Semi-dispersing billiards with an infinite cusp

Marco Lenci
Institute for Mathematical Sciences
SUNY at Stony Brook
Stony Brook, NY 11794-3660
U.S.A.

E-mail: `lenci@math.sunysb.edu`

Abstract

Let $f : [0, +\infty) \longrightarrow (0, +\infty)$ be a sufficiently smooth convex function, vanishing at infinity. Consider the planar domain Q delimited by the positive x-semiaxis, the positive y-semiaxis, and the graph of f.

Under certain conditions on f, we prove that the billiard flow in Q has a hyperbolic structure and, for some examples, that it is also ergodic. This is done using the cross section corresponding to collisions with the dispersing part of the boundary. The relevant invariant measure for this Poincaré section is infinite, whence the need to surpass the existing results, designed for finite-measure dynamical systems.

Mathematics Subject Classification: 37D50, 37D25, 37A40.

1 Historical introduction

There is a long tradition in the study of hyperbolic billiards, especially billiards in the plane. This was initiated by Sinai as early as 1963 [S1], in connection with the Boltzmann Hypothesis in statistical mechanics. In his celebrated 1970 paper [S2], Sinai proved the first cornerstone theorem of the field: the K-property of a billiard in a 2-torus endowed with a finite number of convex scatterers (of positive curvature). The result was polished in a later joint work with Bunimovich [BS] and extended to a larger class of *dispersing* billiards. This terminology was introduced precisely in that paper and designates billiard tables whose boundaries are composed of finitely many convex pieces, when seen from the interior. In particular, the new theorem allowed for positive-angle corners at the boundary.

In the following years this area of research made progress in several directions. For instance, Gallavotti and Ornstein [GO] proved that the Sinai billiard is also

Stony Brook IMS Preprint #2001/13
December 2001

Bernoulli; Bunimovich [B] found an example of a hyperbolic and chaotic billiard with flat and *focusing* (as opposed to dispersing) boundary—the famous Bunimovich stadium. Also, Vaserstein's 1979 paper [V] provided the missing ingredient for the proof of the ergodic properties of a billiard with *cusps*, i.e., zero-angle corners. (In short, the hurdle was that, for cuspidal billiards, the amount of phase-space contraction/expansion after n collisions is not uniformly large as a function of the phase-space point. In jargon, this is a *non-uniformly hyperbolic* system.)

However, it was not until the 1980s that this field was made into an organized theory, the theory of *singular hyperbolic dynamical systems*. We will not even try to summarize the hefty literature concerned with it, but will only mention the works that are most relevant to the present paper. Among these, a prominent place is deserved by the book of Katok and Strelcyn [KS]. One of the most valuable results they proved there is that a very general class of billiards has a *hyperbolic structure*. By this we mean that local stable and unstable manifolds exist almost everywhere and these local foliations are absolutely continuous w.r.t. the invariant measure. In practice, they adapted Pesin's theory for smooth hyperbolic systems [P] to systems with singularities. Actually, again following Pesin, they also showed that such systems admit a very general structure theorem. This claims that the phase space splits into a finite or countable number of ergodic components over which the map has the K-property (more or less—in fact, a power of the map does).

This last theorem, as nice as it is, is not of big practical applicability, since it does not give any characterization of the ergodic components. This is needed, for example, to show that there is only one of them. Researchers have henceforth tried to prove *local ergodicity* theorems, theorems that guarantee that a point with enough properties (usually called a *sufficient* point) has a neighborhood contained in one ergodic component. Any such result is nowadays called a *fundamental theorem* for billiards (the terminology comes from [BS]), and there is a variety of them, formulated in more or less general ways. Sinai and Chernov in 1987 [SC] wrote a version especially tailored for systems of several 2-dimensional discs or 3-dimensional balls (these correspond to higher dimensional semi-dispersing billiards). This result was later improved and extended by Krámli, Simányi and Szász [KSSz]. (It should be mentioned that these two papers are slightly incorrect, for they overlook that a certain technical condition is not verified for billiards in more than two dimensions. This was discovered by the very authors together with others, and fixed in a recent paper [BCSzT], at least for a large class of dispersing billiards.)

Among the more general formulations of the fundamental theorem we have [C1] and [LW]. The latter reference, by Liverani and Wojtkowski, has the nice feature that it uses *invariant cones*. This elegant tool has been very effective in hyperbolic dynamics. Wojtkowski in 1986 wrote a beautiful paper [W] on how to use geometrical optics to define invariant cones for planar billiards. (Here, however, we will use different cones, whose construction dates back to [S2].) This work was improved by Markarian [M1] in 1988.

In the last decade researchers have focused on a yet harder problem, of great importance for the physics that these billiards are supposed to resemble: the decay of correlations. We will be even more superficial on this widely studied question, since it does not concern this note. But it is interesting to recall that it took the community more than one decade to get convinced that a large class of hyperbolic billiards has exponential decay of correlations, and even longer to prove it. This was very recently done by Young [Y], and her techniques better adapted to billiards by Chernov [C2]. The class of systems for which exponential decay holds includes at least all billiards treated in [BS], but does not include dispersing billiards with a cusp, whose decay of correlations is polynomial [CM].

In the present article we study the case of a semi-dispersing billiard with a *non-compact* cusp. Given a function f defined on $[0, +\infty)$ with values in the positive numbers, sufficiently smooth, convex and vanishing at $+\infty$, we consider the table Q delimited by the positive x-semiaxis, the positive y-semiaxis, and the graph of f, as in Fig. 1. For many such billiards we will construct a hyperbolic structure, and for some of them we will also prove ergodicity.

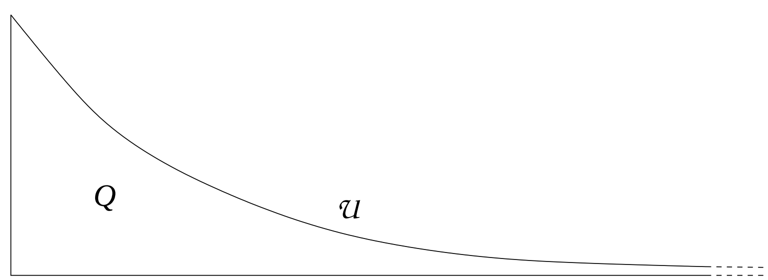

Figure 1: An example of a billiard table Q.

Geometrically speaking, one cannot fail to notice the similarities between this system and the geodesic flow on a non-compact modular surface in the hyperbolic half-plane. The ergodic properties of this flow were proved by Hopf, for both compact and non-compact surfaces [CFS]. However, dynamically speaking, the modular surface is not a good analogue for our billiard, even leaving aside the smoothness of the flow. In fact, in the hyperbolic half-plane, the curvature of the metrics acts as a *constant* source of hyperbolicity. In other words, it does not matter if the particle spends a given amount of time in the "bulk" of the surface or in the cusp. In the case of the billiard, instead, once the particle is deep into the cusp, one does not know *a priori* whether the increased number of rebounds per unit time compensates the flattening of the dispersing boundary and the decrease of the mean free path.

As important as this, is the question of the finiteness of the area. Contrary to the modular surface case, we do not assume Q to have finite area. In fact, our model represents one of the very few treated examples of singular hyperbolic dynamical systems with *infinite* invariant measure. This claim calls for an explanation. There is a serious amount of literature concerning infinite-measure systems, even among dispersing billiards [LM], or billiards with a cusp [L, K, Le1] ([DDL1, DDL2] are polygonal caricatures of the system at hand). Seldom, though, are they studied from the point of view of hyperbolicity and ergodic theory. Perhaps the only close relative to the present model is the Lorentz gas in two dimensions. In recent times, Schmidt [Sch] and Conze [Co] have (independently) proved its recurrence, which yields ergodicity, according to a previous result by Simányi [Si].

But even the Lorentz gas is not a very good analogue, since, thanks to recurrence, one really studies the return map to a finite-measure Poincaré section. Despite our efforts, this did not seem to be a fruitful approach for the infinite-cusp billiard. What we do is consider the return map to the dispersing part of ∂Q. This choice corresponds to an infinite-measure cross section, independently of the integrability of f.

As a matter of fact, although the underlying ideas used here are the same as in the literature previously cited, hardly anything works directly. Particularly disappointing is the fact that we cannot use [KS] to construct a hyperbolic structure. Establishing this becomes our real "fundamental theorem"! We achieve it by introducing two techniques: the *unstable distance* (Section 5) and the concept of *fuzzy boundary* for a sub-cross-section (Section 6). (I have recently learned that a notion of unstable distance was already present in [M2].)

Apart from a number of technical results, the other serious task that we have to take on is to devise our own version of the local ergodicity theorem. We do so at the end of this note, extending the theorem of [LW] to infinite-measure systems.

2 Summary of the results

We give a summary of the paper that presents the main results in some detail. As this section is directed to the reader already familiar with hyperbolic dynamical systems, we will use terminology that is defined precisely only later on.

Section 3 lays down the basic definitions together with their immediate consequences. In particular it includes the definition of $(\mathcal{M}, T, \mu)$, the dynamical system with which we will deal for most of the paper: $\mathcal{M}$ is the Poincaré section corresponding to the rebounds against the dispersing boundary of the billiard; T is the corresponding return map, and μ is the (infinite-mass) invariant measure induced on $\mathcal{M}$ by the billiard flow. In Section 4 we introduce a suitable invariant cone bundle. Section 5 defines the *unstable distance* and describes the neighborhoods of the singularities w.r.t. this distance.

In Section 6 we prove the existence of the local stable and unstable manifolds and in Section 7 their absolute continuity. By local stable manifold, in this context, we mean a smooth topological disk, of the right dimension, that contracts asymptotically in the future, and such that the intersection of two such objects is also a stable manifold (Definition 6.1). The right dimension is determined by the dimension of the cone bundle (1 in our case), as no Osedelec theorem is available for our systems.

In fact, nothing is known about the Lyapunov exponents—not even their existence. One likely possibility, however, is that they are zero; that is, the rate of contraction is sub-exponential. The main idea in the proof of the existence is the concept of *fuzzy boundary* for a sub-cross-section (cf. Theorems 6.2 and 6.5).

Section 8 is devoted to the ergodicity. The main result is a version of the local ergodicity theorem for infinite-measure systems with singularities. Then, for all billiards that verify the hypotheses of this theorem, we derive that the map T is ergodic. Since μ is infinite, it is important to specify what definition of ergodicity we adopt. We choose the Boltzmann original formulation: the time average of every integrable function is constant almost everywhere (Definition 8.1). This is a rather weak notion of ergodicity, as it does not even prevent the existence of two complementary invariant subsets of infinite measure. However, we also prove a much more satisfactory result: the ergodicity of the (finite-measure) Poincaré map corresponding to the returns onto the vertical side of the boundary.

In Section 9 we check that certain examples of billiards do in fact verify all the conditions of the previous section and are thus ergodic.

Finally, some of the less important lemmas are proven in the Appendix.

