Example for Exponential Growth of Complexity in a Finite Horizon Multi-dimensional Dispersing Billiard

Péter Bálint
Mathematical Institute of the Technical University of Budapest
Egry József u. 1, H-1111 Budapest, Hungary
pet@math.bme.hu
Imre Péter Tóth
MTA-BME Stochastics Research group
Egry József u. 1, H-1111 Budapest, Hungary
mogy@math.bme.hu

September 13, 2011

Abstract

We present an example for a periodic point in a 3-dimensional dispersing billiard configuration around which the complexity of singularities grows exponentially with the iteration of the map. This implies the existence of multi-dimensional dispersing billiard configurations with finite horizon where the complexity grows exponentially. We also show that complexity growth can be faster than the minimum expansion of unstable vectors – a phenomenon with far-reaching consequences in the study of mixing properties for these systems. The observed behaviour is in strong contrast with the behaviour of 2-dimensional systems.

Keywords: hyperbolic systems with singularities, complexity of the singularity set, astigmatism, multi-dimensional dispersing billiards

1 Introduction

In the theory of dynamical systems with singularities, understanding the structure of the singularity set is of crucial importance. In this paper we consider a discrete time dynamical system with singularities, and address the question: "How many singularities of the n-th iterate can meet at a single point?" This number is called the *local complexity* of the n-step singularity set. We will make the definition precise in Definition 2.2. The dependence of this quantity on the number of iterates n, that is, the *local complexity growth of the singularity set* plays an important role when studying the ergodic and statistical properties of various kinds of hyperbolic systems with singularities. We will comment more on its importance in Section 2.2.

In the present paper we focus on the case of dispersing billiards with finite horizon and no corner points (see Section 2.3 below for a more detailed description). In the *two dimensional* case, according to a result by Bunimovich (see [BSC] or [Ch]) local complexity grows at most linearly in these systems. This property was exploited in Young's proof of exponential decay of correlations for planar dispersing billiard maps ([Y]). Young's result on the *exponential rate of mixing* was extended by the present authors to *multidimensional dispersing billiard maps* in [BaTo] *under the assumption of subexponential complexity*

growth. However, so far there has been no information available on the actual rate of complexity growth in the multidimensional case. As it usually happens, it turns out that multidimensional dispersing billiards are much more complex then their planar counterparts: in Theorem 2.7 we show that *in high (i.e. at least three) dimensions there exist finite horizon dispersing billiards for which the complexity of the singularity set grows exponentially.*

The rest of the paper is organized as follows. In Section 2 we collect some background material on hyperbolic systems with singularities in general (Section 2.1) and dispersing billiards in particular (Section 2.3). The main result of our paper, Theorem 2.7 is stated in Section 2.4, along with the explicit description of a 3-dimensional billiard configuration with exponentially growing complexity. Section 3 is devoted to 2-dimensional dispersing billiards. We introduce a one parameter family of planar finite horizon dispersing billiard configurations that, on the one hand, mimic some features of infinite horizon planar dispersing billiards, and, on the other hand, provide the building bricks for our multidimensional example with exponentially growing complexity. Finally in Section 4 we summarize some material on multidimensional dispersing billiards and prove Theorem 2.7. Some arguments of more calculatory character are presented in the Appendix.

2 Setting and statement of results

2.1 Hyperbolic dynamical systems with singularities

There are slight variations in the literature on what we mean by a hyperbolic dynamical system with singularities. The setting we describe below is neither the most possible general, nor it is complete in the sense that further assumptions are needed (depending on the studied example) to deduce reasonable statements on the ergodic properties of the system. However, we think that our observations are relevant for many interesting examples.

The two main features of the systems we consider are hyperbolicity and the presence of singularities.

The description of the **singularity structure** is central to our exposition, hence here we elaborate a little more on the details. We stick to discrete time maps. The phase space M of our dynamical system $T : M \to M$ is a (finite union of) D dimensional compact Riemannian manifold(s) (possibly with boundary). The domains of continuity for T are finitely many, pairwise disjoint open sets $O_k \subset M$, $k = 1, \dots, K$, such that

(i) Their closures cover the phase space ($\cup_{k=1}^{K} \overline{O_k} = M$).

(ii) The boundary ∂O_k of any fixed piece is a finite union of smooth 1-codimensional compact submanifolds (possibly with boundary) in M.

(iii) Restricted to any fixed open piece, $T_k = T|_{O_k}$ is a C^2 smooth one-to-one map onto its image. Furthermore, T_k has a continuous extension to the closed piece $\overline{O_k}$. (However, its derivatives may have no finite limit at the boundary ∂O_k.)

(iv) For $x \in \overline{O_k} \cap \overline{O_l} (= \partial O_k \cap \partial O_l)$, $k \neq l$, Tx is defined to coincide with one of the possible continuous extensions, i.e. – somewhat sloppily – as $T_k x$ or $T_l x$.

Remark 2.1. *Note that the set $\bigcup_{k \neq l} \partial O_k \cap \partial O_l$ is a finite union of smooth 1 codimensional submanifolds, in particular, it is of zero Lebesgue measure. Hence, from the viewpoint of ergodic properties, it is irrelevant how exactly the dynamics is defined on that set. Condition (iv) above is only included to avoid ambiguities.*

The set of phase points where T is possibly non-continuous, $\cup_{k=1}^{K} \partial O_k$ is called the *singularity set* of T. To understand the asymptotic properties of the dynamics, the singularities of all iterates T^n, $n \in \mathbb{Z}^+$ must be taken into account: our aim is to "count singularities" that meet at a single point. However, the framework we are using is not really suitable for counting singularities, because the whole singularity set is treated as one, components are not indexed. Instead, we will *count the domains of continuity*. This can

be done precisely, and then one can loosely say that $m-1$ singularity components, that meet at a point, cut the phase space locally into m pieces.

For this reason we introduce some notation: for $n \geq 1$, and $\mathbf{i} = (i_1, i_1, \ldots, i_n) \in \{1, \ldots, K\}^n$ we let $T^n_{\mathbf{i}} = T_{i_n} \circ \cdots \circ T_{i_1}$, which is defined on $O_{\mathbf{i}}$, where $O_{(i_1)} = O_{i_1}$, and

$$O_{\mathbf{i}} = O_{(i_1,\ldots,i_n)} = \{q \in O_{i_1} \text{ such that } T_{i_1}(q) \in O_{(i_2,\ldots,i_n)}\}\,. \tag{2.1}$$

Note that some of the $O_{\mathbf{i}}$ may be empty. In fact, our aim is exactly to study how fast the possible combinations for nonempty $O_{\mathbf{i}}$ grow.

Definition 2.2. *For every $x \in M$, the n-step complexity index of x, $m_n(x)$ is the number of possible distinct $\mathbf{i} \in \{1, \ldots, K\}^n$ such that $x \in \overline{O_{\mathbf{i}}}$. The n-step complexity of the singularity set, m_n is the supremum of $m_n(x)$ for all $x \in M$.*

In contrast to the description of the singularity structure, it has little significance for the present paper how exactly the **hyperbolicity** of the dynamics is formulated. The only important feature is that hyperbolicity is uniform, that is, there is a constant $\Lambda > 1$ such that unstable perturbations are expanded – and stable perturbation are contracted – at least by a factor Λ with each iteration of the map. This can be expressed, for example, in terms of cones, as done in the next paragraph, which can be skipped at a first reading of the paper.

There exist two transversal continuous cone fields in the tangent bundle of M, the d_s-dimensional stable cone field $C^s(x)$ and the $d_u(= D - d_s)$ dimensional unstable cone field $C^u(x)$, $x \in M$, which satisfy the following properties. Consider $x \in O_k$ for some k (so that the derivative DT_x is well-defined), then $DT_x(C^u(x)) \subset C^u(Tx)$ and $(DT_x)^{-1}C^s(Tx) \subset C^s(x)$. Furthermore, hyperbolicity is uniform: there exists some constant $\Lambda > 1$ such that for any x, and any vector $v \in C^u(x)$ we have $|DT_x v| > \Lambda|v|$, while for any $w \in C^s(Tx)$ we have $|(DT_x)^{-1}w| > \Lambda|w|$. The constant $\Lambda > 1$ (which is uniform in the sense that it depends only on the dynamical system itself) and its relation to the combinatorial properties of the singularities plays an important role in the studies of the ergodic behaviour of the system.

Dynamical systems with the above discussed properties have been in the focus of investigation for several decades. Main questions of interest are the existence of at most finitely many ergodic SRB measures along with their statistical properties, such as rates of mixing and probabilistic limit laws. Various methods, which may study the system directly or apply some coding, and may use various sophisticated tools such as modern functional analysis or probabilistic coupling, have been implemented to study these problems. Not aiming for completeness, we mention [Y], [DL], [CD], [CZ], [BG1], [BG2] as some of the most recent and best known examples. These works may have various purposes in mind and correspondingly, the results obtained also have a different range of applicability. However, it is a common feature in all of them that, to prove strong statistical properties for the SRB-measure(s), one has to ensure that "hyperbolicity dominates singularities" in an appropriate sense.

The most convenient case is when the growth of m_n is subexponential:

Definition 2.3. *The singularity set has* sub-exponential complexity growth *if $m_n = o(\lambda^n)$ for every $\lambda > 1$.*

However, when one studies statistical behaviour in *uniformly hyperbolic* systems, a somewhat weaker property is usually enough:

Definition 2.4. *The singularity set has* sub-expansion complexity growth *– or, in other words,* hyperbolicity dominates singularities *if*

$$m_n = \mathcal{O}(\lambda^n) \text{ for some } \lambda < \Lambda, \tag{2.2}$$

where Λ is the factor of least expansion for unstable vectors.

Since it is typically hard to get explicit lower bounds on Λ, sub-expansion complexity growth is often proven as a trivial consequence of sub-exponential growth.

