|\, \rho(x,E) \le \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \le \varepsilon\}\right) \cup \\ \cup &\left(\{x \in W_r \,|\, \rho(x,E) \le \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \le \varepsilon\}\right). \end{aligned}$$

Both terms can be estimated using Sublemma B.1, and the result is

$$\begin{aligned} &\mathrm{Leb}\left(\{x \in W \,|\, \rho(x,E) \le \xi\varepsilon\} \setminus \{x \in W \,|\, \rho(x,\partial W) \le \varepsilon\}\right) \le \\ \le\ &\xi \mathrm{Leb}(\{x \in W_r \,|\, \rho(x,\partial W) < \varepsilon\}) + \xi \mathrm{Leb}(\{x \in W_l \,|\, \rho(x,\partial W) < \varepsilon\}) \end{aligned}$$

which is what we need. □

Proof. of Sublemma 4.15

We use the notation $H_\varepsilon = \{x \in W \,|\, \rho(x,\partial W) \le \varepsilon\}$, $H_{k\varepsilon} = \{x \in W \,|\, \rho(x,\partial W) \le k\varepsilon\}$, $V_\varepsilon = \mathrm{Leb}(H_\varepsilon)$, $V_{k\varepsilon} = \mathrm{Leb}(H_{k\varepsilon})$. We use the property of the Lebesgue measure that it can be obtained as the infimum of sums of volumes of spheres in a countable covering:

$$V_\varepsilon = \inf\left\{\sum_{i=1}^{\infty} \Gamma_m r_i^m \,|\, H_\varepsilon \subset \bigcup_{i=1}^{\infty} B_{r_i}(y_i), y_i \in \mathbb{R}^m, r_i \in \mathbb{R}^+\right\},$$

where $B_r(y)$ denotes the sphere of radius r centered at y, and Γ_m is the volume of the m-dimensional unit sphere. So for any $\delta > 0$ there exist $\{y_i\}_{i=1}^{\infty}$ and $\{r_i\}_{i=1}^{\infty}$ such that $H_\varepsilon \subset \bigcup_{i=1}^{\infty} B_{r_i}(y_i)$ and $\Gamma_m \sum_{i=1}^{\infty} r_i^m < V_\varepsilon + \delta$.

Now for every i, let x_i be one of the points of ∂W which is the closest to y_i. Such a point exists, since ∂W is compact, so the infimum defining $\rho(y_i, \partial W)$ is obtained. Define $z_i = x_i + k(y_i - x_i)$ and $B_i' = B_{kr_i}(z_i)$. That is, B_i' is obtained with the magnification of $B_{r_i}(y_i)$ with a factor k, but *with x_i as the center of the magnification.* See Figure 8 for the notation.

We claim that $H_{k\varepsilon} \subset \bigcup_{i=1}^{\infty} B_i'$. This immediately implies the statement of the lemma, since it means that

$$V_{k\varepsilon} \le \Gamma_m \sum_{i=1}^{\infty} (kr_i)^m = k^m \Gamma_m \sum_{i=1}^{\infty} (r_i)^m \le k^m (V_\varepsilon + \delta)$$

by the choice of $\{y_i\}_{i=1}^{\infty}$ and $\{r_i\}_{i=1}^{\infty}$, and this holds for every $\delta > 0$.

To see the claim, choose any point $c \in H_{k\varepsilon}$. Let a be one of the points of ∂W closest to c – again, such a point exists. Define $b = a + \frac{c-a}{k}$. We can see that $\rho(b,\partial W) = \rho(b,a)$, because the existence of a point $d \in \partial W$ with $\rho(b,d) < \rho(b,a)$ would imply $\rho(c,d) < \rho(c,a)$, which contradicts the choice of a.