Of course, all these results require assumptions on the function f. For the convenience of the reader, we list them all at the beginning of Section 3 and reference them appropriately in the formulation of each theorem. For now, let us anticipate that two classes of examples, $f(x) = Ce^{-kx}$ $(k > 0)$ and $f(x) = Cx^{-p}$ $(p > 0)$, have a hyperbolic structure, in the sense of Sections 6 and 7; furthermore, the latter family is also ergodic. This is particularly interesting since, within this family, we find tables of infinite area. For these systems, one cannot hope to avoid the hurdles associated to infinite-measure dynamical systems by studying a 3-dimensional flow, instead of a 2-dimensional map. (Not that this approach would be effortless, anyway.)

A large class of examples of ergodic billiards is found in [Le2].

3 Mathematical preliminaries

A billiard is a dynamical system defined by the free motion of a material point inside a domain, subject to the Fresnel law of reflection at the boundary, i.e., the angle of reflection is equal to the angle of incidence. Here we are concerned with domains (otherwise referred to as *tables*) of the form: $Q := \left\{ (x,y) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ \mid 0 \le y \le f(x) \right\}$ (see Fig. 1), where $f : \mathbb{R}_0^+ \longrightarrow \mathbb{R}^+$ is three times differentiable, bounded and convex. Thus the table is semi-dispersing. The angle at the vertex $V := (0, f(0))$ is $\pi/2 + \arctan f'(0^+)$ and is allowed to be zero; so the billiard might have a compact cusp, together with the non-compact cusp at $x = +\infty$.

For many geometrical proofs, throughout the paper, it will be convenient to introduce two more domains in the plane, $Q_2 := \left\{ (x,y) \in \mathbb{R}_0^+ \times \mathbb{R} \mid |y| \le f(x) \right\}$, a two-fold copy of Q, and $Q_4 := \{ (x,y) \in \mathbb{R} \times \mathbb{R} \mid |y| \le f(|x|) \}$, a four-fold copy (see Figs. 2 and 3 later on).

Before moving on, we list all the assumptions on f that we use in the rest of this paper, specifying which result requires which hypothesis. To this purpose, we give an extra definition, that will be introduced with the aid of Fig. 2.

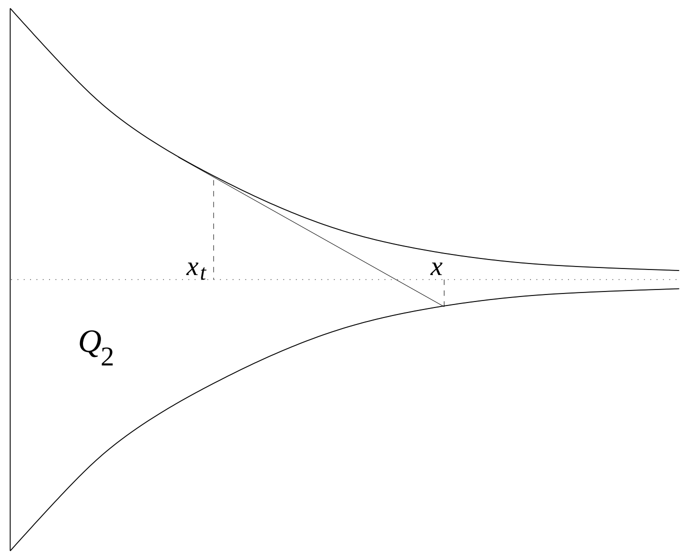

Figure 2: The definition of x_t.

In Q_2, for $x > 0$, consider the straight line passing through $(x, -f(x))$ and tangent to ∂Q (i.e., to the part of ∂Q_2 that lies in the first quadrant). Denote by $x_t = x_t(x)$ the abscissa of the tangency point. This is uniquely determined by the equation

$$\frac{f(x) + f(x_t)}{x - x_t} = -f'(x_t). \tag{3.1}$$

In the sequel, for $f, g \geq 0$, we use the notation $f(x) \ll g(x)$ to indicate that there is constant C such that $f(x) \leq C\, g(x)$, as $x \to +\infty$; likewise for $\gg$. Also, $f(x) \sim g(x)$ means that, when $x \to +\infty$, $f(x)/g(x)$ is bounded away from 0 and $+\infty$. Later on, where there is no danger of confusion, we use the same symbols for different asymptotics, such as $\varepsilon \to 0^+$, and so on.

Assumptions for the existence of the local stable and unstable manifolds:

$$f''(x) \to 0; \tag{A1}$$

$$|f'(x_t)| \ll |f'(x)|; \tag{A2}$$

$$\frac{f(x) f''(x)}{(f'(x))^2} \gg 1; \tag{A3}$$

$$\frac{|f'''(x)|}{f''(x)} \ll 1. \tag{A4}$$

Assumption for the absolute continuity of the local stable and unstable manifolds:

$$|f'(x)| \gg (f(x))^{\theta}, \text{ for some } \theta > 0. \tag{A5}$$

Examples of ergodicity:

$$f(x) = Cx^{-p}, \text{ with } C, p > 0, \tag{E1}$$

Remark 3.1 We have worked out the ergodicity result for a class of examples, rather than giving general conditions, because such conditions would be rather unhandy to write down and to verify (see Section 9). On the other hand, it is a safe conjecture to say that ergodicity must hold for many other examples as well, including $f(x) = Ce^{-kx}$.

We notice that (A1) and (A4) imply, via a De l'Hôpital-like argument, that

$$\frac{|f'(x)|}{f(x)} \ll \frac{f''(x)}{|f'(x)|} \ll 1, \tag{3.2}$$

since f'' (hence f and f') vanish at infinity. It is evident that none of the above assumptions depend on multiplicative constants in front of f. We now check that $f(x) = x^{-p}$, with $p > 0$, and $f(x) = e^{-kx}$, with $k > 0$, verify (A1)-(A5).

The other conditions being trivially satisfied, we only need to worry about (A2). For $f(x) = x^{-p}$, after a simple algebraic manipulation, (3.1) reads

$$\frac{\left(\dfrac{x_t}{x}\right)^p + 1}{1 - \dfrac{x_t}{x}} = p \left(\frac{x_t}{x}\right)^{-1}, \tag{3.3}$$

which gives $x_t/x = const$, whence (A2). For $f(x) = e^{-kx}$, (3.1) amounts to

$$\frac{e^{-k(x-x_t)} + 1}{x - x_t} = k; \tag{3.4}$$

hence $x - x_t = const$, yielding (A2) for this case, too.

A state in our system is completely specified by the position (x, y) and velocity v of the point. Since the kinetic energy is a constant, one can assume that $|v| = 1$. Hence the natural phase space for the flow is $Q \times S^1$. In the terminology of [S2], points in this space are called *line elements*. The relevant Liouville measure here is the Lebesgue measure, which is therefore left invariant by the flow.

It is customary, especially if one is only interested in the ergodicity, to work with a cross-section. For billiards, one usually chooses the cross-section corresponding to rebounds of the point against ∂Q. In our case, the geometry of the billiard (see Q_4) suggests that we restrict to rebounds against the dispersing part of ∂Q, which we

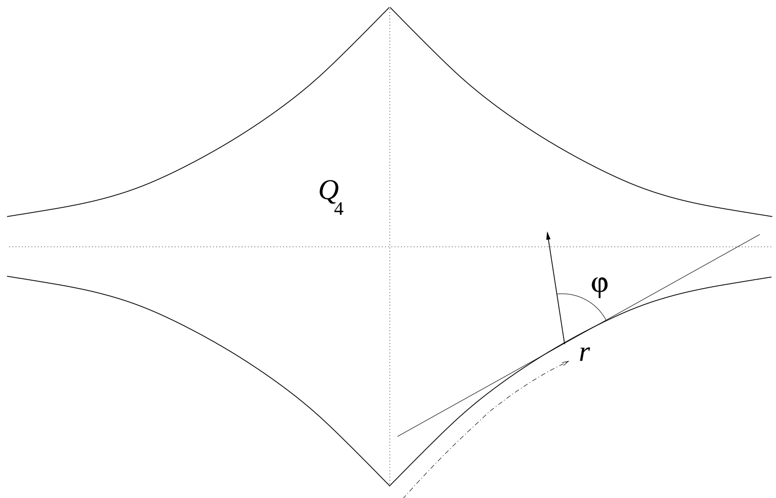

Figure 3: The definition of r and φ.

denote by $\mathcal{U}$. More precisely, we consider only unit vectors based in $\mathcal{U}$ and pointing towards the interior of Q. We parameterize these line elements by $z := (r, \varphi)$, where $r \in [0, +\infty)$ is the arc length variable along $\mathcal{U}$, and $\varphi \in [0, \pi]$ measures the angle between $\partial/\partial r$ and the velocity vector (Fig. 3).

So the manifold over which we define our dynamical system is $\mathcal{M} := (0, +\infty) \times (0, \pi)$ (it will be clear in the sequel why it is a good idea to exclude the boundary of $\mathcal{M}$). The billiard flow defines on $\mathcal{M}$ a Poincaré return map T which, according to an easy classical result, preserves the measure $d\mu(r, \varphi) = \sin\varphi \, dr d\varphi$.

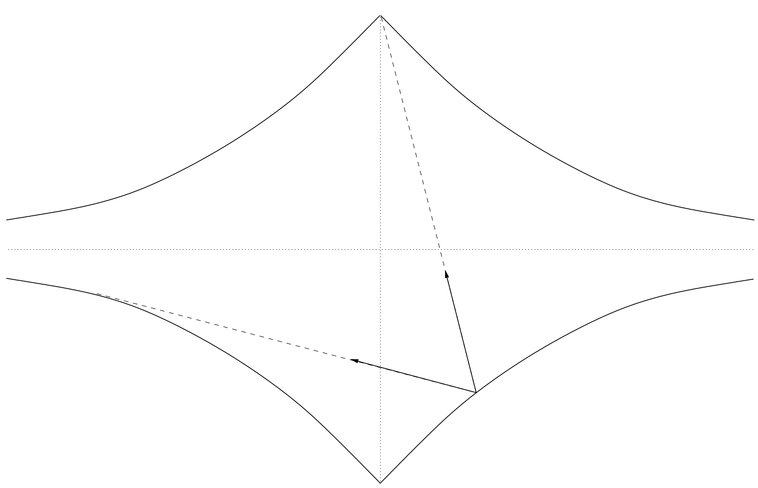

Figure 4: The geometrical meaning of the singularities.

We do not define T at those points of $\mathcal{M}$ that would end up in V or hit $\mathcal{U}$ tangentially (see Fig. 4); that is, we morally exclude the set "$T^{-1}\partial\mathcal{M}$". In fact, these points make up the discontinuity set of T. We will see later that they are often singularities as well—every time they correspond to a tangency. They are arranged in two curves, depicted in Fig. 5. $\mathcal{S}^{1+}$ corresponds to tangencies to ∂Q_4 in the third quadrant (i.e., in Q, tangencies to $\mathcal{U}$ after a rebound on the vertical side); this curve is a regular as f. As for $\mathcal{S}^{2+}$, its first part corresponds to line elements pointing into V (in Q, after a rebound on the horizontal side); as r increases, these become elements tangent to ∂Q. The border between these two behaviors is the only non-regular point of $\mathcal{S}^{2+}$. We denote $\mathcal{S}^+ := \mathcal{S}^{1+} \cup \mathcal{S}^{2+}$.