2.2 The role of complexity growth

To indicate the significance of (2.2), let us mention that we are unaware of any proof of strong statistical properties (such as exponential decay of correlations) for a truly hyperbolic system (i.e. when there are both stable and unstable directions) with singularities, that does not use this condition (or some essentially equivalent version of it). It is an interesting open question whether there exist models at all with an exponentially mixing unique SRB measure for which, however, condition (2.2) fails.

On the other hand, in the example we are presenting – where complexity growth is too fast – it may still happen that correlation decay will once turn out to be exponential – although all presently existing methods of proof break down. This is a very important open question and a highly interesting specific system for further detailed study.

At this point it may be worth commenting briefly on piecewise expanding maps, i.e. when every direction is unstable. Here the collection of results is much larger, see eg. [BG1] for a list of references. There is, however, a big difference between piecewise expanding *interval* maps ($D = 1$) and *higher dimensinal systems*. In the former, the singularity set consists of finitely many points, hence, as already observed in [LY], practically no additional assumptions are needed to prove strong statistical properties. On the contrary, in the multidimensional case (if $D \geq 2$) interesting counterexamples by Tsujii ([T]) and Buzzi ([Bu]) indicate that some assumptions in the spirit of (2.2) are necessary. Without any assumptions, not even the existence of an absolutely continuous invariant measure (which plays the role of SRB measures in the context of expanding maps) can be ensured.

2.3 Multidimensional dispersing billiards

In this paper we consider one of the most popular class of hyperbolic systems with singularities, dispersing billiards with finite horizon (or Sinai billiards). For this particular class of dispersing billiards, the natural absolutely continuous invariant measure – the Liouville-measure – is known to be ergodic and mixing (under the natural condition that the scatterers are defined by algebraic equations, see [BChSzT2]). However, exponential decay of correlations was only proven under the additional assumption of sub-expansion complexity growth (cf. [BaTo]). It is natural to ask whether particular billiard configurations have this property or not. Here we present an example of a multidimensional dispersing billiard configuration with finite horizon for which complexity is shown to grow exponentially, and actually, (2.2) also fails. Hence the rate of mixing for the dynamics in our particular billiard configuration is unknown.[1]

Before describing our counterexample we collect some material on the class of dispersing billiards studied in this paper. Let us consider a connected 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 make some further assumptions.

The *boundary* $\partial\mathbb{Q}$ is assumed to be a finite collection of *pairwise disjoint*, compact $d-1$ dimensional C^3*-smooth* submanifolds in $\mathbb{T}^d$: $\partial\mathbb{Q} = \cup_{k=1}^{K} \partial\mathbb{Q}_k$. Furthermore, the billiard is *strictly dispersing*. That is, the boundary components, as viewed from the exteriour, are strictly convex.

In continuous time, the dynamics can be described as a flow $\Phi^t : \mathcal{M} \to \mathcal{M}$, where $\mathcal{M} = \mathbb{Q} \times \mathbb{S}^{d-1}$. Here $\mathbb{S}^{d-1}$ is the $d-1$ dimensional unit sphere, and for $q \in \partial\mathbb{Q}$, phase points with incoming ($\langle n(q), v\rangle \leq 0$) and outgoing ($\langle n(q), v\rangle \geq 0$)velocities are identified (here $n(q)$ is the unit normal vector of $\partial\mathbb{Q}$ at q). Nonetheless there is a natural Poincaré section for this flow corresponding to collision moments. Consider $q \in \partial\mathbb{Q}$ (a configuration point) and $v \in \mathbb{S}^{d-1}$ with $\langle n(q), v\rangle \geq 0$ (an outgoing velocity). The pair $\mathbf{w} = (q, v)$ is a phase point in our dynamical system for which we may consider the *free flight function*: $\tau(\mathbf{w})$ measures the distance along the straight line that starts out of q in the direction of v until it reaches $\partial\mathbb{Q}$ again. 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 $\mathbf{w} = (q, v)$ we have $\tau(\mathbf{w}) \leq \tau_{max}$. At this point it is worth recalling

[1] We remark that the opposite direction is still completely open: it is unknown whether there exists a multi- (not 2) dimensional configuration with sub-expansion complexity growth.

that the smooth components of the boundary (the scatterers) are assumed to be disjoint, hence *there are no corner points*. 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 $\mathbf{w} = (q, v)$ we also have $\tau(\mathbf{w}) \geq \tau_{\min}$.

In a summary, we consider dispersing billiards with finite horizon and no corner points. Let us describe briefly why these models belong to the above defined class of hyperbolic systems with singularities. We investigate the dynamics in discrete time, that is, we work with the Poincaré section (or collision) phase space:

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

where, as already mentioned above, $n(q)$ denotes the (unit) normal vector of $\partial\mathbb{Q}$ at q pointing inward $\mathbb{Q}$. It can be shown that M is indeed a finite collection of $D = 2d - 2$ dimensional Riemannian manifolds (actually, semi-sphere bundles, see [BChSzT1]) corresponding to the smooth components of the boundary, in other words, to the scatterers. Hence, roughly: $M = \cup_{k=1}^{K} M_k = \cup_{k=1}^{K} \partial\mathbb{Q}_k \times S_+^{d-1}$, where S_+^{d-1} is the semisphere of outgoing velocities. The boundary $\partial M = \cup_{k=1}^{K} \partial M_k$ of the phase space corresponds to tangential collisions ($\langle v, n(q) \rangle = 0$) which determine the singularities of the system (see below). We will consider the component of the phase space corresponding to a fixed scatterer, say M_1, other components can be treated similarly.

The dynamics $T : M \to M$ determines the state of the particle at the next collision as a function of the present state. Hyperbolicity of T is related to the strict convexity of the boundary, on details of this mechanism see eg. [CM] or [BaTo]. Some important features of the hyperbolicity in dispersing billiards, in particular, phenomena relevant to the present paper, will be discussed in course of our exposition in sections 3 and 4.

Concerning singularities, it is easy to see that the dynamics is discontinuous at preimages of tangential collisions: if $\mathbf{w} \in M_1$ such that $T\mathbf{w} \in \partial M_i$ for some $i \in \{1, \ldots, K\}$, then there are $\mathbf{w}_1$ and $\mathbf{w}_2$ arbitrary close to $\mathbf{w}$ such that $T\mathbf{w}_1 \in \mathrm{int}M_i$ while $T\mathbf{w}_2 \in \mathrm{int}M_j$ for some $j \neq i$. The lack of corner points guarantees that preimages of tangential collisions are the only singularities.

Remark 2.5. *The billiard configuration on $\mathbb{T}^d$ lifts to a $\mathbb{Z}^d$-periodic billiard configuration on $\mathbb{R}^d$. Actually, it is the later that is relevant for the singularity structure of T. However, as long as the horizon is finite, it makes no essential difference to consider the configuration on the torus. Indeed, the scatterers of the infinite periodic configuration can be indexed by $(i, \mathbf{k})$, where $i \in \{1, \ldots, K\}$ and $\mathbf{k} = (k_1, ..., k_d) \in \mathbb{Z}^d$. Any $\mathbf{w} \in M_1$ lifts to some $\tilde{\mathbf{w}} \in M_{(1,\mathbf{0})}$ (where $\mathbf{0} = (0, ..., 0)$ is chosen to avoid ambiguity). Finiteness of the horizon implies that there are only finitely many scatterers $M_{(i,\mathbf{k})}$ visible from $M_{(1,\mathbf{0})}$. Hence, by choosing the fundamental cell of the periodic configuration larger – and, consequently, increasing K, the number of distinct scatterers contained in configuration on $\mathbb{T}^d$ – it can be ensured that for any $i \in \{1, \ldots, K\}$ there is a unique $\mathbf{k}_i \in \mathbb{Z}^d$ such that any $\mathbf{w} \in M_1$ with $T\mathbf{w} \in M_i$ lifts uniquely to $\tilde{\mathbf{w}} \in M_{(1,\mathbf{0})}$ and $T\tilde{\mathbf{w}} \in M_{(i,\mathbf{k}_i)}$. (Actually, the entries of any such $\mathbf{k}_i$ are 0 or ± 1.)*

Hence we have open sets $O_i \subset M_1$, $i = 1, \ldots, K$ defined as $O_i = \{\mathbf{w} \in M_1 \mid T\mathbf{w} \in \mathrm{int}M_i\}$. It is known (see eg. [CM], or [BChSzT1] and [BaTo] specifically on the multidimensional case) that the open pieces O_i (restricted onto which the dynamics is continuous) fit the assumptions for hyperbolic systems with singularities, cf. the description in Section 2.1.

To address the issue of complexity, we consider a phase point $\mathbf{w} \in M_1$ such that none of the first n collisions is tangential on the trajectory of $\mathbf{w}$. Then the *symbolic collision sequence* of length n for $\mathbf{w}$ is the sequence of symbols $\mathbf{i}_n(\mathbf{w}) = (i_1, \ldots, i_n) \in \{1, \ldots, K\}^n$ that lists the indices of the consecutive scatterers along the trajectory of $\mathbf{w}$, that is, for which $T^l\mathbf{w} \in \mathrm{int}M_{i_l}$, $l = 1, \ldots, n$. The open pieces restricted onto which the n-th iterate of the dynamics is continuous can be identified with symbolic collision sequences. That is, for $\mathbf{j}_n = (j_1, \ldots, j_n) \in \{1, \ldots, K\}^n$ we have $O_{\mathbf{j}_n} = \{\mathbf{w} \in M_1 \mid \mathbf{i}_n(\mathbf{w}) = \mathbf{j}_n\}$.

The above observations can be summarized as follows:

Lemma 2.6. *For every $\mathbf{w} \in M$, the complexity $m_n(\mathbf{w})$ equals the number of distinct symbolic collision sequences $\mathbf{i}_n$ that can be realized by trajectories starting out of phase points arbitrary close to $\mathbf{w}$.*

2.4 Description of the counterexample

Now we are in the position that we can state the main result of our paper.