Notice that $b \in H_\varepsilon$, because $\rho(b,\partial W) = \rho(b,a) = \frac{\rho(c,a)}{k} = \frac{\rho(c,\partial W)}{k} \le \frac{k\varepsilon}{k} = \varepsilon$ by the choice of c, and $b \in W$ because $c \in W$ and $\rho(c,\partial W) > \rho(c,b)$. This means that there is

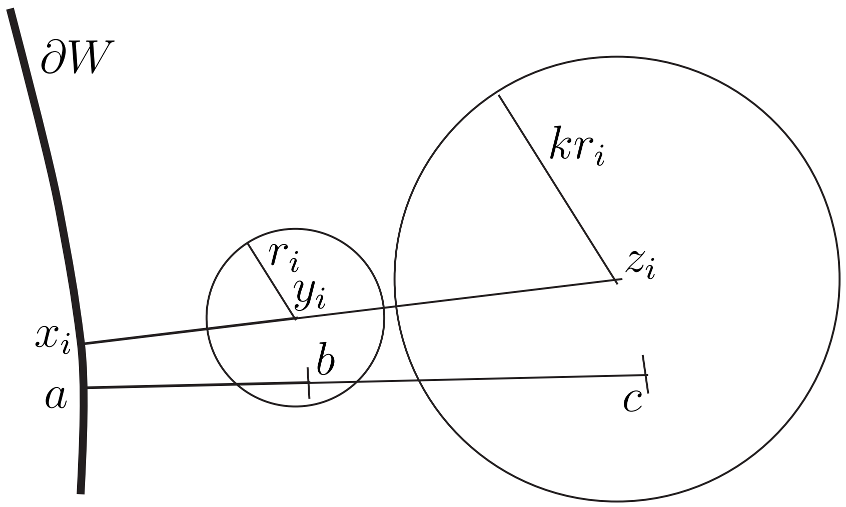

Figure 8: notation for the proof of Sublemma 4.15

an i for which $b \in B_{r_i}(y_i)$. We will show that for the same i, $c \in B_i'$, and this completes the proof.

To make the proof of $c \in B_i'$ transparent, we introduce the vectors $e = x_i - a$, $f = b - a$ and $g = y_i - x_i$. We will make use of the choice of x_i and a through the inequalities

$$\rho(y_i, a) \geq \rho(y_i, x_i) \quad ; \quad \rho(c, x_i) \geq \rho(c, a).$$

Our statement follows from

$$\rho(z_i, c) \leq k\rho(y_i, b).$$

With the vectors introduced, the conditions can be written as

$$|e + g| \geq |g| \quad ; \quad |kf - e| \geq |kf|$$

and the statement becomes

$$|e + kg - kf| \leq k|e + g - f|.$$

The conditions can be further rewritten as

$$e^2 + 2eg \geq 0 \quad ; \quad e^2 - 2kef \geq 0,$$

and the statement becomes (using $k > 1$)

$$ke^2 + e^2 + 2keg - 2kef \geq 0.$$

In this form, the statement is just the sum of k times the first condition and the second. $\square$

References

[AS] D. V. Anosov and Ya. G. Sinai *Some smooth ergodic systems*, Russ. Math. Surveys **22** (1967) 103–167.

[BBT] P. Bálint, P. Bachurin and I. P. Tóth, *Local Ergodicity of Systems with Growth Properties including Multi-dimensional Dispersing Billiards*, to appear in Israel Journal of Mathematics; preprint available at `http://www.renyi.hu/~bp/pub.html`

[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)

[B] P. Batchourine; *On the structure of the singularity manifolds of dispersing billiards* Arxiv preprint math.DS/0505620, 2005 - arxiv.org

[Ch1] N. Chernov; *Statistical Properties of Piecewise Smooth Hyperbolic Systems in High Dimensions*, Discrete and Continuous Dynamical Systems, **5**, 425–448, (1999).

[Ch2] N. Chernov, *Decay of correlations and dispersing billiards*, J. Statist. Phys. **94** (1999), 513–556.

[CD] N. Chernov and D. Dolgopyat, *Brownian Brownian motion I.* to appear in Memoirs AMS, available at
`http://www.math.umd.edu/~dmitry/brown11.pdf`

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

[CY] N. Chernov and L.S. Young, *Decay of correlations of Lorentz gases and hard balls*, in *Hard ball systems and the Lorentz gas*, ed. Domokos Szász, Springer, Berlin (2000) 89–120

[KSSz] A. Krámli, N. Simányi and D. Szász, *A "Transversal" Fundamental Theorem for Semi-Dispersing Billiards*, Commun. Math. Phys. **129** (1990), 535–560.

[KN] S. Kobayashi, K. Nomizu; *Foundations of differential geometry* Wiley 1963–1996.

[S] B. Saussol; *Absolutely continuous invariant measures for multidimensional expanding maps* Israel J. Math. **116** (2000), 223–248.

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

[Y] L.S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Math. **147** (1998), 585–650.