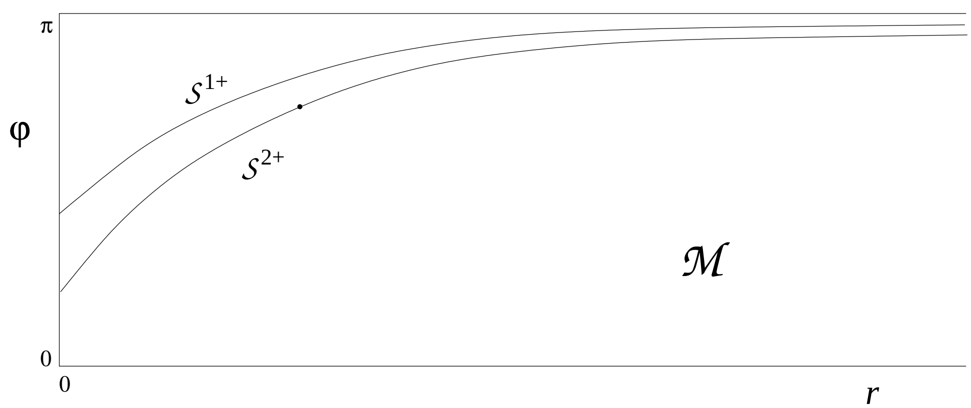

Figure 5: The singularity lines $\mathcal{S}^{1+}$ and $\mathcal{S}^{2+}$ in $\mathcal{M}$.

Analogously we name $\mathcal{S}^- := \mathcal{S}^{1-} \cup \mathcal{S}^{2-}$, where the $\mathcal{S}^{i-}$ are the singularity lines of T^{-1}, obtained from $\mathcal{S}^{i+}$ through the application of the time-reversal operator: $(r, \varphi) \mapsto (r, \pi - \varphi)$. Whenever the superscript $\pm$ is dropped we will always mean $\mathcal{S}^+$.

For all points in $\mathcal{M} \setminus \mathcal{S}^+$, the differential of T is known [LW, §14], and not too hard to compute, anyway. If $(r_1, \varphi_1) := z_1 := Tz = T(r, \varphi)$, then

$$DT_z = \begin{bmatrix} -\dfrac{\sin\varphi}{\sin\varphi_1} - \dfrac{k\tau}{\sin\varphi_1} & \dfrac{\tau}{\sin\varphi_1} \\[2ex] k + k_1\dfrac{\sin\varphi}{\sin\varphi_1} + \dfrac{kk_1\tau}{\sin\varphi_1} & -1 - \dfrac{k_1\tau}{\sin\varphi_1} \end{bmatrix}, \tag{3.5}$$

where $\tau = \tau(z)$ is the traveling time (equivalently, the distance in Q_4) between the two collision points. Also, k (resp. k_1) is the curvature of ∂Q in z (resp. z_1). We adopt the convention that in our semi-dispersing billiard the curvature is non-negative. Another geometric relation that can be simply checked is

$$\varphi_1 = \varphi \pm \beta, \tag{3.6}$$

β being the angle between the tangent lines to ∂Q_4 at z and z_1 (see Fig. 6). The sign depends on whether these two tangent lines meet to the right or to the left of the segment joining z and z_1.

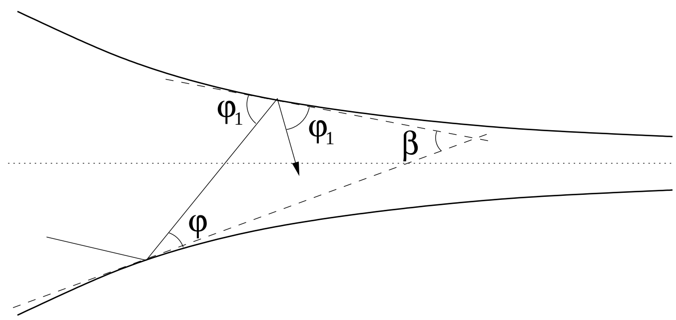

Figure 6: The relation between φ and φ_1.

We conclude this preliminary section by presenting what is known about the behavior of the trajectories in the vicinity of the cusp(s) or the corner. We return to the language of the billiard flow (as opposed to the Poincaré map), and recall the convention that the material point stops when it hits a vertex.

Proposition 3.2 *No billiard semiorbit falls into the vertex V or the "vertex at infinity" unless it shoots directly there. In other words, if $(x(t), y(t))$ is an orbit in Q,*

$$\lim_{t\to\pm\infty} (x(t), y(t)) = V \iff (x(t), y(t)) = V, \ \forall t \geq (\leq)\, t_0;$$
$$\lim_{t\to\pm\infty} x(t) = +\infty \iff y(t) = 0, \ \forall t;$$

the last condition specifying the trajectory that runs along the x-semiaxis.

PROOF. As far as V is concerned, if we have a corner there, the result comes from elementary considerations that are, anyway, mentioned in [BS]; one can even give an upper bound on the number of rebounds that the trajectory performs before moving away from V. In the case that V has a cusp, the assertion is contained in [V] and can be easily deduced by the "a gallon of water won't fit a pint-sized cusp" argument of [K].

For the non-compact cusp, the result was first derived by Leontovich in 1962 [L] (see also [Le1]). Q.E.D.

4 The cone field

We use here the construction of stable and unstable directions for dispersing billiards, as originally developed by Sinai in [S2, §2] (found in clean form, e.g., in [G, §2]), although we re-express those results in the more modern language of invariant cones.

On $\mathrm{T}\mathcal{M}$, the tangent bundle of $\mathcal{M}$, let us define

$$\mathcal{C}^+(z) := \mathcal{C}(z) := \{(dr, d\varphi) \in \mathrm{T}_z\mathcal{M} \mid dr \cdot d\varphi \leq 0\}\,; \tag{4.1}$$

that is, the second and fourth quadrant of $\mathrm{T}_z\mathcal{M}$, using the natural basis $\{\partial/\partial r, \partial/\partial\varphi\}_z$. This endows $\mathrm{T}\mathcal{M}$ with a continuous bundle of Lagrangian cones, sometimes referred to as *unstable cones*. The alternate signs of DT_z in (3.5) show that $DT_z\mathcal{C}(z) \subset int\mathcal{C}(Tz)$, where $int\mathcal{C}$ denotes the interior of $\mathcal{C}$ *together with 0*. We say that the above cones are *strictly invariant* under the action of T [W, LW].

Remark 4.1 This definition of the cone field does not correspond to the one devised by Wojtkowski in [W] for dispersing billiards. Rather, it is the transliteration of the long-known properties of monotonic curves in $\mathcal{M}$ under the action of $T^{\pm1}$ [S2, Cor. 2.2]. These cones could be rightfully called Sinai cones.

Later on, we will also need a cone bundle for T^{-1}. Since the time-reversed map can be treated in the same way as T, its cones $\mathcal{C}^-(z)$ will be defined analogously, as the first and third quadrant of $\mathrm{T}_z\mathcal{M}$.

Remark 4.2 The fact that the two cone bundles $\mathcal{C}^+$ and $\mathcal{C}^-$ are (almost) complementary to each other is purely accidental. It is true that when the dynamical system is Hamiltonian (more precisely, when it preserves a symplectic form whose volume element is absolutely continuous w.r.t. the standard volume element, and viceversa), as in our case, defining $\mathcal{C}^-$ as the closure of the complement of $\mathcal{C}^+$ still leads to essentially all the sought results [LW]. But, when possible, it is more convenient to define each cone bundle separately, based on the map that makes it invariant.

Denote $\mathcal{S}_n^- := \bigcup_{i=0}^{n-1} T^i\mathcal{S}^-$. For $(r_{-n}, \varphi_{-n}) := z_{-n} := T^{-n}z = T^{-n}(r,\varphi)$, we define the n-th nested cone as

$$\mathcal{C}_n(z) := DT^n_{z_{-n}}\mathcal{C}(z_{-n}) = \left\{ (dr, d\varphi) \in \mathrm{T}_z\mathcal{M} \;\middle|\; a_n \leq \frac{1}{\sin\varphi}\frac{d\varphi}{dr} \leq b_n \right\}, \tag{4.2}$$

It can be computed from (3.5) and (4.1) that

$$a_n = -\frac{k}{\sin\varphi} + \cfrac{1}{-\tau_{-1} + \cfrac{1}{-\cfrac{2k_{-1}}{\sin\varphi_{-1}} + \cfrac{1}{\ddots\, \cfrac{1}{-\tau_{-n}}}}} \tag{4.3}$$

and

$$b_n = -\frac{k}{\sin\varphi} + \cfrac{1}{-\tau_{-1} + \cfrac{1}{-\cfrac{2k_{-1}}{\sin\varphi_{-1}} + \cfrac{1}{\ddots \cfrac{1}{-\tau_{-n} + \cfrac{1}{-\cfrac{2k_{-n}}{\sin\varphi_{-n}}}}}}}. \tag{4.4}$$

Here τ_{-i} and k_{-i} are respectively $\tau(z_{-n})$ and $k(z_{-n})$, for $i = 1, 2, ..., n$ (see also [G, §2]). Hence a_n and b_n are the left and right approximants of the continued fraction defined by (4.3)-(4.4). The next result shows that this continued fraction converges.

Proposition 4.3 *For all $z \notin \mathcal{S}^-_\infty := \bigcup_{i=0}^\infty T^i \mathcal{S}^-$,*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n =: \chi^u(z).$$

Hence, as $n \to \infty$, $\mathcal{C}_n(z)$ converges, in the sense of decreasing sets, to a subspace $E^u(z) \subset \mathrm{T}_z\mathcal{M}$. We call it the unstable subspace. This is a line of slope $\chi^u(z)\sin\varphi$ w.r.t. the basis $\{\partial/\partial r, \partial/\partial\varphi\}_z$.

Proof. Since the terms of the continued fraction (4.3)-(4.4) are negative, classical results [S2, §2] show that a necessary and sufficient condition for convergence is

$$\sum_{i=-1}^{-\infty} \left(\tau_i + \frac{2k_i}{\sin\varphi_i} \right) = +\infty. \tag{4.5}$$

Assuming the contrary, τ_i would tend to zero. This would imply that the backward orbit of z zig-zags into V or into the cusp at infinity, in contradiction with Proposition 3.2. Q.E.D.

It will be very important in the remainder to estimate the expansion of a vector in $(dr, d\varphi) \in \mathcal{C}(z)$. From (3.5), using the notation previously introduced:

$$(dr_1)^2 \;\geq\; \left(\frac{\sin\varphi}{\sin\varphi_1}\right)^2 \left(1 + \frac{k\tau}{\sin\varphi}\right)^2 dr^2; \tag{4.6}$$

$$(d\varphi_1)^2 \;\geq\; \left(1 + \frac{k_1\tau}{\sin\varphi_1}\right)^2 d\varphi^2. \tag{4.7}$$

Sometimes the pseudo-metric $(dr\sin\varphi)^2$ is called *Z-metric* [KSSz]; it has the property that it is strictly increasing for vectors in the unstable cone.