Theorem 2.7. *For every $d \geq 3$ there exists a d-dimensional dispersing billiard configuration with finite horizon and no corner points that has exponentially growing complexity. Actually, any multidimensional finite horizon dispersing billiard can be modified, by adding finitely many scatterers, in such a way that the new configuration has exponentially growing complexity.*

Actually, our result is not merely an existence statement, we can explicitly construct a billiard configuration to which Theorem 2.7 applies. For simplicity, we restrict to the three dimensional case, generalization to higher dimensions is automatic.

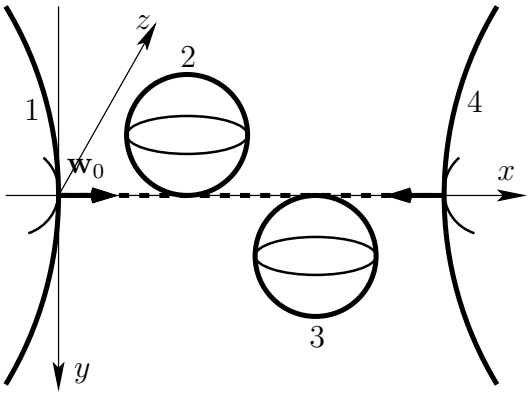


Figure 1: 3D configuration with exponential complexity

The configuration (see Figure 1) consists of four spherical scatterers $1, 2, 3, 4$, and a periodic orbit of a phase point $\mathbf{w}_0$ that bounces back and force between the two "big" scatterers 1 and 4. Between two consecutive perpendicular bounces on the big scatterers, the orbit will collide tangentially on both of the two "small" scatterers 2 and 3. The configuration has the following parameters:

– r is the radius of the "small" scatterers,

– R is the radius of the big scatterers,

– L_1 is the length of the free flight (along the trajectory of $\mathbf{w}_0$) between the regular collision on 1 and the tangential collision on 2,

– L_2, the length of the free flight (along the trajectory of $\mathbf{w}_0$) between the regular collision on 4 and the tangential collision on 3,

– and finally D, the length of the free flight (along the trajectory of $\mathbf{w}_0$) between the tangential collisions on 2 and 3.

Note that $L = L_1 + D + L_2$ is the distance of scatterers 1 and 4, and thus, it is (half of) the length of the periodic trajectory of $\mathbf{w}_0$. We will assume, only for the convenience of notation, that $L_1, L_2, D > r$, which ensures that all the spheres are disjoint.

Specifically

– scatterers 1 and 4 are spheres of radius R in $\mathbb{R}^3$, centered around the points $(-R, 0, 0)$ and $(L + R, 0, 0)$, respectively;

– scatterers 2 and 3 are spheres of radius r in $\mathbb{R}^3$, centered around the points $(L_1, r, 0)$ and $(L_1 + D, -r, 0)$, respectively;

– the point $\mathbf{w}_0 = (q_0, v_0)$ (the phase point from which the singular periodic orbit originates) has configurational coordinate $q_0 = (0, 0, 0)$ and velocity $v_0 = (1, 0, 0)$.

For our demonstration of exponential complexity to work, we will need to assume a further relation between the geometric parameters (see Theorem 2.8) meaning essentially that R is big enough compared to the other parameters.

Our example could be generalized in several ways as long as the following crucial features of the configuration are maintained:

(i) the periodic trajectory collides on the spheres 1 and 4 perpendicularly,

(ii) it touches scatterers 2 and 3 tangentially between two such perpendicular bounces,

(iii) there is a plane containing the trajectory such that scatterers 2 and 3 are on the opposite sides of that plane.

Now we can give an explicit description of the set of collision sequences that can be realized arbitrary close to the orbit of $\mathbf{w}_0$, which we will denote by Σ_n. Let

$$A_{1,2} = \{(1, 2), (2, 1)\} \qquad \text{and} \qquad A_{3,4} = \{(3, 4), (4, 3)\}.$$

Then

$$\Sigma_n = \{(2, w_1, w_2, \ldots, w_{n-1}) \,|\, w_i \in A_{3,4} \text{ for odd } i, \text{ and } w_i \in A_{1,2} \text{ for even } i\}.$$

In words, an (almost) tangential collision on scatterer 2 should be followed by an (almost) tangential collision on scatterer 3. However, this collision on scatterer 3 may take place either before or after the perpendicular bounce on scatterer 4. Similarly, an (almost) tangential collision on scatterer 3 should be followed by an (almost) tangential collision on scatterer 2. However, this collision on scatterer 2 may take place either before or after the perpendicular bounce on scatterer 1.

Theorem 2.8. *Consider the above 3-dimensional billiard configuration, the phase point $\mathbf{w}_0$ and the set Σ_n of symbolic collision sequences. Assume also that $L_1 = L_2$ and $R > 169\frac{L^3}{(2L_1 - r)^2}$. Then for any $n \geq 1$ and any $\delta > 0$ there exists some point $\mathbf{w} \in B_\delta(\mathbf{w}_0)$ (in the δ-neighbourhood of $\mathbf{w}_0$) with symbolic collision sequence $\mathbf{i}_{2n-1}(\mathbf{w}) = \mathbf{j}$. Thus $m_{2n-1}(\mathbf{w}_0) \geq |\Sigma_n| = 2^{n-1}$.*

Remark 2.9. *In course of our proof of Theorem 2.8, given $n \geq 1$ we will first choose a constant $\delta_0(n) > 0$ such that in the δ_0-neighbourhood of $\mathbf{w}_0$ we can give a description for a rich class of trajectories up to the n-th collision. Then for any $\delta > 0$ we let $\delta_1 = \min(\delta_0, \delta)$ and construct $\mathbf{w}$ with symbolic sequence $\mathbf{j} \in \Sigma_n$ in the δ_1-neighbourhood of $\mathbf{w}_0$.*

The rest of our paper is devoted to the proof of Theorem 2.8. Theorem 2.7, and the following two statements, are easy consequences, which we prove at the end of Section 4 and in the Appendix, respectively.

To get an example where this exponential growth of complexity is surely not sub-expansion, we need an observation about the unstable expansion in our example:

Lemma 2.10. *The minimum unstable expansion factor Λ in the above billiard configuration can be made arbitrarily close to 1 by choosing the geometric parameters appropriately, that is, by choosing R big enough compared to L.*

Corollary 2.11. *The above 3-dimensional billiard configuration, if R is big enough compared to L, does not have the sub-expansion complexity growth property.*

3 Planar dispersing billiards

To understand our three-dimensional example easier, it is useful to first look at certain two-dimensional dispersing billiard configurations. We will see how the iterates of the billiard map can have exponentially many singularity manifolds – or, with other words, exponentially many possible symbolic collision sequences – "near" a reference phase point. To avoid confusion, it is important to mention that (according to a result by Bunimovich [BSC], see also [Ch]) in a 2D finite horizon dispersing billiard with no corner points, complexity grows at most linearly. Hence this "near" cannot be exactly zero distance, as that would mean exponential complexity for a 2D system. However, the mechanism of the growth in the number of nearby singularities is the same as it will be in our true three-dimensional example.

3.1 Infinite horizon

For the first step, consider a periodic configuration of circular scatterers on the (infinite) two-dimensional plane. The interesting part is the so-called *corridor* near a flow phase point with infinite free flight. This is shown in Figure 2.

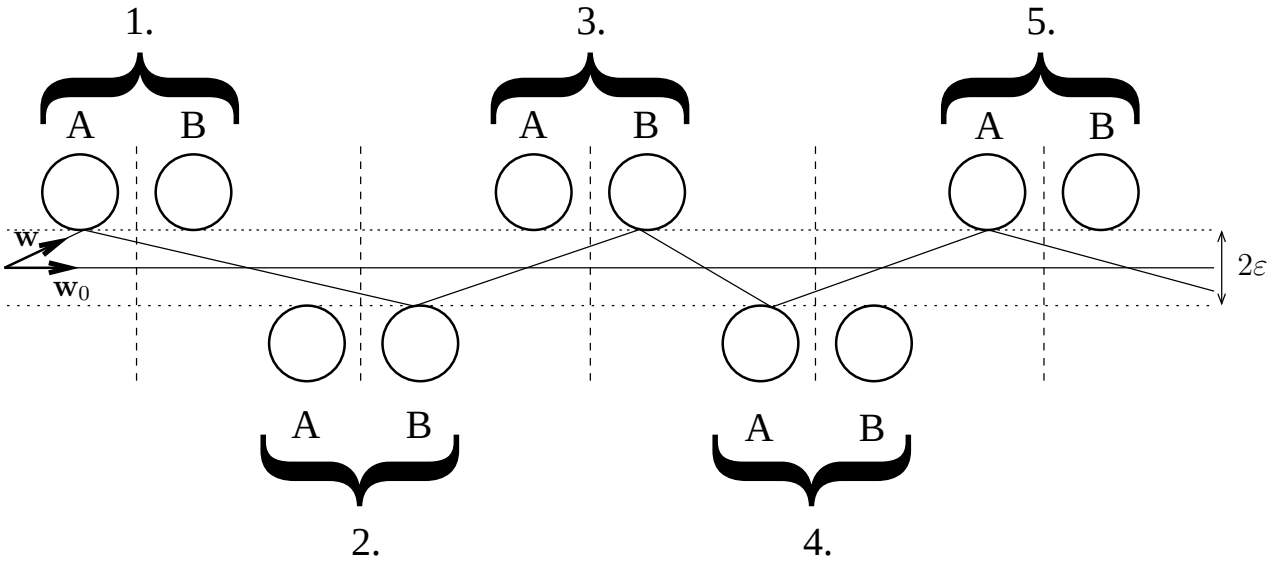


Figure 2: The corridor in a 2D infinite horizon billiard

The figure shows a reference phase point $\mathbf{w}_0$ with an infinite free flight, and a "nearby" phase point $\mathbf{w}$ with a pretty arbitrary collision history. Actually, if the geometric parameters are chosen properly, then $\mathbf{w}$ is indeed quite close to $\mathbf{w}_0$. In particular, their distance is of order ε, where 2ε denotes the "width" of the corridor, that is, the distance of the two (dotted) lines tangent to the infinitely many scatterers (see the figure). We will use the phrase "optical axis" for the (solid) line in the middle of the corridor which is the (configuration trace of the) trajectory of $\mathbf{w}_0$. We will also use the phrase "horizontal" for the direction of the optical axis, and the words "below", "above", "left" and "right" accordingly.