5 Neighborhoods of the singularity lines

The problem with the direction field defined above is that it is hard to integrate, due to its scanty continuity properties. Morally its integral lines are the unstable manifolds, curves that shrink when T^{-1} is iterated many times (see Definition 6.1 for the precise details).

The techniques we deploy here to construct these objects stem from the original idea of Sinai, which relies upon measure estimates of the tubular neighborhoods of the singularity set. However, a big complication arises: since the $\mathcal{S}^{i\pm}$ are curves of infinite length, and are nicely embedded in $\mathcal{M}$, every ε-neighborhood has infinite measure. At least if one uses the ordinary Riemannian distance. On the other hand, in all of the arguments, it suffices to measure distances along the unstable direction. To this purpose let us define the *unstable distance* $d^u(z,w)$ between two points z and w as the infimum length of differentiable curves $t \mapsto \gamma(t)$ joining z with w, and such that $d\gamma/dt(t) \in \mathcal{C}_1(\gamma(t))$ (check definition (4.2)). Every such curve is henceforth referred to as an *unstable curve*. Having introduced $d^u(z,w)$, the unstable distance between a point and a set is defined in the usual way.

In order to state the central result of this section, let us introduce some notation. Denote $\mathcal{S}^{0+} := \mathbb{R}_0^+ \times \{\pi\}$ and $\mathcal{S}^{0-} := \mathbb{R}_0^+ \times \{0\}$. Also, for $A \subset \mathcal{M}$ and $\varepsilon > 0$,

$$A_{[\varepsilon]} := \{ z \in \mathcal{M} \mid d^u(z,A) \le \varepsilon \} . \tag{5.1}$$

Theorem 5.1 *Let* Leb *denote the Lebesgue measure on* $\mathcal{M}$*, and assume* (A1)-(A4). *For* $i = 0,1,2$*,* $\mu(\mathcal{S}^{i\pm}_{[\varepsilon]}) \le \mathrm{Leb}(\mathcal{S}^{i\pm}_{[\varepsilon]}) \ll \varepsilon$*, as* $\varepsilon \to 0^+$*.*

Proof. We prove the statement for $\mathcal{S}^{2+}$. Later we will see how to adapt the proof to neighborhoods of the other curves.

Let γ be an unstable curve, shorter than ε, having one endpoint (say γ_0) on $\mathcal{S}^2$. Suppose for the moment that the curve lies beneath $\mathcal{S}^2$. γ has negative slope, thus can be reparameterized to be the graph of a function $\varphi : [r_0, r_0 + \delta] \longrightarrow (0,\pi)$, with $\varphi(r_0) = \varphi_0$ and $(r_0, \varphi_0) = \gamma_0$. (The reader will forgive the abuse of notation regarding φ.) It follows that $\delta \le \varepsilon$. By definition of unstable curve, the tangent vector to γ at z belongs to $\mathcal{C}_1(z)$. Hence, applying (4.3) with $n = 1$ gives

$$\frac{d\varphi}{dr} \ge -k(r) - \frac{\sin\varphi}{\tau_{-1}(r,\varphi)}; \qquad \varphi(r_0) = \varphi_0. \tag{5.2}$$

We consider this differential inequality in $U := [r_0, r_0 + \varepsilon_0] \times [\varphi_0 - b, \varphi_0]$, ε_0 to be fixed later, and b an unimportant sufficiently large number. The situation is shown in Fig. 7.

Define $a(r,\varphi) := \tau_1(r,\varphi)\sin\varphi$. This is the length of the "almost vertical" segment depicted in Fig. 8. Inequality (5.2) becomes

$$\frac{d\varphi}{dr} \ge -k(r) - \frac{\sin^2\varphi}{a(r,\varphi)}. \tag{5.3}$$

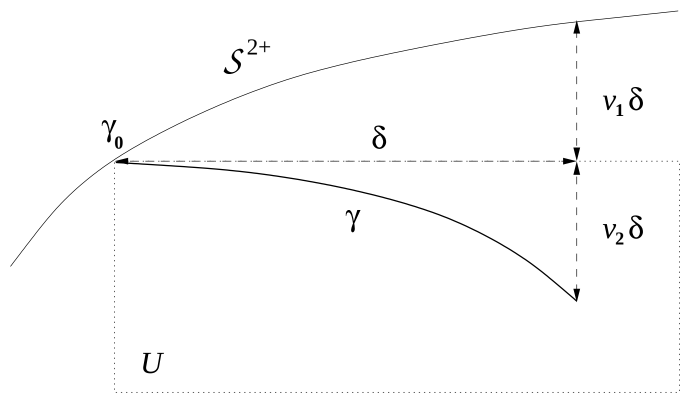

Figure 7: The unstable curve γ.

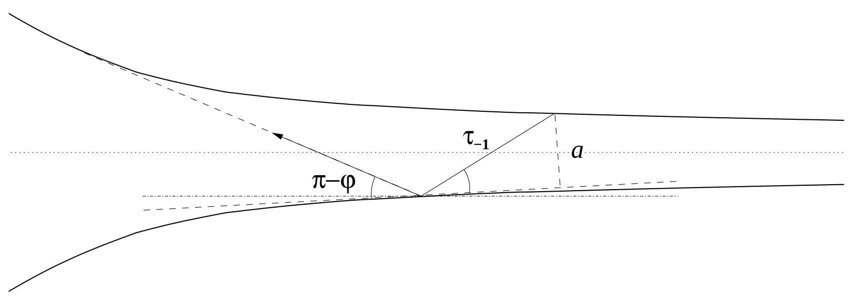

Figure 8: The definition of a, in the case $(r, \varphi) \in \mathcal{S}^2$.

We want to replace the r.h.s. of the above with a simpler bound, in order to solve the differential inequality. First of all, we substitute $k(r)$ with $k_M := k(\bar{r}) := \max_{[r_0, r_0+b]} k$. Then we notice that $a(r, \varphi)$ is decreasing in φ, hence its minimum on U is achieved at some point $(\hat{r}, \varphi_0) \in \partial U$: we denote it by a_m and use it to turn (5.3) into

$$\frac{d\varphi}{dr} \geq -k_M - \frac{\sin^2 \varphi}{a_m} \geq -k_M - const\, \frac{(\pi - \varphi)^2}{a_m} \tag{5.4}$$

in U. Introducing $\widetilde{\varphi}(r) := \pi - \varphi(r + r_0)$ and absorbing *const* into a_m (with no damage for the proof) we finally get the differential inequality

$$\frac{d\widetilde{\varphi}}{dr} \leq k_M + \frac{\widetilde{\varphi}^2}{a_m}; \qquad \widetilde{\varphi}(0) = \widetilde{\varphi}_0, \tag{5.5}$$

which is solved by

$$\widetilde{\varphi}(r) \leq \tan\left(\sqrt{\frac{k_M}{a_m}} r + \arctan\left(\frac{\widetilde{\varphi}_0}{\sqrt{k_M a_m}} \right) \right) \sqrt{k_M a_m}. \tag{5.6}$$

What we are interested in, here, is the maximum vertical distance between $\mathcal{S}^2$ and γ, when γ is defined in the domain $[r_0, r_0 + \delta]$. As illustrated in Fig. 7, this is the sum of two quantities, which we conveniently denote $v_1\delta$ and $v_2\delta$. The v_i's $(i = 1, 2)$ are actually functions of γ_0, the initial condition of γ, and of δ. For reasons that will be clear later, we need to show that the dependence on δ can be eliminated by two upper bounds that are integrable, as functions of $\gamma_0 = (r_0, \varphi_0) \in \mathcal{S}^2$. More precisely, this means as functions of x, where $x = x(r_0)$ is the x-coordinate of the point of ∂Q otherwise parameterized by r. In other words, $x(r)$ is the inverse of

$$r(x) := \int_0^x \sqrt{1 + (f'(t))^2}\, dt. \tag{5.7}$$

In the reminder, we will liberally switch from r_0 to x.

Now, as far as v_1 is concerned, we notice that $\mathcal{S}^2$ is the graph of an increasing function $g : \mathbb{R}^+ \longrightarrow (0, \pi)$. g has at most one point of non-regularity, and thus is eventually convex down. Therefore $v_1(x, \delta)\, \delta \le g'(x)\, \delta$, and g' is obviously integrable. As regards v_2, $\forall \delta \le \varepsilon_0$,

$$\begin{aligned} v_2(x, \delta) &\le \max_{[r_0, r_0+\delta]} \left| \frac{d\varphi}{dr} \right| = \frac{d\widetilde{\varphi}}{dr}(\varepsilon_0) \le \\ &\le \left(1 + \tan\left(\sqrt{\frac{k_M}{a_m}}\varepsilon_0 + \arctan\left(\frac{\widetilde{\varphi}_0}{\sqrt{k_M a_m}} \right) \right)^2 \right) k_M. \end{aligned} \tag{5.8}$$

The last inequality was obtained by plugging (5.6) into (5.5). We study the asymptotics of the quantities contained above. First of all, if $\bar{x} := x(\bar{r})$,

$$k_M = k(\bar{r}) = \frac{f''(\bar{x})}{(1 + (f'(\bar{x}))^2)^{3/2}} \sim f''(\bar{x}). \tag{5.9}$$

Now integrate (A4) to get $\forall \bar{x} \ge x$, $f''(\bar{x}) \le f''(x) e^{c(\bar{x}-x)}$, for some positive c. Since $\bar{r} - r_0 \le \varepsilon_0$, then $\bar{x} - x < \varepsilon_0$. Therefore $f''(\bar{x}) \sim f''(x)$, i.e.,

$$k_M \sim f''(x) \tag{5.10}$$

Moving on, it is clear from Fig. 8 that

$$\widetilde{\varphi}_0 = \pi - \varphi_0 = \arctan|f'(x)| + \arctan|f'(x_t)|. \tag{5.11}$$

Therefore

$$\widetilde{\varphi}_0 \sim |f'(x)| + |f'(x_t)| \sim |f'(x)|, \tag{5.12}$$

the last relation coming from (A2).

The case of $a_m = a(\hat{r}, \varphi_0)$ is a little more involved. We have already established that a decreases when φ increases, hence

$$a_m > \tau_{-1}(\hat{r}, \hat{\varphi}) \sin\hat{\varphi}, \tag{5.13}$$

with $\hat{\varphi}$ such that $(\hat{r}, \hat{\varphi}) \in \mathcal{S}^2$. Define $\hat{x} := x(\hat{r})$ and $\hat{x}_t := x_t(\hat{x})$.

$$\sin \hat{\varphi} \sim |f'(\hat{x})| + |f'(\hat{x}_t)| \sim |f'(\hat{x})| + 2|f'(\hat{x}_t)| \sim \arctan |f'(\hat{x})| + 2 \arctan |f'(\hat{x}_t), \tag{5.14}$$

where the first relation is the analogue of (5.12) and the second comes from (A2). Looking back at Fig. 8, we see that the rightmost term in (5.14) is the slope of the segment of trajectory from $(\hat{r}, \hat{\varphi}) \in \mathcal{S}^2$ to its backward image, a point that we denote by (r_{-1}, φ_{-1}). The length of said segment is $\tau_{-1}(\hat{r}, \hat{\varphi})$. Therefore, from (5.13), if we call $x_{-1} := x(r_{-1})$,

$$a_m \gg f(\hat{x}) + f(x_{-1}) > f(\hat{x}). \tag{5.15}$$

In analogy with (5.9)-(5.10), we integrate the second inequality of (3.2) to obtain an exponential estimate for f, this time from below. $\forall \hat{x} \geq x$, $f(\hat{x}) \geq f(x) e^{-c(\hat{x}-x)}$. Therefore (5.15) implies that $a_m \gg f(x)$. On the other hand, considering the case when φ is approximately $\pi/2$, it is evident that $a_m < 2f(x)$. Hence,

$$a_m \sim f(x). \tag{5.16}$$

Armed with (5.10), (5.12) and (5.16), we can now consider (5.8). The argument of the arctan is bounded by (A3), and so the arctan is less than some $\lambda < \pi/2$. Also, from (3.2), $f''(x) \ll f(x)$, thus $k_M/a_m \ll 1$. Let us fix ε_0 so small that $\sqrt{k_M/a_m}\, \varepsilon_0 < \pi/2 - \lambda$. In this way we obtain that $v_2(x, \delta) \ll k_M \sim f''(x)$, which is integrable.

We are finally ready to use the integrability of v_1 and v_2. Denote by $\mathcal{S}'_{[\varepsilon]}$ the portion of $\mathcal{S}^2_{[\varepsilon]}$ that lies beneath $\mathcal{S}^2$. For a given $r > 0$ let us estimate the thickness of $\mathcal{S}'_{[\varepsilon]}$ at abscissa r. Fig. 7 tells us that this is $\delta[v_1(r-\delta, \delta) + v_2(r-\delta, \delta)]$, where $\delta < \varepsilon$ is the length of the support of the unstable curve γ that connects (r, φ) in the lower part of $\partial \mathcal{S}'_{[\varepsilon]}$ to (r_0, φ_0) in $\mathcal{S}^2$ (ε is the length of γ and $\delta = r - r_0$); δ depends on r. But it is a clear fact that $v_i(r, \delta)$ is decreasing in r and increasing in δ. Therefore, by the above estimates,

$$\begin{aligned} \mathrm{Leb}(\mathcal{S}'_{[\varepsilon]}) &< \int_0^\infty \varepsilon [v_1(r-\varepsilon, \varepsilon) + v_2(r-\varepsilon, \varepsilon)]\, dr \leq \\ &\leq \mathit{const} \cdot \varepsilon \int_0^\infty [g'(x-\varepsilon) + f''(x-\varepsilon)]\, dx \leq \mathit{const} \cdot \varepsilon, \end{aligned} \tag{5.17}$$

since, by (5.7), $dr/dx \to 1$, as $x \to +\infty$. For the sake of rigor, let us remark that for $x \in [-\varepsilon_0, 0)$ we have defined $v_i(x, \varepsilon), g(x), f(x)$ in an arbitrary suitable way.

We now must prove the same result for the part of $\mathcal{S}^2_{[\varepsilon]}$ that stands above $\mathcal{S}^2$. This will be even easier. In fact, differential inequality (5.2) continues to hold. Let us consider it in $U := [r_0 - \varepsilon_0, r_0] \times [\varphi_0, \varphi_0 + b]$. We simplify it by introducing the maximum of k there, $k_M := k(\bar{r})$ and the minimum of τ_{-1}, $\tau_m := \tau_{-1}(\hat{r}, \varphi_0)$, for some $\bar{r}, \hat{r} \in [r_0 - \varepsilon_0, r_0]$. This is so because the latter function is increasing in φ.

Performing again the change of coordinate $\widetilde{\varphi}(r) := \pi - \varphi(r + r_0)$, the inequality is turned into

$$\frac{d\widetilde{\varphi}}{dr} \le k_M + \frac{\widetilde{\varphi}}{\tau_m}; \qquad \widetilde{\varphi}(0) = \widetilde{\varphi}_0. \tag{5.18}$$

We do not even need to solve it. In fact, since $d\widetilde{\varphi}/dr > 0$ (because the tangent to γ always belongs to the unstable cone), it is evident that

$$\max_{[-\delta,0]} \frac{d\widetilde{\varphi}}{dr} = \frac{d\widetilde{\varphi}}{dr}(0) = k_M + \frac{\widetilde{\varphi}_0}{\tau_m} =: v_2(x, \delta), \tag{5.19}$$

which is the analogue of (5.8). Concerning the asymptotics of these quantities, we see that $k_M \sim f''(x)$, using another exponential estimate. Also, $\widetilde{\varphi}_0 \sim |f'(x)|$, by (5.12). As to τ_m, one has that $\tau_m > \tau_{-1}(\hat{r}, \hat{\varphi})$, with $(\hat{r}, \hat{\varphi}) \in \mathcal{S}^2$. Then $a' := \tau_{-1}(\hat{r}, \hat{\varphi}) \sin\hat{\varphi}$ is completely similar to a_m in the previous case. We can use (5.14)-(5.16) to show that $\widetilde{\varphi_0}/\tau_m \sim \widetilde{\varphi}_0^2/a' \ll k_M$. Therefore (5.19) implies that $v_2(x, \delta) \ll f''(x)$, which is of course integrable.

The integrability of v_1 is straightforward, since the convexity of g implies $v_1(x, \delta)\, \delta \le g'(x - \varepsilon_0)\, \delta$.

This completes the proof of the result for the set $F^{2+}_{[\varepsilon]}$. Proving the corresponding statement for all the other neighborhoods is now just a corollary of the above. Take for instance $F^{1+}_{[\varepsilon]}$. The r.h.s. of (5.2) is decreasing in φ (in other words the cones shrink as we approach the upper boundary vertically). Hence the solutions will have smaller slope (in absolute value) and v_2 (and v_1 too, for that matter) will be smaller, making (5.17) still hold true.

The case of $F^{2-}_{[\varepsilon]}$ is also "overestimated" by the above computations. In fact, passing from a neighborhood of $\mathcal{S}^{2+}$ to its symmetrical $\mathcal{S}^{2-}$, we observe that $\sin(\pi - \varphi) = \sin\varphi$ and $\tau_{-1}(r, \pi - \varphi) = \tau(r, \varphi) > \tau_{-1}(r, \varphi)$, for $\varphi > \pi/2$ and r large. Once again, the r.h.s. of (5.2) becomes smaller and the previous estimate of v_2 largely suffices. Furthermore, the maximum vertical distance between γ and $\mathcal{S}^{2-}$, in this case, is $|v_2 - v_1|$, which makes the bound even more redundant. The other cases are now clear. Q.E.D.

Corollary 5.2 *There exists a measure π defined on the singularity set, of finite mass, such that for every* closed $A \subseteq \mathcal{S}^+ \cup \mathcal{S}^-$,

$$\mathrm{Leb}(A_{[\varepsilon]}) \le \pi(A)\, \varepsilon + o(\varepsilon), \qquad \textit{as } \varepsilon \to 0^+.$$

This measure is absolutely continuous w.r.t. the one-dimensional Lebesgue measure on the singularity set.

Proof. This assertion is what Euclid would have called a *porism*, in the sense that it is derived from the proof of the previous theorem, rather than its statement. At any rate, the argument is rather obvious. Consider, without loss of generality, a closed arc $A \subseteq \mathcal{S}^{2+}$. Say that $A := \{(r, g(r)) \mid b \le r \le c\}$, with the same g as in proof of Theorem 5.1. Define

$$\pi(A) := \int_b^c [g'(x) + f''(x)] \, \frac{dr}{dx} \, dx. \tag{5.20}$$

For simplicity, we estimate only the part of $A_{[\varepsilon]}$ that lies below $\mathcal{S}^{2+}$; we call it $A'_{[\varepsilon]}$. From the line of reasoning that led to (5.17) it is clear that. for small ε,

$$\mathrm{Leb}(A'_{[\varepsilon]}) \le \pi(A)\, \varepsilon + o(\varepsilon). \tag{5.21}$$

The remainder term is essentially $\mathrm{Leb}(\{c\}_{[\varepsilon]})$ and thus depends on A. Q.E.D.

6 Local stable and unstable manifolds

The existence and absolute continuity of the invariant manifolds for a hyperbolic system is given by *Pesin's theory*. It was first developed by Pesin for smooth systems [P] and has been later adapted and generalized in many ways (see, e.g., [PS] and its references).

In the case of billiards, the dynamics is complicated by the presence of the singularities, which generate a whole new mechanism for two nearby points to have far-away forward (or backward) images. Extending Pesin's theory to non-smooth hyperbolic systems was the goal of Katok and Strelcyn's 1986 major work [KS]. Among their results, the one that concerns us here is the existence and absolute continuity of local stable and unstable manifolds (LSUMs) at almost every point, for a very general class of systems; a fine-tuned theorem that requires only reasonable conditions on the non-regularity set. This is the main, irreplaceable, ingredient in the proof of local ergodicity using Hopf's idea.

The problem is that this theorem, like most in the literature, assumes the invariant measure to be finite, which is not the case of $\mathcal{M}$! Practically speaking, this means that we have to work out our own results for the hyperbolic structure of $(\mathcal{M}, T, \mu)$. This section is devoted to the existence of LSUMs and the following section to their absolute continuity.

We start by giving a definition of LSUM which is sufficient for our purposes. In many instances, one can postulate (and then prove) much more about these objects [KH, §6].

Definition 6.1 *Given a point $z \in \mathcal{M} \setminus \mathcal{S}^{+(-)}$, we define a* local (un)stable manifold $W^{s(u)}$ for T at *z to be a C^1 topological disk containing z, and such that:*

(a) The tangent space to $W^{s(u)}$, at every point, is included in the (un)stable cone of every order, i.e., $\forall w \in W^{s(u)}$, $\mathrm{T}_w W^{s(u)} \subset \bigcap_n \mathcal{C}_n(w)$, and it has the maximal dimension there (1 in our case);

(b) $\forall w \in W^{s(u)}$, $|T^n w - T^n z| \to 0$, *as* $n \to +\infty(-\infty)$*;*

(c) If $W_0^{s(u)}$ *is another such manifold, then so is* $W^{s(u)} \cap W_0^{s(u)}$.

If the convergence in (b) is exponential, we say that $W^{s(u)}$ *is* exponentially *(un)stable.*

The above conditions ensure that, if W^u is a LUM at z, then a subdisk of $T^{-1}W^u$ is a LUM at $T^{-1}z$. In fact, by assumption, T^{-1} is smooth in a neighborhood of z and so at least for a sufficiently small neighborhood U of $T^{-1}z$, $T^{-1}W^u \cap U$ is a smooth topological disk of the right dimension. Furthermore, for each w in this disk and each n, $\mathrm{T}_w T^{-1} W^u \subset \mathcal{C}_n(w)$, since $\mathrm{T}_{Tw} W^u \subset \mathcal{C}_{n+1}(Tw)$. In the same way, *(b)* and *(c)* are easily seen to hold for $T^{-1}W^u \cap U$.