We will show in a proposition that if ε is small enough (w.r.t the other geometric parameters), then indeed many symbolic collision sequences are possible. To state the proposition, let us introduce the following notations: r is the radius of the (identical) scatterers, $2L$ is the horizontal periodicity of our configuration (hence L is the distance between the midpoints of two consecutive *pairs* of scatterers, located on the two opposite sides of the corridor) and $d(< L - 4r)$ is the (horizontal) distance between the two scatterers "A" and "B" within a pair.

Proposition 3.1. *For any choice of r, L and d there is some ε_0 such that whenever the width of the corridor is $\varepsilon < \varepsilon_0$, the following property holds. For any finite sequence $\sigma = (\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n)$ of "A"-s and "B"-s it is possible to choose some $\mathbf{w}$ in the $C\varepsilon$-neighbourhood of $\mathbf{w}_0$ such that the n-long symbolic collision sequence of $\mathbf{w}$ is equal to σ. In other words, it is possible to hit exactly scatterer σ_i from each*

couple i as shown in Figure 2. So the number of possibilities grows exponentially with the length of the piece of the corridor considered. Here C is a numerical constant that could be determined explicitly, its value is, however, not important.

The proof of this proposition is based on a lemma about the evolution of fronts in the corridor. To this end, let us first recall (cf. [CM], [BChSzT1]) that in billiard theory a *front* is a one codimensional smooth, compact and oriented submanifold $W \subset \mathbb{Q}$ which, when supplied with its unit normal vector field (the velocities), can be regarded as a submanifold of the billiard flow phase space. In the two dimensional case, W is simply a curve in $\mathbb{Q}$ (supplied with its normal vector, i.e. the velocity, in each of its points). The class of fronts is preserved by the billiard flow. To state the lemma, first we introduce an appropriate subclass by defining the temporary notion of a "good front in the corridor". Roughly speaking, this means

– a front which contains many phase points traveling along the corridor (thus having nearly horizontal velocities);

– the directions of the velocities at the endpoints are such that the phase points in between are guaranteed to collide at many different points of the next *two* scatterers;

– and the front is far enough from the next scatterer so that these collisions are nearly tangential.

We will use the notation introduced above, see also Figure 3. K denotes a large constant that depends on r, L and d, and by "nearly horizontal velocity" we mean a unit vector with vertical component at most $K^2\varepsilon$, see Lemma 3.3 and its proof.

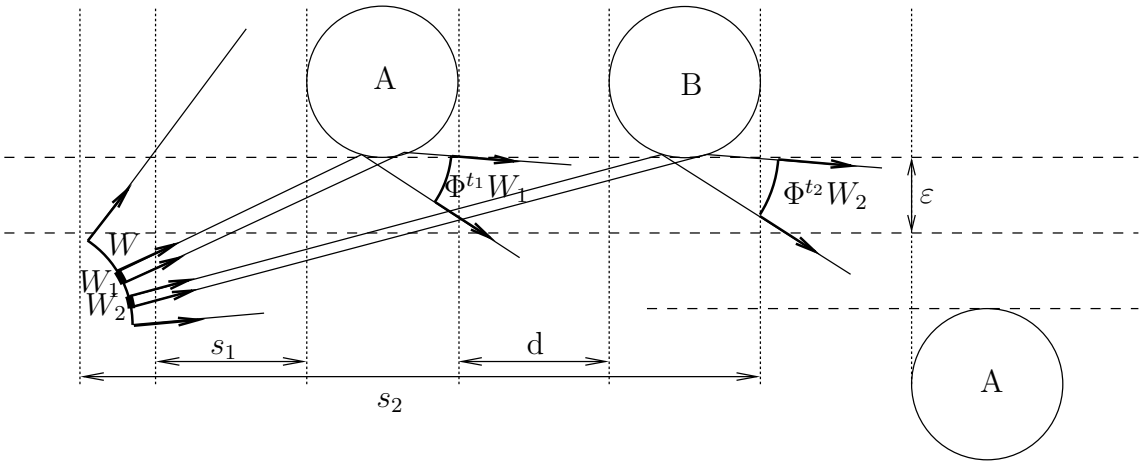


Figure 3: Evolution of a good front in the corridor

Definition 3.2. *A front is called a* good front in the corridor, tending up, *if*

- *it consists of phase points with nearly horizontal velocities flying to the right,*
- *it is entirely below the optical axis,*
- *the distance of any of its points from the optical axis is at most 2ε,*
- *the tangent of the velocity at its bottom endpoint is $\leq \varepsilon/K$,*
- *the tangent of the velocity at its top endpoint is $\geq K\varepsilon$,*
- *the next two scatterers along the corridor approached by the front (denoted by A and B) are at the top of the corridor,*
- *the horizontal distance of any of its points from any point of A or B is between $L/2$ and $3L$. (With the notation of Figure 3: $L/2 \leq s_1, s_2 \leq 3L$.)*

A good front in the corridor, tending down *is defined analogously. It is the mirror image (reflected over the optical axis) of a good front tending up, shifted horizontally with L so that the next two scatterers approached are at the bottom of the corridor.*

For better transparency of the arguments we can assume (although this is not needed) that a good front is convex.

Lemma 3.3. *For any $K \in \mathbb{R}^+$ large enough there exists an $\varepsilon_0 > 0$ such that for every $\varepsilon < \varepsilon_0$ the following holds: If W is a good front in the corridor tending up, then W contains sub-fronts $W_1, W_2 \subset W$ and there are times t_1, t_2 with $L/2 \le t_1, t_2 \le 2L$ such that*

- *W_1 and W_2 only intersect at their common endpoint that hits scatterer A tangentially,*
- *points of W_1 collide (only) on scatterer A before time t_1 and miss scatterer B after that,*
- *points of W_2 miss scatterer A and collide on B before time t_2,*
- *$\Phi^{t_1} W_1$ and $\Phi^{t_2} W_2$ are good fronts in the corridor, tending down.*

Proof. To prove the lemma, we only need to understand the mechanism of infinite expansion of billiard dynamics near tangent collisions, as sketched in Figure 4. This shows that if we have two parallel incoming trajectories of some small distance h, one colliding tangentially, the other non-tangentially at the same scatterer, then the two trajectories are diverted by the collision to an extent much greater than h. Namely, $h = r(1 - \cos\alpha) \le r\alpha^2$ (for small enough h or α), so the slope of the post-collision trajectory line is

$$\tan 2\alpha \ge 2\alpha \ge 2\sqrt{\frac{h}{r}} = \frac{2}{\sqrt{r}}\sqrt{h}. \tag{3.1}$$

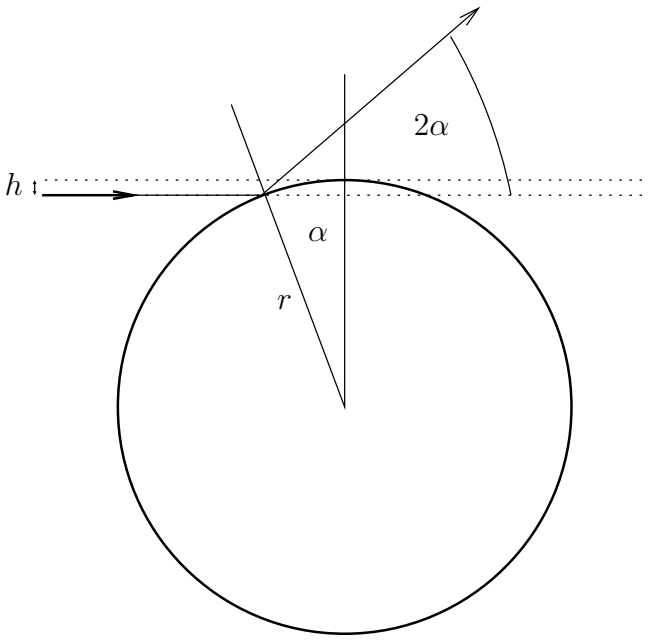


Figure 4: Greatly diverted trajectory near a tangent one

The requirements on the position of W and the direction of the velocities at the endpoints ensure that the upper endpoint hits A at a point at least $(KL/2 - 3)\varepsilon \ge 100\varepsilon$ higher than the bottom point of A, while the lower endpoint of W misses both A and B (if K is chosen large enough in terms of L). Due to the smoothness of the front, the points in between will scan the entire surface of A and B that is visible from W.

First, let W_2 be the set of points in W which do not hit A but hit B. (In Figure 3 only a subset of this is shown.) Let t_2 be a time shortly after all points of W_2 collided on B. The phase point with largest collision time is, actually, the lower endpoint of the incoming W_2, and it is hitting B tangentially. To see that $\Phi^{t_2}W_2$ is indeed a good front in the corridor tending down, it is enough to see that the endpoints satisfy the requirements for the direction of the velocity. For the upper endpoint (of the outgoing front) this is obvious, since the slope is required to be greater than $-\varepsilon/K$ (the trajectory can descend only slightly), but the velocity is actually positive (the outgoing trajectory is ascending). For the lower endpoint of the outgoing front (which is the Φ^{t_2}-image of the upper endpoint of the incoming front W_2) we have to check that A does not block too much of B. However, the geometric requirements ensure that this endpoint (after just avoiding a tangent collision on A) hits B at a point $h \geq \frac{d}{3L}\varepsilon$ high, so (3.1) guarantees that the slope after collision is

$$\leq -\frac{2}{\sqrt{r}}\sqrt{\frac{d}{3L}\varepsilon} = -2\sqrt{\frac{d}{3Lr}}\sqrt{\varepsilon} \leq -K\varepsilon$$

(descending steeply), if ε is small enough. (The fact that the incoming trajectory is not horizontal – unlike in Figure 4 – makes the situation even better.)