Property *(c)* guarantees "uniqueness", in a certain sense. For this reason, we allow the (customary) abuse of notation and call an (un)stable manifold at z, *the* (un)stable manifold at z, denoting it by $W^{s(u)}(z)$.

We move on to the theorem of existence of LSUMs. As we have mentioned later, the technical problem is that $\mu(\mathcal{M}) = \infty$. On the other hand, the singularity lines, responsible for "cutting" the invariant manifolds, are contained in a finite-measure set.

Suppose we choose a connected set $\mathcal{M}_0$ that includes the singularities (see, for instance, Fig. 9 later on). It is not hard to realize that the region of $\mathcal{M}$ above $\mathcal{S}^{2+}$ (which we will call $\mathcal{M}_3$ at the end of Section 8) is the set of those line elements whose billiard trajectory hits the left part of the vertical side of Q. By Proposition 3.2 every semi-trajectory that is well defined (in the future or in the past) does this; therefore almost every T-orbit intersects T_0. So one might think of working with T_0, the map induced by T on $\mathcal{M}_0$, prove the existence of LSUMs for this subsystem and then "pull them backward" to the whole of $\mathcal{M}$ by using T, since there are no singularities there.

But this is not so simple, since T_0 has a much bigger discontinuity set than T. In fact, the set $\mathcal{M}_0 \cap T^{-n}\partial\mathcal{M}_0$ represent the borderline between points that take n T-iterations to come back to $\mathcal{M}_0$, and points that take $n+1$ iterations. Therefore a new discontinuity is induced there.

We will use a strategy, however, that relies on the idea above, and the additional fact that these new discontinuities do not really have to do with the "physics" of the system. The new boundary was put there arbitrarily and, somehow, can be deformed whenever is convenient. For this reason, $\partial\mathcal{M}_0$ might deserve the name of *fuzzy boundary.*

These concepts will be made rigorous in the next theorem. Recall the notation (5.1).

Theorem 6.2 *Let* $\mathcal{M}$ *be a Riemannian manifold, embedded in* $\mathbb{R}^N$*, and* $(\mathcal{M}, T, \mu)$ *an invertible, recurrent, dynamical system on it. Denote the discontinuity set of* T *by* $\mathcal{S}$*. Assume that for some* α*, the following holds:*

(a) $\mu\left((\mathcal{S}\cup\partial\mathcal{M})_{[\varepsilon]}\right) \ll \varepsilon^{\alpha}$, *for* $\varepsilon \to 0^{+}$.

(b) *These exists a continuous, invariant cone bundle* $\mathcal{C}$, *such that* $\forall z \in \mathcal{M}$, $\bigcap_n \mathcal{C}_n(z) = E^u(z)$, *a subspace of* $\mathrm{T}_z\mathcal{M}$. *(*$\mathcal{C}_n(z)$ *is defined as in* (4.2)*.)*

(c) *There exists an increasing norm* $\|\cdot\|$ *for cone vectors, that is,* $\forall z \in \mathcal{M}\setminus\mathcal{S}$, $\exists\kappa(z) > 1$ *such that* $\forall v \in \mathcal{C}(z)$, $\|DT_z v\|_{Tz} \geq \kappa(z)\,\|v\|_z$.

(d) *Let us denote by* H *the set where the* expansion factor κ *is not bounded away from 1, i.e.,* $H := \{z \mid \exists z_n \to z, \kappa(z_n) \to 1\}$. *Then* $\mu\left(H_{[\varepsilon]}\right) \ll \varepsilon^{\alpha}$, *when* $\varepsilon \to 0^{+}$.

(e) *Denoted by* $|\cdot|$ *the Riemannian norm on* $\mathrm{T}\mathcal{M}$, *and taken two functions* $0 < p \leq q$ *such that* $\forall z \in \mathcal{M}\setminus\mathcal{S}$,

$$p(z)\,\|\cdot\|_z \leq |\cdot|_z \leq q(z)\,\|\cdot\|_z,$$

then p *is locally bounded below, and* $q(z) \ll [d^u(z,\mathcal{S}\cup\partial\mathcal{M})]^{-\beta}$.

Then, for μ*-a.e.* z, *the local unstable manifold* $W^u(z)$ *exists.*

Furthermore, let us take a $\mathcal{M}_0 \subseteq \mathcal{M}$, $\mu(\mathcal{M}_0) < \infty$, *such that* $(\mathcal{S}\cup\partial\mathcal{M})_{[\varepsilon_0]} \subseteq \mathcal{M}_0$, *for some* $\varepsilon_0 > 0$. *Then the* $W^u(z)$ *are exponentially expanding w.r.t. the return times to* $\mathcal{M}_0$. *This means that, given a* $z \in \mathcal{M}_0$ *for which* $W^u(z)$ *exists, and denoted by* $\{-n_k\}_{k\in\mathbb{N}}$ *the sequence of its return times in the past, then* $\exists C, \lambda > 0$ *such that*

$$\forall w \in W^{s(u)}, \quad |T^{-n_k}w - T^{-n_k}z| \leq C\,e^{-\lambda k}, \quad \text{as } k \to \infty.$$

Remark 6.3 The statement of the theorem looks rather cumbersome because we have tried to present it in a certain generality. Since, as we have recalled, the literature is not very generous in terms of results for infinite-measure dynamical systems, the hope is that this theorem finds application beyond the scope of the present paper.

At any rate, in concrete examples one may expect some of the conditions to be trivially satisfied. For instance, in our case, verifying *(a)-(d)* will be immediate, as we will see later. However, it turns out that hypothesis *(e)* does not hold! Indeed we must substitute it with a different set of requirements. Nevertheless, for the sake of clarity—and in the spirit of the above paragraph—we have decided to state Theorem 6.2 in the given form. This contains all the relevant ideas, and most technical points. Theorem 6.5 will work out the necessary adaptations for use on our billiard.

Proof of Theorem 6.2. Without loss of generality, assume that also $H_{[\varepsilon_0]} \subseteq \mathcal{M}_0$. As anticipated, we denote by T_0 the return map onto $\mathcal{M}_0$. This is well defined due to the recurrence. Let us define

$$\begin{aligned} A_k &:= \left\{ z \in \mathcal{M}_0 \;\middle|\; d^u(T_0^{-k} z, H \cup \mathcal{S} \cup \partial\mathcal{M}) < \frac{1}{(k+1)^{2/\alpha}} \right\} = \\ &= T_0^k \left((H \cup \mathcal{S} \cup \partial\mathcal{M})_{[(k+1)^{-2/\alpha}]} \right), \end{aligned} \tag{6.1}$$

having used notation (5.1). Hypotheses *(a), (d)* guarantee that $\mu(A_k) \ll (k+1)^{-2}$. Therefore, denoting

$$A := \{ \{A_k\} \text{ infinitely often } \} := \bigcap_{m \in \mathbb{N}} \bigcup_{k \geq m} A_m, \tag{6.2}$$

we have that $\mu(A) = 0$ by the Borel-Cantelli Lemma. Via an easy argument, then, $\forall z \in \mathcal{M}_0 \setminus A$, $\exists C_1(z) > 0$ such that, $\forall k \in \mathbb{N}$,

$$d^u(T_0^{-k} z, H \cup \mathcal{S} \cup \partial\mathcal{M}) \geq \frac{2C_1}{(k+1)^{2/\alpha}}. \tag{6.3}$$

By hypothesis, H is a null-measure set. On its complement we define

$$\psi(z) := \inf_{w \in B^u(z, d^u(z,H)/2)} \log \kappa(w) > 0, \tag{6.4}$$

It is crucial to notice here that the ball in the above definition is a d^u-ball of $\mathcal{M}$ and not $\mathcal{M}_0$! In other words, we seek the infimum of $\log\kappa$ in a neighborhood of z that can exceed $\mathcal{M}_0$.

Applying Lemma A.1 in the Appendix to the dynamical system $(\mathcal{M}_0, T_0, \mu)$, gives that there is a $B \subset \mathcal{M}_0$, $\mu(B) = 0$, such that, if $z \in \mathcal{M}_0 \setminus B$,

$$\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \psi(T_0^{-k} z) > 0. \tag{6.5}$$

Now let us fix a $z \in \mathcal{M}_0 \setminus (A \cup B)$, and let $\{-n_k\}$ be its sequence of past return times, as in the statement of the theorem. Since ψ is positive, (6.5) implies that $\exists \lambda > 0$ such that, $\forall m \in \mathbb{Z}^+$,

$$\sum_{k=1}^{m} \psi(T^{-n_k} z) \geq \lambda m. \tag{6.6}$$

A point w in a neighborhood of $T^{-n_m} z$ is "good" if, for all $j = 0, ..., m$,

$$T^{n_m - n_j} w \in B^u \left(T^{-n_j} z, \frac{C_1}{(j+1)^{2/\alpha}} \right). \tag{6.7}$$

Thus, using (6.3),

$$T^{n_m - n_j} w \in B^u \left(T^{-n_j} z, \frac{d^u(T^{-n_j} z, H \cup \mathcal{S} \cup \partial\mathcal{M})}{2} \right). \tag{6.8}$$

In particular, if w is good, the lower bound (6.4) holds generous for w and its first m forward images. The dynamics we are talking about here is $\{T^{n_m - n_j}\}$, the one generated by the return times of z. These are not necessarily the same return times as $T^{n_m} w$: in fact, the balls that appear in (6.8) may very well exceed $\mathcal{M}_0$. This implements the aforementioned idea of $\partial \mathcal{M}_0$ as a fuzzy boundary.

For w good and $v \in \mathcal{C}(w)$,

$$\|DT^{n_m} v\|_{T^{n_m} w} \geq \exp\left\{ \sum_{j=0}^{m-1} \psi(T^{n_m - n_j} w) \right\} \|v\|_w \geq e^{\lambda m} \, \|v\|_w, \tag{6.9}$$

because of (6.6). We naturally call good an unstable curve γ which contains $T^{-n_m} z$ and is made up of good points. If we denote by $\ell_{\|}$ the length of a curve in the $\| \cdot \|$ metric, then (6.9) implies that a good curve γ verifies

$$\ell_{\|}(T^{n_m} \gamma) \geq e^{\lambda m} \, \ell_{\|}(\gamma). \tag{6.10}$$

A sufficiently short (in the sense of $\ell_{\|}$) γ is obviously good. We claim that γ can be elongated in such a way that it remains good and also

$$\begin{array}{rcll} \ell_{\|}(T^{n_m - n_j} \gamma) & \leq & C_2 \, e^{-\lambda j}, & \qquad j = 1, ..., m; \\ \ell_{\|}(T^{n_m} \gamma) & = & C_2, & \end{array} \tag{6.11}$$

for some $C_2 = C_2(z)$ to be determined as follows: By (6.8) and (6.3),

$$d^u(T^{n_m - n_j} w, \mathcal{S} \cup \partial \mathcal{M}) \geq \frac{C_1}{(j+1)^{2/\alpha}}. \tag{6.12}$$

Therefore, by hypothesis *(e)*, there is a C_3 such that

$$q(T^{n_m - n_j} w) \leq C_3 \, (j+1)^{2\beta/\alpha}. \tag{6.13}$$

We use the above to define C_2 as any number that verifies, for all $k \in \mathbb{N}$,

$$C_2 \, C_3 \, (k+1)^{2\beta/\alpha} \, e^{-\lambda k} \leq \frac{C_1}{(k+1)^{2/\alpha}}. \tag{6.14}$$

We proceed to prove (6.11). Assume first that γ stays good as the elongation is performed. Violating (6.11) amounts to finding a j, $1 \leq j \leq m$, such that $\ell_{\|}(T^{n_m - n_j} \gamma) = C_2 \, e^{-\lambda j}$ and $\ell_{\|}(T^{n_m} \gamma) < C_2$, but this would contradict (6.10) with j replacing m and $T^{n_m - n_j} \gamma$ (also a good curve) replacing γ_j.