Second, let W_1 be the set of points in W which hit A and then miss B. (Again, in Figure 3 only a subset of this is shown.) Let t_1 be a time shortly after all points of W_1 collided on A. The phase point with largest collision time is, actually, the lower endpoint of the incoming W_1: after colliding on A, it is hitting B tangentially. Again, the only thing to check is the post-collision direction of the velocities at the endpoints. Note that the lower endpoint of the outgoing front is the Φ^{t_1}-image of the upper endpoint of the incoming front W_1, which is hitting A at a height at least 100ε if K is chosen sufficiently large. Hence it is easy to see that its velocity slope is $\leq -K\varepsilon$. For the upper endpoint one has to see that not too much of the outgoing front is blocked by B. An argument analogous to the previous one gives that we can guarantee a slope $\geq -\text{const}\varepsilon^2 \geq -\varepsilon/K$ (descending slowly) if ε is small enough. □

Proof of Proposition 3.1. Let K be so large and ε_0 be so small that the statement of Lemma 3.3 holds, and let $\varepsilon < \varepsilon_0$. For every $m \in \mathbb{R}$ let e_m denote the unit vector with slope m. We write $\mathbf{w}_0 = (q_0, v_0)$, so q_0 is the configuration part of the reference phase point $\mathbf{w}_0$. Let q_0' denote the configuration point which is $\frac{\varepsilon}{100}$ below q_0. Now consider the following set of phase points:

$$W_\emptyset := \Phi^{\frac{\varepsilon}{100K}}\left\{(q_0', e_m) \,:\, 0 \leq m \leq K\varepsilon\right\}.$$

In words: take a configuration point just a little bit below q_0, equip it with all possible velocities with slopes $0 \leq m \leq K\varepsilon$, and let these phase points evolve under the free flight for a very short time. The resulting set $W_\emptyset$ is a good front in the corridor tending up, all of whose points are within the $2K\varepsilon$-neighbourhood of $\mathbf{w}_0$. So we set $C = 2K$ and claim that $W_\emptyset$ contains the desired phase point $\mathbf{w}$ with the appropriate symbolic collision sequence σ – actually it contains an entire sub-front W_σ of such points. This completes the proof of the proposition.

We prove the claim by induction.

a.) For $n = 1$, Lemma 3.3 ensures the claim, with the choice $W_{(A)} := W_1$, $W_{(B)} := W_2$.

b.) Assume inductively that the claim holds for $n-1$ with some $n > 1$. Let $\sigma = (\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n)$. Assume first that $\sigma_1 = A$. Apply first Lemma 3.3 to obtain $W_{(\sigma_1)} = W_{(A)} = W_1$. The lemma ensures that $\Phi^{t_1}W_1$ is a good front, so the inductive assumption can be applied with $\sigma' = (\sigma_2, \sigma_3, \ldots, \sigma_n)$ to obtain the sub-front $W'_{\sigma'}$ of $\Phi^{t_1}W_1$ which has symbolic collision sequence σ'. Now $W_\sigma := \Phi^{-t_1}W'_{\sigma'}$ is a sub-front of W with symbolic collision sequence σ. The case when $\sigma_1 = B$ is exactly the same with $W_{(\sigma_1)} = W_{(B)} = W_2$, with $W'_{\sigma'}$ being a subfront of the good front $\Phi^{t_2}W_2$ and $W_\sigma := \Phi^{-t_2}W'_{\sigma'}$. This finishes the inductive proof of the claim, and the proof of the proposition.

□

3.2 Finite horizon

One could ask why we need to construct phase points with such arbitrary collision sequences, since in the infinite horizon case already the (first iterate of the) billiard map has infinitely many singularities accumulating at a single point, as shown by Figure 5. However, what we are truly interested in is the *finite* horizon case which we get by adding "transparent walls" to the infinite horizon configuration. These transparent walls are indicated by dashed lines in the figures. When one of these walls is reached, the velocity of the flying particle is unchanged, but a collision event is counted. So the example of Figure 5 does not work: many transparent walls are crossed before a distant scatterer can be hit, so the singular collisions shown in Figure 5 are only seen by high iterates of the (artificially-constructed) finite horizon billiard map. Indeed, the singularities of Figure 5 only contribute a linearly growing complexity.

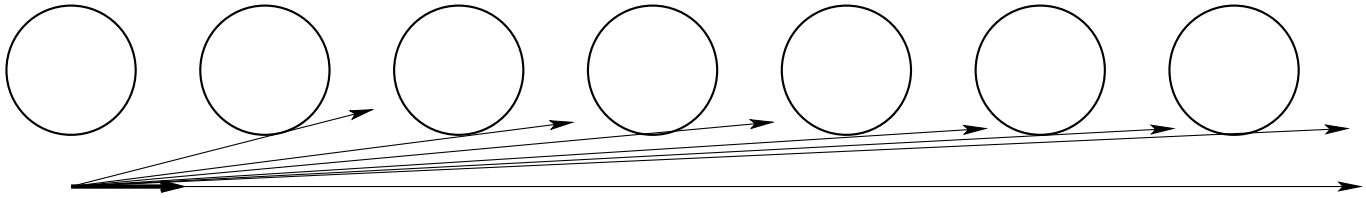
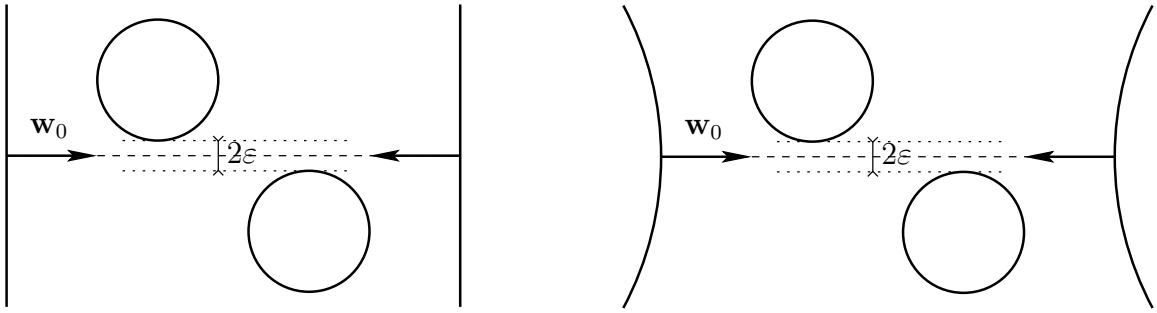


Figure 5: A bad example: infinitely many singularities of the infinite horizon billiard map accumulating at a point

On the other hand, many of the collision sequences of Proposition 3.1 are guaranteed to be seen by relatively low iterates of the map. Indeed, until reaching the k-th transparent wall, at least 2^k symbolic collision sequences are shown to exist, all of which are at most $2k$ long. If all these symbolic collision sequences were possible *arbitrarily* close to $\mathbf{w}_0$ for a fixed ε, it would mean that the complexity of T^n at $\mathbf{w}_0$ is at least $2^{n/2}$.

Next we consider further slight variations of the two dimensional configuration displayed in Figure 2. First let us replace the dashed lines by real straight lines as scatterers, such that the velocity of the billiard particle is reflected when it reaches these straight lines (see Figure 6(a)). The billiard dynamics obtained



(a) straight reflecting walls (b) curved reflecting walls

Figure 6: 2D finite horizon examples with many symbolic collision sequences

this way is equivalent to that of Figure 2, hence Proposition 3.1 applies. Now as a final variation, let us replace the straight lines by some strictly convex scatterers (say circles of some big radius), on which $\mathbf{w}_0$ bounces perpendicularly, see Figure 6(b). This way we obtain a two dimensional configuration with finite horizon and strictly dispersing scatterers, to which again Proposition 3.1 applies. To see this, note that this proposition relied on Lemma 3.3 which ensured that there is enough amount of scattering, so that good fronts are sufficiently expanded to realize both of the two possible continuations of the symbolic collision sequence. If instead of straight lines we have strictly convex scatterers, there is additional scattering, and the good fronts are expanded even stronger, so that the exponential growth of realizable sequences is maintained.

Again, we would like to emphasize that Figure 6(b) is *not* an example for the exponential growth of complexity. Here ε (the width of the corridor) is a fixed positive parameter, and the high number of

distinct collision sequences can be observed in an $\sim \varepsilon$-neighbourhood of $\mathbf{w}_0$. However, by adding one more dimension, we can realize this scenario with $\varepsilon \to 0$: this is exactly what happens in the three-dimensional configuration of Figure 1, as we discuss in Section 4.

4 Multidimensional dispersing billiards

In the present section we consider the three dimensional billiard configuration of Figure 1 and prove Theorem 2.8. In accordance with Remark 2.9, the essence of our reasoning is that given n (i.e. the length of the time interval we would like to control) it is possible to choose $\delta_0(n)$ such that for a large class of phase points $\mathbf{w}$ that are δ_0-close to the reference point $\mathbf{w}_0$, we can give an explicit description of how the trajectory of $\mathbf{w}$ deviates from that of $\mathbf{w}_0$ (for the n steps we are interested in). Once this explicit description is given, we will argue that the 2-dimensional picture in Figure 6 is of high relevance (cf. Proposition 4.3 below). The explicit description we give is accurate up to "negligible error" which should be interpreted as follows. Consider the distance $d_t = d(\Phi^t\mathbf{w}, \Phi^t\mathbf{w}_0)$ at a specific moment t within the studied time interval. Note that, by continuity of the flow Φ^t, if δ_0 is chosen sufficiently small, d_t can be regarded as a small parameter. Now by "negligible error" we mean that the actual $\Phi^t\mathbf{w}$ differs form our explicit description at most by $o(d_t)$: that is, our explicit description can be made arbitrarily accurate by choosing δ_0 sufficiently small.

A key feature of the configuration of Figure 1, on which our explicit description relies, is the following fact:

Lemma 4.1. *For phase points sufficiently close to $\mathbf{w}_0$, up to negligible error, perturbations in the y and z directions evolve independently in course of collisions with any of the four scatterers of Figure 1.*

We postpone the rigorous proof of this lemma to the Appendix. Here we only mention that the key to the proof is the well-understood behaviour of the two kinds of collisions along the trajectory:

- Collisions on the big scatterers are *(nearly) orthogonal* collisions on *spheres*. This has a scattering effect in both the y and z direction, but the y and z components of any perturbation evolve independently.
- Collisions on the small scatterers are *nearly tangential* collisions with the *surface normal vectors pointing (nearly) in the y direction*. This causes unbounded scattering in the y direction, but no scattering at all in the z direction. Again the two components of any perturbation evolve independently.

To proceed, recall that the parameter δ_0 in Remark 2.9 – which is the maximum distance of our phase point $\mathbf{w}$ from the singular periodic point $\mathbf{w}_0$ – can be chosen depending on the length n of the trajectory segment we consider, and we only need to consider a fixed n. So for any such fixed n, the distance of $\Phi^t\mathbf{w}$ from $\Phi^t\mathbf{w}_0$ can be made as small as we need during the *entire time period considered*, just by choosing δ_0 small enough. To see this, we need nothing more than

- the continuity of the flow Φ^t, which is *always* true, even at the multiply singular point $\mathbf{w}_0$.
- the boundedness (for a fixed n) of the time interval considered, which follows from the finiteness of the horizon.

This means that if $\delta_0 = \delta_0(n)$ is small enough, then, during the entire time period considered, the following picture about the trajectory is correct up to negligible error:

- Evolution of the x coordinate:
  As in the previous sections, x can be thought of either as a periodically changing, or as a monotonously increasing function of time – in fact, $x = t$, up to a negligible error, if δ_0 is small enough. We choose this latter point of view. This has the advantage that both y and z can be thought of as functions of time as the independent variable, or equivalently, as functions of x.

- Evolution of z:
  - Nearly tangential collisions on $2, 3$ have negligible effect, if δ_0 is small enough.
  - nearly orthogonal collisions on $1, 4$ cause scattering in the z direction, which is, however, independent of the evolution of y.

  So to understand the behaviour of z, consider the planar dispersing billiard configuration obtained by intersecting the $3D$ configuration of Figure 1 with the $x - z$ plane. This is shown in Figure 7. Indeed, the evolution of z is – up to negligible error – exactly the same as the evolution of z in this two-dimensional configuration.

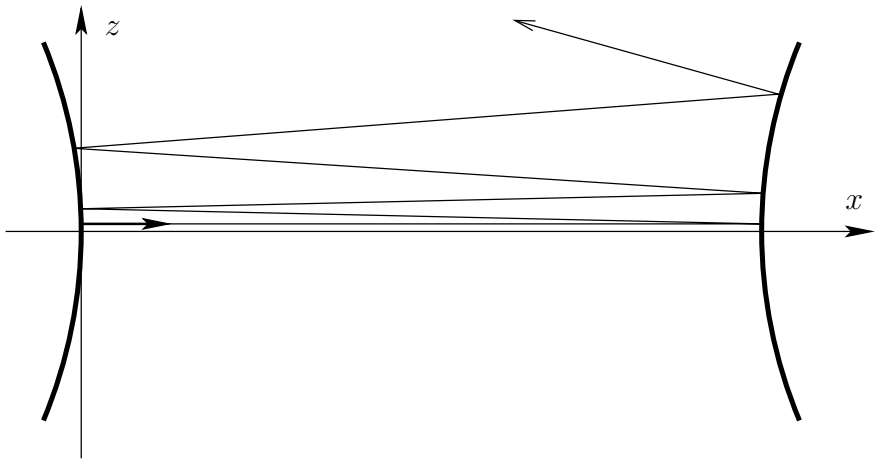


Figure 7: the evolution of z near the reference phase point

Before considering the evolution of y, we will first fix the z component of the initial perturbation of $\mathbf{w}_0$ to a specific value, and thus fix the entire evolution of z. In particular:

Let the pair $(\delta z, \delta v_z)$ be an *unstable* perturbation of the reference phase point $\mathbf{w}_0$. (4.1)

Then z grows along

$$z(t) \approx \delta z \, e^{\lambda t} \tag{4.2}$$

(without any decaying stable component, but approximated, of course, by a polygonal line), where λ is exactly the positive Lyapunov exponent of the reference trajectory in the configuration displayed on Figure 7. Later we will need the value of this unstable Lyapunov exponent, so we state it as a lemma:

Lemma 4.2. *The unstable Lyapunov exponent of the periodic point in the 2-dimensional billiard configuration of Figure 7 is*

$$\lambda = \frac{1}{L} \log(1 + \frac{L}{R} + \sqrt{\frac{L^2}{R^2} + \frac{2L}{R}}).$$

The proof is a routine calculation in billiard theory, and we postpone it to the Appendix.

- Evolution of y:
  - Consider the intersection of the configuration in Figure 1 with the $\{z = z_0\}$ plane for some $0 \neq |z_0| \ll 1$. This is exactly the configuration shown in Figure 6, right, with $\varepsilon = \frac{1}{2r} z_0^2 + o(z_0^2)$ (from the geometry of scatterers 2 and 3). If z were fixed and equal to z_0, y would evolve as in this planar configuration.

– However, the width ε of the corridor shown there increases in time, due to the growth of z, since at each time moment the intersection of the 3-dimensional configuration with a specific $\{z = const\}$ plane is relevant. Having fixed the t-dependence of z (by the choice of the z component of the initial perturbation) as (4.2), we have in fact "fixed the scenario" in which y evolves.

– Recall our convention on the description of the evolution of the x coordinate, according to which $x = t$, up to negligible error. As a consequence, (4.2) becomes

$$z(x) \approx \delta z \, e^{\lambda x}. \tag{4.3}$$

So the 2-dimensional configuration, according to which y evolves, is like in Figure 2, but the two dotted lines are not straight – instead, they are nearly horizontal, and their distance – although always small – grows as

$$\varepsilon(x) = \frac{1}{2r} z(x)^2 + o(z(x)^2) = \frac{(\delta z)^2}{2r} e^{2\lambda x}, \tag{4.4}$$

up to negligible error within the first n "half-periods" of length L, if the initial perturbation $(\delta z, \delta v_z)$ is small enough. This is shown in Figure 8. In particular, during a half-period L, the width of the corridor grows with a factor

$$M = (e^{\lambda L})^2. \tag{4.5}$$

Having fixed $(\delta z, \delta v_z)$ as (4.1) and thus the scenario for the evolution of y, we now vary the y component of the intial point $\mathbf{w}$ to get all the symbolic collision sequences in Σ_n. With this, the proof of the theorem will be complete.

To see that the symbolic collision sequences in Σ_n are all possible, it is enough to show that in this scenario, the modified version of Proposition 3.1 holds. Namely:

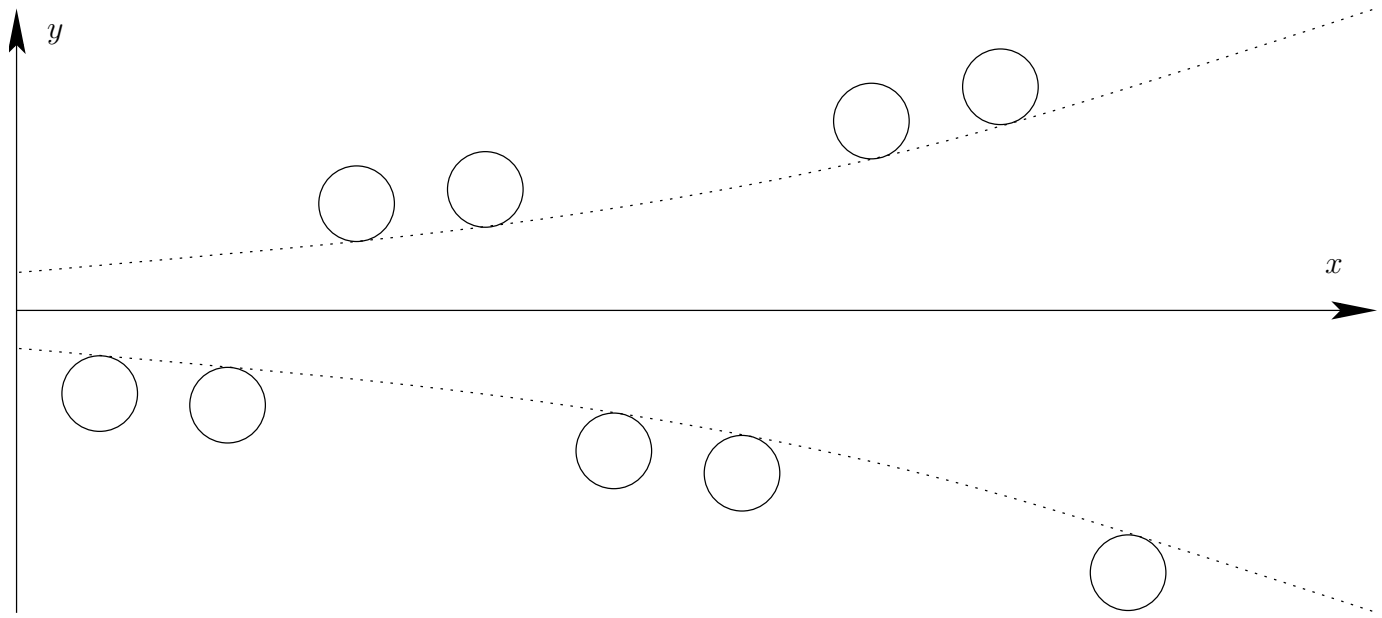


Figure 8: Expanding corridor as seen by the evolution of y

Proposition 4.3. *Consider the billiard configuration shown in Figure 8. Fix an $n \in \mathbb{N}$. Suppose that*

(i) the width of the corridor (the distance of the scatterers from the optical axis) grows monotonically with x,

(ii) there is some constant $M < \infty$ such that the width of the corridor grows at most by a factor M during one half-period (of length L) in x,

(iii) $M < 1 + \frac{d}{L}\,(< 2)$.