It remains to show that γ remains good until it reaches situation (6.11), i.e., as long as $\ell_{\|}(T^{n_m - n_j} \gamma) \leq C_2 \, e^{-\lambda j}$, $j = 0, ..., m$. From (6.13),

$$\ell(T^{n_m - n_j} \gamma) \leq C_3 \, (j+1)^{2\beta/\alpha} \, \ell_{\|}(T^{n_m - n_j} \gamma), \tag{6.15}$$

recalling that ℓ is the Riemannian length. Therefore, with the help of (6.14),

$$\ell(T^{n_m - n_j} \gamma) \leq C_2 \, C_3 \, (j+1)^{2\beta/\alpha} \, e^{-\lambda j} \leq \frac{C_1}{(j+1)^{2/\alpha}}. \tag{6.16}$$

Hence, by definition of d^u, all points of $T^{n_m - n_j} \gamma$ are as close to $T^{-n_j} z$ as (6.7) prescribes. This completes the proof of the claim.

Remark 6.4 Notice that, by (6.8), none of the $T^{n_m-n_j}\gamma$, for $j=0,...,m$, can be cut by $\mathcal{S}$. This is all the more true for the other iterates $T^{n_m-n}\gamma$, $n=0,...,n_m$, $n\neq n_j$. In fact, for $n_j<n<n_{j+1}$, $T^{-n}z\notin\mathcal{M}_0$ and, by construction, $T^{-n}z$ is even further away from $\mathcal{S}\cup\partial\mathcal{M}_0$ than $T^{-n_j}z$. Therefore, by *(e)*, (6.13), (6.15) hold with n in the place of n_j. Add that $\ell_\|(T^{n_m-n}\gamma)\le\ell_\|(T^{n_m-n_j}\gamma)$, and (6.16) holds too with n replacing n_j.

We have now worked out the most technical part of this proof: we have managed to reconduct to a situation in which certain curves are exponentially contracting (in the Riemannian metric) up to some time in the past, as seen in (6.16). The only peculiarity is that this rate of contraction is attained w.r.t. the return times of a point z to $\mathcal{M}_0$. Furthermore, these curves do not see the singularity lines. Pesin's key idea is precisely that disks made up of such curves are natural approximations for the unstable manifolds. So, apart from the fact that we use a different time scale, the reasoning will now become very standard in the context of Pesin's theory and thus the exposition will be a little less detailed.

Given $z\in\mathcal{M}\setminus(A\cup B)$ as above, let Δ_m be a smooth topological disk, centered in $T^{-n_m}z$, good, lying "inside" the cone bundle (i.e., $\forall w\in\Delta_m$, $\mathrm{T}_w\Delta_m\subset\mathcal{C}(w)$), and with the maximal dimension there. Call this dimension $\nu\in\mathbb{Z}^+$.

Although Δ_m needs to be small to be good, we can choose it large enough that $T^{n_m}\Delta_m$ is "macroscopic". More precisely, we require Δ_m to contain a topological disk of radius $C_2e^{-\lambda m}$ in the $\|\cdot\|$ metric. This means that

$$\inf\left\{\ell_\|(\gamma)\;\middle|\;\gamma \text{ smooth curve }\subset\Delta_m\text{ linking }T^{-n_m}z\text{ to }\partial\Delta_m\right\}\ge C_2\,e^{-\lambda m}. \tag{6.17}$$

By the first inequality of *(e)*, $\ell_\|(\,\cdot\,)\le C_4\,\ell(\,\cdot\,)$ in a neighborhood of z. This guarantees that there is a topological disk of Riemannian radius C_2/C_4 inside $T^{n_m}\Delta_m$, that we can name $\mathcal{B}_m$. In fact, assume the contrary. Then there exists a curve $\eta\subset T^{n_m}\Delta_m$, starting at z and reaching $\partial T^{n_m}\Delta_m$, such that $\ell(\eta)<C_2/C_4$, which gives $\ell_\|(\eta)<C_2$. Therefore $T^{-n_m}\eta$ is good; hence, by (6.10), $\ell_\|(T^{-n_m}\eta)<C_2e^{-\lambda m}$. But $T^{-n_m}\eta$ links $T^{-n_m}z$ to some $\partial\Delta_m$, and this contradicts (6.17).

In compliance with the plan we have anticipated, let us define:

$$W^u(z):=\lim_{m\to\infty}\mathcal{B}_m. \tag{6.18}$$

The limit here is intended in a certain C^1 Hausdorff distance. More precisely, if $\mathcal{B}$ and $\mathcal{B}'$ are two C^1 (closed) disks, then their distance is defined as

$$dist(\mathcal{B},\mathcal{B}'):=\max_{z\in\mathcal{B}}dist(z,\mathcal{B}')+\max_{w\in\mathcal{B}'}dist(w,\mathcal{B}), \tag{6.19}$$

where, with abuse of notation, we have denoted by the same symbol the distance between two sets and the distance between a point and a set. The latter is

$$dist(z,\mathcal{B}'):=\min_{w\in\mathcal{B}'}\{|z-w|+d_{G_\nu}(\mathrm{T}_z\mathcal{B},\mathrm{T}_w\mathcal{B}')\}. \tag{6.20}$$

The norm and the distance on r.h.s. above are those inherited by the embedding in $\mathbb{R}^N$. More specifically, d_{G_ν} is any distance in $G_\nu(\mathbb{R}^N)$, the space of ν-dimensional planes in $\mathbb{R}^N$. The distance defined by (6.19) is complete. We will use Lemma A.2 in the Appendix to prove that limit (6.18) exists.

For all $m \geq 0$, denote by K_m the compact neighborhood of z made up of points z' such that $T^{-n_m} z'$ is good. In other words, K_m is the intersection, for $j = 0, \cdots, m$ of the forward images of the balls in the r.h.s. of (6.7). Of course $K_{m+1} \subseteq K_m$. In particular, points of K_m stay away from $\mathcal{S}^-_{n_m} = \bigcup_{i=0}^{n_m - 1} \mathcal{S}^-$, so T^{-n_m} is a diffeomorphism between K_m and its image. This implies that $\mathcal{C}_{n_m}$ varies continuously on K_m. Moreover, if $g_m(z')$ denotes the size of $\mathcal{C}_{n_m}(z')$, i.e.,

$$g_m(z') := \max \left\{ d_{G_\nu}(X - Y) \mid X, Y \; \nu\text{-dim. subspace of } \mathcal{C}_{n_m}(z') \right\}, \tag{6.21}$$

then one has $g_m(z') \searrow 0$, as $m \to +\infty$, by condition *(b)*. At this point Lemma A.2 tells us that this convergence occur somehow "uniformly", although on shrinking sets. But this is sufficient to see that, for $j > m$ large enough, $dist(\mathcal{B}_m, \mathcal{B}_j) \leq \varepsilon$. In fact, both $T^{n_m} \Delta_m$ and $T^{n_j} \Delta_j$ are contained in K_m and their tangent spaces are uniformly close over the two disks; furthermore, since the disks have at least z in common, their points are also close.

The completeness of $dist$ proves that $\mathcal{B}_m$ has a limit. Moreover, one can see that the limit does not depend on the choice of Δ_m. In fact the above argument works as well if we replace Δ_j with some other good Δ'_j, so that also $dist(\mathcal{B}_m, \mathcal{B}'_j) \leq \varepsilon$, for the same m and j. This gives $W^u(z) = \lim_m \mathcal{B}_m = \lim_j \mathcal{B}'_j$.

For almost every $z \in \mathcal{M} \setminus \mathcal{M}_0$, one obviously defines $W^u(z)$ as $T^{-n} W^u(T^n z)$, if n is the smallest positive integer s.t. $T^n z \in \mathcal{M}_0$.

It remains to show that $W^u(z)$ verifies the axioms of Definition 6.1. *(a)* is just obvious by construction. *(c)* is more or less as direct: in fact, fixed $z \in \mathcal{M}_0$ for simplicity, and taken given another $W^u_0(z)$, one can construct Δ'_j simply by taking $T^{-n_j} W^u_0(z)$, and possibly by extending it in an arbitrary way, should it be smaller than the size prescribed by (6.17). By the above argument, re-applying T^{n_j} and taking the limit gives again $W^u(z)$. Lastly, estimating the middle term of (6.16) with some $C_5 \, e^{\lambda' j}$, for a certain $\lambda' < \lambda$, shows that, $\forall z' \in W^u(z)$, $|T^{-n_j}(z') - T^{-n_j}(z)| \leq C_5 \, e^{\lambda' j}$, since the distance in $\mathbb{R}^N$ is certainly less than or equal to the unstable distance on $\mathcal{M}$. This proves the last statement of Theorem 6.2. The fact that $|T^{-n}(z') - T^{-n}(z)|$ becomes small even for $n \neq n_j$ has been explained in Remark 6.4. This verifies Definition 6.1, *(b)*, whence the theorem. Q.E.D.

Let us check that hypotheses *(a)*-*(d)* of Theorem 6.2 hold for our non-compact billiard: *(a)* is true by Theorem 5.1. *(b)* holds by the results in Section 4, in particular Proposition 4.3. The increasing norm in *(c)* is

$$\|(dr, d\varphi)\|^2_{(r,\varphi)} := \sin^2 \varphi \, dr^2 + d\varphi^2, \tag{6.22}$$

see (4.6)-(4.7). Also, $H = \emptyset$, and so there is nothing to prove in *(d)*.

From (6.22), $p \equiv 1$, and $q(r, \varphi) = 1/\sin\varphi$. Therefore, as one can easily see, *(e)* fails to hold in our case. The next theorem circumvents this problem. Actually, the arguments will even resemble more closely those used for finite-measure dynamical systems. Except that, for infinite measure, the formulation is heavier than Theorem 6.2.

Theorem 6.5 *The assertions of Theorem 6.2 also hold if the estimate on q in (e) is replaced by the following:*

(f) *There exists a $\beta > 0$ such that $\int_{\mathcal{M}_0} q^\beta \, d\mu < \infty$.*

(g) *There exist $\varepsilon_0, C' > 0$ such that*

$$\sup_{w \in B^u(z, d^u(z, \mathcal{S} \cup \partial\mathcal{M})/2)} q(w) \leq C' \, q(z),$$

uniformly in z, every time $d^u(z, \mathcal{S} \cup \partial\mathcal{M}) \leq \varepsilon_0$.