Then (for any R) there is some $\varepsilon_0 = \varepsilon_0(n)$ such that for every $\varepsilon_1 < \varepsilon_0$, whenever the width of the corridor is $\varepsilon(x) < \varepsilon_1$ for all $x < nL$, the following property holds. For any finite sequence $\sigma = (\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_n)$ of "A"-s and "B"-s it is possible to choose some $\mathbf{w}$ in the $C\varepsilon_1$-neighbourhood of $\mathbf{w}_0$ such that the n-long symbolic collision sequence of $\mathbf{w}$ is equal to σ. In other words, it is possible to hit exactly scatterer σ_i from each couple i as shown in Figure 8. So the number of possibilities grows exponentially with the length of the piece of the corridor considered. Here C is a numerical constant that could be determined explicitly, its value is, however, not important.

Proof. We do not write the proof of this proposition in full detail, because it contains no new idea compared to the proof of Proposition 3.1, but the increased amount of notation (due to the variable width of the corridor) could hide the essence. Instead, we point out the only element of the proof where more care is needed.

The proof of the proposition is based on the notion of "good fronts in the corridor" and Lemma 3.3 describing their evolution. The definition of this notion contains the parameter ε, which denoted (half of) the width of the corridor. In the new setting the width is not constant. However, our assumption (ii) guarantees that it changes at most by a factor $M^2 < 4$ within a period of length L. Since the mentioned lemma only concerns the evolution of a front within such a period, the half-width there grows from some $\varepsilon_{\text{left}}$ to some $\varepsilon_{\text{right}}$ with $1 < \frac{\varepsilon_{\text{right}}}{\varepsilon_{\text{left}}} < 4$. To get this notion of "good fronts in the corridor" for the setting of the growing corridor, the ε-s in the definition can be replaced by any value within the interval $[\varepsilon_{\text{left}}, \varepsilon_{\text{right}}]$. So Lemma 3.3 remains valid, and the proof is unchanged[2], except for one detail:

Since the width of the corridor is growing, to construct W_2 (the sub-front that hits B) we have to make sure that those phase points of W that hit B at the appropriate place, actually miss A before. In other words, the scatterer A should not cover B, as seen from the good front. This is ensured by condition (iii).

Once we have the appropriate version of Lemma 3.3, the inductive proof of the proposition, based on the lemma, goes through without modification. □

Proof of Theorem 2.8. To complete the proof of Theorem 2.8, all we need is to verify that the conditions of Proposition 4.3 are satisfied – in particular, condition (iii) needs to be checked, where M is given by (4.5), so we need

$$(e^{\lambda L})^2 < 1 + \frac{d}{L} \tag{4.6}$$

Qualitatively it is obvious that if R is chosen big enough (with L and d given), then the Lyapunov exponent λ of the reference trajectory in Figure 7 can be made sufficiently small to satisfy this requirement. To get the quantitative statement, we refer to Lemma 4.2 for the value of λ, and we write it in the form

$$\lambda = \frac{1}{L} \log(1 + \frac{L}{R}(1 + \sqrt{1 + 2\frac{R}{L}})).$$

Let us now demand that $\frac{L}{R} < 1$. This implies that $\frac{L}{R} < \sqrt{\frac{L}{R}}$. Using this estimate several times, we get

$$e^{\lambda L} = 1 + \frac{L}{R}(1 + \sqrt{1 + 2\frac{R}{L}}) < 1 + (1 + \sqrt{3})\sqrt{\frac{L}{R}},$$

and

$$(e^{\lambda L})^2 < 1 + [2(1 + \sqrt{3}) + (1 + \sqrt{3})^2]\sqrt{\frac{L}{R}} < 1 + 13\sqrt{\frac{L}{R}}.$$

This means that the condition (4.6) is guaranteed to hold if

$$1 + 13\sqrt{\frac{L}{R}} < 1 + \frac{d}{L} \quad (\text{and } \frac{L}{R} < 1),$$

[2]All estimates containing ε are generous enough to remain valid with the above uncertainty in choosing ε.

which follows from

$$R > 169\frac{L^3}{d^2},$$

which is exactly the condition assumed in Theorem 2.8. So condition (iii) of the proposition is verified.

To complete the proof of Theorem 2.8, for given δ and n-long symbolic collision sequence $\mathbf{j} \in \Sigma_n$, we need to specify a phase point $\mathbf{w}$ in the δ-neighbourhood of $\mathbf{w}_0$ with $\mathbf{i}_{2n-1}(\mathbf{w}) = \mathbf{j}$. In the spirit of Remark 2.9, first we choose δ_0 according to two requirements:

- $\delta_0(n) \le 2C\varepsilon_0(n)$, where C and $\varepsilon_0(n)$ are the constants that appear in the statement of Proposition 4.3,
- $\delta_0(n)$ should be small enough so that, in the δ_0-neighbourhood of $\mathbf{w}_0$, the above description of the evolution of y and z coordinates remains true with the required precision.[3]

Then we let $\delta_1 = \min(\delta, \delta_0)$ and construct $\mathbf{w}$ with symbolic sequence $\mathbf{j}$ in the δ_1-neighbourhood of $\mathbf{w}_0$ in two steps.

- As a first step we consider a phase point $\mathbf{w}'$ which is a perturbation of $\mathbf{w}_0$ in the $x - z$ plane by an unstable vector $(\delta z, \delta v_z)$, cf. Figure 7 and the description preceding Lemma 4.2. The length of this unstable perturbation, which we denote by δ_2, should satisfy the following two requirements:
  (i) $\delta_2 < \delta_1/2$,
  (ii) δ_2 should be so small that throughout the time interval considered (i. e. for $0 \le x \le nL$) the width of the corridor in the $x - y$ plane, which is determined by Formula (4.4), should not exceed $\frac{\delta_1}{2C}$.
- Now as a second step, we apply Proposition 4.3 with $\varepsilon_1 = \frac{\delta_1}{2C}$. Our requirement (ii) on δ_2 ensures that there exists a phase point $\mathbf{w}$ in the $C\varepsilon_1$-neighbourhood of $\mathbf{w}'$ with symbolic sequence $\mathbf{j}$.

Finally, by our requirement (i) on δ_2, we have

$$d(\mathbf{w}_0, \mathbf{w}) \le d(\mathbf{w}_0, \mathbf{w}') + d(\mathbf{w}', \mathbf{w}) \le \frac{\delta_1}{2} + C\varepsilon_1 = \delta_1,$$

which completes the proof of Theorem 2.7. □

The proof of Lemma 2.10 (along with its Corollary 2.11) is given in the Appendix as it is directly related to some calculations presented there. It remains to prove Theorem 2.8.

Proof of Theorem 2.7. Take a multi-dimensional dispersing billiard configuration with finite horizon. Such configurations exist, even without corner points, which follows e.g. form the results of [BoTa]. Then add to this configuration a (suitably small) copy of the configuration given in Theorem 2.8, so that the newly inserted spheres do not intersect any of the old scatterers. This configuration obviously satisfies the claim of the theorem. □

Appendix

Here we present the proofs of lemmas 4.2, 4.1, 2.10 and Corollary 2.11.

[3]That is, the existence of phase points with all specified symbolic collision sequences, as described in Proposition 4.3, should remain valid (note that this is an open condition).

Proof of Lemma 4.2. Of course, the *Ljapunov exponent* we are looking for is a quantity corresponding to the billiard *flow*. When using it later, we will translate it into the appropriate *expansion factor* of the dicrete time *map*.

To prove the lemma, we need to consider the evolution of perturbations of the periodic point on the billiard configuration of Figure 7. Since this is a planar configuration, general perturbations are two dimensional. It is convenient to follow a common convention in billiard theory: instead of using directly the coordinates on the flow phase space $\mathcal{M}$, we describe the evolution of such two dimensional perturbations in terms of the coordinates of the local orthogonal section to the flow. That is, any non-neutral perturbation of the periodic point gives rise to an infinitesimal local orthogonal front (see Section 3 on the description of local orthogonal fronts). Such an infinitesimal front can be characterized by a two dimensional vector $(\delta q, \delta v)$, the two components of which are the configuration and the velocity components of the perturbation on the front.

Remark 4.4. *The fact that we are looking for the stable and unstable perturbations in a 2-dimensional sub-plane of the tangent space correspond to the fact that*

- *perturbations changing the length of v are not allowed: they would change the energy and thus lead out of the phase space $\mathcal{M}$, which is a* constant energy surface*.*
- *Perturbations of the configuration point in the velocity direction are obviously neutral perturbations independent of both the stable and the unstable direction.*

So it remains to consider perturbation for which both the configuration and the velocity component is orthogonal to the velocity – one dimension each.

Now an unstable perturbation (and also a stable perturbation) $(\delta q, \delta v)$ is characterized by the fact that it is mapped into a constant $\tilde{M}^2$ times itself by the derivative of the dynamics during an entire period. Moreover, $\tilde{M}^2 = e^{2L\lambda}$ where $2L$ is the length of the periodic orbit and λ is the Lyapunov exponent. Since the considered configuration is symmetric, it is enough to consider a half-period of length L, during which the perturbation $(\delta q, \delta v)$ is mapped into $(\tilde{M}\delta q, \tilde{M}\delta v)$ with $\tilde{M} = e^{L\lambda}$. The evolution during the half-period consists of two steps:

- During the free flight, $(\delta q, \delta v)$ evolves into $(\delta q', \delta v') = (\delta q + L\delta v, \delta v)$.
- During the collision, $(\delta q', \delta v')$ evolves into $(\delta q'', \delta v'') = (\delta q', \delta v' + \frac{2}{R}\delta q')$. This comes directly from the differentiation of the collision formula $v'' = v' - 2\langle n, v'\rangle n$, since the collision normal vector n has the differential $\delta n = \frac{1}{R}\delta q'$ and $\langle n, v'\rangle = 1$.

Putting these together, we need to find the nontrivial solutions of

$$(\tilde{M}\delta q, \tilde{M}\delta v) = (\delta q + L\delta v, \delta v + \frac{2}{R}(\delta q + L\delta v)).$$

Introducing the notation $\kappa = \frac{\delta v}{\delta q}$ (for the curvature of the front), this leads to $\tilde{M} = 1 + L\kappa$ and

$$\kappa = \frac{1}{R} \pm \sqrt{\frac{1}{R^2} + \frac{2}{RL}}.$$

All in all,

$$\tilde{M} = 1 + \frac{L}{R} \pm \sqrt{\frac{L^2}{R^2} + \frac{2L}{R}}.$$

Of course, the greater value (with the $+$) corresponds to the unstable perturbation, and the other one to the stable, so the unstable Lyapunov exponent is

$$\lambda = \frac{1}{L}\log(1 + \frac{L}{R} + \sqrt{\frac{L^2}{R^2} + \frac{2L}{R}}).$$

□

Proof of Lemma 2.10. This lemma is about the minimum unstable expansion factor in a 3-dimensional billiard. What "minimum expansion factor" exactly means can slightly vary depending on the way how the hyperbolic properties are formulated for our system (e.g. how the unstable cones are defined). Fortunately we do not need to go into the details, since – whatever formulation of hyperbolicity is used – the minimum expansion factor can never exceed the expansion factor for any direction in the unstable subspace (an *unstable* direction for short). In particular we have

$$\Lambda \le e^{\tilde{\lambda}}$$

if $\tilde{\lambda}$ is an unstable Lyapunov exponent of some periodic point (for the discrete time map). We prove our statement by looking at an unstable Lyapunov exponent of the periodic point $\mathbf{w}_0$. Namely, this point clearly has an unstable perturbation which is entirely in the $x-z$ plane: this is exactly the unstable perturbation of the periodic point in the 2-dimensional billiard of Figure 7. The corresponding Lyapunov exponent is given by Lemma 4.2 as $\lambda = \frac{1}{L}\log(1+\frac{L}{R}+\sqrt{\frac{L^2}{R^2}+\frac{2L}{R}})$ *for the flow*, so we get $\tilde{\lambda} = L\lambda = \log(1+\frac{L}{R}+\sqrt{\frac{L^2}{R^2}+\frac{2L}{R}})$ for the map. This gives

$$\Lambda \le 1 + \frac{L}{R} + \sqrt{\frac{L^2}{R^2}+\frac{2L}{R}}.$$

The right hand side tends to 1 as $\frac{L}{R}$ approaches zero, which completes the proof of the lemma. □

Proof of Corollary 2.11. By Theorem 2.8, complexity of the singularity set at $\mathbf{w}_0$ grows as $m_n(\mathbf{w}_0) \ge c2^{n/2}$ for some fixed $c > 0$. On the other hand, by Lemma 2.10 for R big enough we have $\Lambda < \sqrt{2}$. This completes the proof of the Corollary. □

Proof of Lemma 4.1. Throughout the argument, we will use the notion of negligible error in the sense described at the beginning of Section 4. The reference trajectory (of $\mathbf{w}_0$) is characterized by the following property: for any phase point $\Phi^t\mathbf{w}_0 = (q_t^0, v_t^0)$ along the trajectory the y and z coordinates of both the position q_t^0 and the velocity v_t^0 are identically zero. By following a "perturbed" trajectory "near" this reference trajectory for a fixed time interval $[0, T]$ we mean that for $\mathbf{w}$ sufficiently close to $\mathbf{w}_0$ we keep track of the y and z components of both the position and the velocity of $\Phi^t\mathbf{w}$, as long as $0 \le t \le T$. When we claim that the y and z directions evolve independently (in course of any single collision) we mean that the post-collision values of z and v_z can be calculated from the pre-collision values of z and v_z only, and similarly, the post-collision values of y and v_y can be calculated from the pre-collision values of y and v_y only – up to negligible error. In what follows we show that all errors that arise are indeed negligible.

First note that during free flight the velocity of the particle is unchanged, while the evolution of y and z is determined by the value of v_y and v_z, respectively. In course of a (momentary) collision, the position of the particle is unchanged – that is, the collision has an immediate effect only on the velocity. Hence it is enough to investigate the post-collision values of v_y and v_z, the part of the statement which concerns the evolution of y and z will follow automatically. As mentioned just after the statement of Lemma 4.1, there are two types of collisions along the reference trajectory: nearly orthogonal and nearly tangential collisions. In both cases, the independence of the evolution of the two components of a perturbation can be seen from the collision rule

$$v^+ = v^- - 2\langle n, v^-\rangle n = v^- - 2\cos\varphi\, n,$$

where v^- is the pre-collision velocity, v^+ is the post-collision velocity, n is the surface normal vector at the collision point and $\varphi = \sphericalangle(n, v^+) \in [0, \frac{\pi}{2}]$ is the collision angle.

- In case of the nearly orthogonal collisions (on scatterers 1 and 4), we have $\cos\varphi \approx 1$ within arbitrarily small error, so

  $$v^+ = v^- - 2n.$$

  But in this case – again, within arbitrarily small relative error –

  $$n = n_0 + Kq,$$

  where

– n_0 is the surface normal for the reference trajectory – that is, $n_0 = (1; 0; 0)$ for collisions on sphere 1 and $n_0 = (-1; 0; 0)$ for collisions on sphere 4,

– $q = (0, y, z)$ is the relevant component of the collision configuration point (that is, the component orthogonal to v_0),

– K is the curvature operator (or second fundamental form) of the scatterer (at the collision point of the reference trajectory).

The crucial point is that the operator K is *diagonal* – in fact, it is the scalar $K = \frac{1}{R}$, since the scatterers are *spheres*. This means that – again, within arbitrarily small relative error –

$$v^+ = v^- - 2n_0 - \frac{2}{R}q,$$

so indeed, the change in the y and z component of the velocity depends only on y and z, respectively.

- In case of the nearly tangential collisions (on scatterers 2 and 3), the scatterers are *spheres* centered somewhere in the $x - y$ plane, so the z coordinate of n is simply proportional to the z coordinate of the collision point – in particular, $n_z = z/r$. So we have

$$v_z^+ = v_z^- - 2\cos\varphi\,\frac{z}{r}.$$

Since for these collisions $\cos\varphi \approx 0$ within arbitrarily small (absolute) error, this gives

$$v_z^+ = v_z^-$$

within an error arbitrarily small (relative to z). On the other hand, the y coordinate of n is nearly ± 1, so $\langle n, v\rangle \approx \pm y$, which leads to

$$v_y^+ = v_y^- \mp 2y,$$

again within arbitrarily small relative error.

□

Acknowledgements

This research has been partially supported by ERC via the University of Helsinki (IPT), by the Bolyai Scholarship of the Hungarian Academy of Sciences (both authors), and by OTKA (Hungarian National Fund for Scientific Research) grants PD73609 (IPT), F60206 (PB), and K71693 (both authors).

References

[BG1] V. Baladi and S. Gouëzel; *Good Banach spaces for piecewise hyperbolic maps via interpolation*, Annales de l'Institut Henri Poincaré / Analyse non linaire **26** (2009) 1453–1481.

[BG2] V. Baladi and S. Gouëzel; *Banach spaces for piecewise cone hyperbolic maps*, Journal of Modern Dynamics, **4** (2010) 91–137.

[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** (2002) 451 – 482.

[BaTo] P. Bálint and I. P. Tóth, *Exponential Decay of Correlations in Multidimensional Dispersing Billiards*, Annales Henri Poincaré **9** (2008) 1309–1369.

[BoTa] K. Böröczky, Jr. and G. Tardos, *The longest segment in the complement of a packing*, Mathematika, **49** (2002), 45–49.

[BSC] L. Bunimovich, Ya. Sinai and N. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Russian Mathematical Surveys **46** (1991) 47–106

[Bu] J. Buzzi, *No or infinitely many a.c.i.p. for piecewise expanding C^r maps in higher dimensions*, Communications in Mathematical Physics **222** (2001) 495-501.

[Ch] N. Chernov, *Sinai billiards under small external forces*, Annales Henri Poincaré, **2** (2001) 197–236

[CD] N. Chernov and D. Dolgopyat, *Brownian Brownian motion I.*, Memoirs of the American Mathematical Society, Volume 198, Number 927 (2009).

[CM] N. Chernov and R. Markarian, *Chaotic billiards*, Mathematical Surveys and Monographs **127** (2006) AMS, Providence, RI.

[CZ] N. Chernov anf H.-K. Zhang, *On statistical properties of hyperbolic systems with singularities*; Journal of Statistical Physics, **136** (2009), 615–642.

[DL] M. Demers and C. Liverani, *Stability of Statistical Properties in Two-dimensional Piecewise Hyperbolic Maps*, Transactions of the American Mathematical Society **360** (2008), 4777–4814.

[LY] A. Lasota and J. Yorke, *On the existence of invariant measures for piecewise monotonic transformations*, Transactions of the American Mathematical Society **186** (1973), 481–488.

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

[T] M. Tsujii, *Piecewise expanding maps on the plane with singular ergodic properties*, Ergodic Theory Dynam. Systems **20** (2000) 1851-1857.

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