(h) *Fixed a $z \in \mathcal{M}_0$, with $\{-n_j\}$ the sequence of its past return times to $\mathcal{M}_0$, then, for $n_j < n < n_{j+1}$, $q(T^{-n}z) \leq C'' q(T^{-n_j} z)$, with C'' not depending on z.*

Proof of Theorem 6.5. The main fact that we lose, if we give up *(e)* in Theorem 6.2, is (6.13). That is, we do not know whether, for w good, the ratio between the two metrics grows polynomially along the backword orbit of $T^{n_m} w$. When this happens, the growth is eventually tamed by the exponential contraction in the increasing norm.

We have to reconstruct this situation: By *(f)* and a suitable Chebychev-type inequality,

$$\mu\left(\left\{ z \in \mathcal{M}_0 \;\middle|\; q(z) > k^{2/\beta} \right\}\right) \ll k^{-2}. \tag{6.23}$$

In a way totally analogous to (6.1)-(6.3), one concludes that, for z outside a null-measure set, $\exists C_6 = C_6(z)$ such that

$$q(T_0^{-k} z) = q(T^{-n_k} z) \leq C_6 \, k^{2/\beta}. \tag{6.24}$$

Fixed any such z, fast-forward to (6.7)-(6.8): if w is good and $m \geq j \geq j_0$, for some $j_0 = j_0(\varepsilon_0)$, then

$$d^u(T^{n_m - n_j} w, T^{-n_j} z) \leq \min\left\{ \varepsilon_0, \frac{d^u(T^{-n_j} z, \mathcal{S} \cup \partial\mathcal{M})}{2} \right\}. \tag{6.25}$$

Hence

$$q(T^{n_m - n_j} w) \leq C_5 \, j^{2/\beta}. \tag{6.26}$$

In fact, for $j \geq j_0$, the above comes from (6.24) and *(g)*, with $C_5 = C' \, C_6$; the finitely many remaining values of j can be included by adjusting C_5. (6.26) replaces (6.13), which is what we wanted to do.

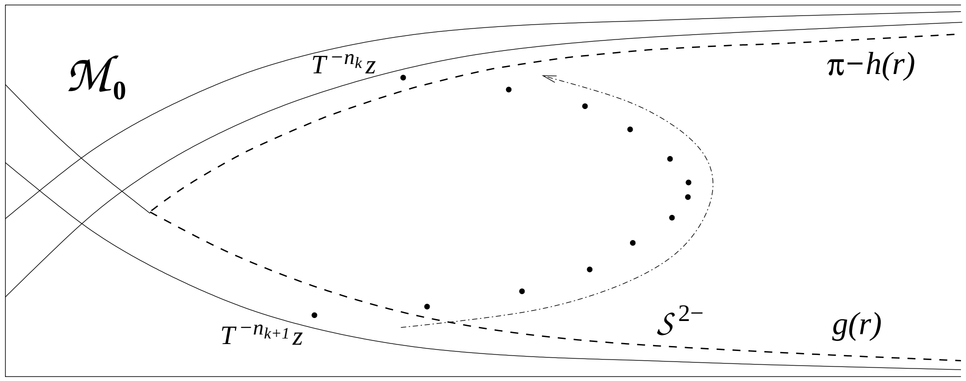

Figure 9: A suitable choice of $\mathcal{M}_0$.

The remaining hypothesis has to do with Remark 6.4. Without *(h)* it might happen that, for $n_j < n < n_{j+1}$, $T^{n_m - n}\gamma$ is so long, in the Riemannian length, that it can reach $\mathcal{S}$. Instead, by *(h)* and (6.26), (6.13) holds as well with n in the place of n_j, and so do (6.15)-(6.16). Q.E.D.

In order to use the above result, we have to define a suitable $\mathcal{M}_0$. Let us chose it like in Fig. 9. More precisely, we define it in an arbitrary way in a compact region of $\mathcal{M}$ and, for r (or x) large, we ask that $\partial\mathcal{M}_0 \cap int\mathcal{M}$ be composed of two curves. The lower curve is given by $\mathcal{S}^{2-}$. With the usual correspondence $r \longleftrightarrow x$—see (5.7)—this is the graph of the function

$$r \mapsto g(r) := \arctan|f'(x)| + \arctan|f'(x_t)| \sim |f'(x)|; \tag{6.27}$$

see (5.11)-(5.12). (Notice that we have already encountered this function in the proof of Theorem 5.1, although what we denoted g there was the graph of $\mathcal{S}^{2+}$, which is $\pi - g$ here.) As concerns the upper curve, this must lie below a certain neighborhood $\mathcal{S}^{2+}_{[\varepsilon_0]}$. From (5.8), (5.10) and (5.17), we know that the thickness of $\mathcal{S}^{2+}_{[\varepsilon_0]}$ is of the order of $f''(x)$. By (3.2) this is asymptotically bounded by $|f'(x)|$. Hence the sought curve can be chosen as the graph of a function $r \mapsto \pi - h(r)$, with $h(r) \sim g(r) \sim |f'(x)|$. Therefore *(f)* holds with $\beta = 1$ since, by definition

$$\int_{\mathcal{M}_0} q(r,\varphi)\, d\mu(r,\varphi) = \int_{\mathcal{M}_0} dr d\varphi = \mathrm{Leb}(\mathcal{M}_0) < \infty. \tag{6.28}$$

Verifying *(g)* will be a trifle more boring. First of all, we can forget about $\mathcal{S}$, since q only diverges at $\partial\mathcal{M}$. Fig. 10 shows that $B^u(z, d^u(z,\partial\mathcal{M})/2)$ looks like a "bowtie". If $z = (r,\varphi)$ is sufficiently close to $\partial\mathcal{M}$ (in the Riemannian sense) the bowtie is fairly horizontal and the maximum of $q(w)$ is achieved at the indicated point $z_1 = (r_1, \varphi_1)$. This point cuts the curve γ in two parts of equal length. γ is

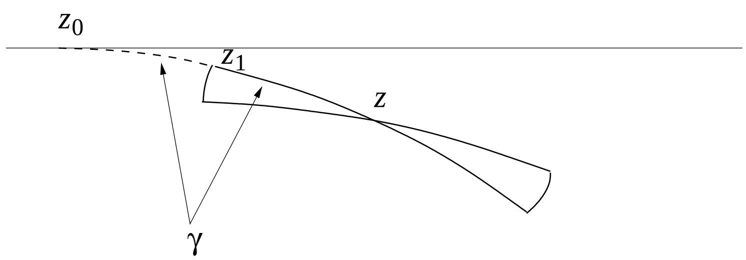

Figure 10: The "bowtie" $B^u(z, d^u(z, \partial\mathcal{M})/2)$.

exactly the type of unstable curve that we have studied in the proof of Theorem 5.1. It is the graph of a function $r \mapsto \varphi(r)$ that satisfies

$$\frac{d\varphi}{dr} = -k(r) - \frac{\sin\varphi}{\tau_{-1}(r,\varphi)}; \qquad \varphi(r_0) = \pi; \tag{6.29}$$

see (5.2). We need to prove that

$$\frac{q(z_1)}{q(z)} = \frac{\sin\varphi}{\sin\varphi_1} = \frac{\sin\varphi(r)}{\sin\varphi(r_1)} \le C'. \tag{6.30}$$

Defining $\widetilde{\varphi}(r) := \pi - \varphi(r + r_0)$, as in Section 6, and noting that $\sin\varphi = \sin\widetilde{\varphi} \sim \widetilde{\varphi}$, for φ close to π, a sufficient condition for (6.30) is

$$\frac{\widetilde{\varphi}(2\bar{r})}{\widetilde{\varphi}(\bar{r})} \le C_7, \tag{6.31}$$

having called $\bar{r} := r_1 - r_0$. The above holds since $\widetilde{\varphi}$ is a convex increasing function and $2(r_1 - r_0) > r - r_0$. A suitable version of the Lagrange Mean Value Theorem ensures that $\exists \hat{r} \in (0, \bar{r})$ such that

$$\frac{\widetilde{\varphi}(2\,\bar{r}_1)}{\widetilde{\varphi}(\bar{r}_1)} = \frac{2\dfrac{d\widetilde{\varphi}}{dr}(2\hat{r})}{\dfrac{d\widetilde{\varphi}}{dr}(\hat{r})}. \tag{6.32}$$

The denominator is bounded below by $k(r_0 + \hat{r})$, via (6.29). As to the numerator, that is the r.h.s. of a differential equation that we know how to estimate, from the proof of Theorem 5.1—at least for γ shorter than some ε_0. Indeed, fix ε_0 as in Theorem 5.1. If we regard γ as having an initial point in $(\hat{r}, \widetilde{\varphi}(\hat{r}))$, then (5.8)-(5.16) say that $d\widetilde{\varphi}/dr(2\hat{r}) \le C_8 k_M$, C_8 not depending on z_0. But $k_M = \max_{[r_0+\hat{r}, r_0+\varepsilon_0]} k$ was shown to be asymptotically of the same order as $k(r_0 + \hat{r})$. Hence (6.32) is bounded above.

We now show that *(h)* holds. Denote $z_{-n} := (r_{-n}, \varphi_{-n}) := T^{-n}z$. When z_{-n_k} belongs to any compact subset of $\mathcal{M}_0$, the portion $\{z_{-n}\}_{n=n_k+1}^{n_{k+1}-1}$ of its orbit lies in a

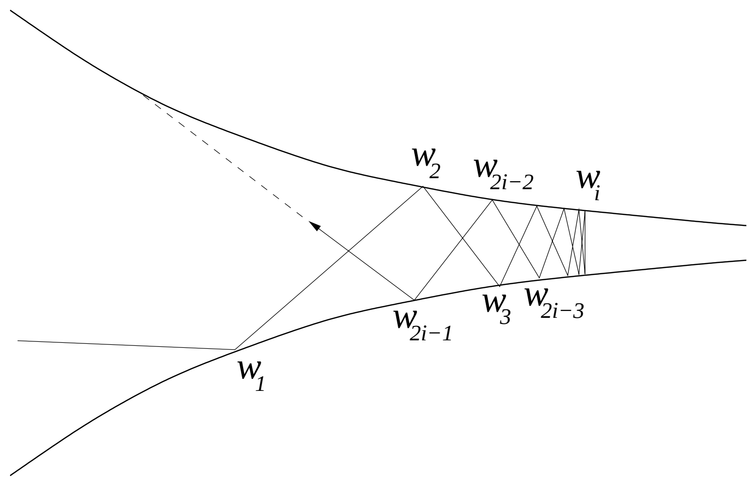

Figure 11: An example of a billiard trajectory moving towards the cusp and coming back. (This trajectory correspond exactly to the orbit of Fig. 9, provided $w_1 := T^{-n_{k+1}}z$ and $w_{2i-1} := T^{-n_k}z$.) This illustration shows how one can think of the two halves of the trajectory (the part moving right and the part moving left) as the boundary of a dispersing beam of trajectories originated at w_i. Denote $(r_j, \varphi_j) := w_j$. Then $j \mapsto \varphi_j$ is increasing, with φ_i very close to $\pi/2$. Also, for $j < i$, the arc (r_{j+2}, r_{2i-j-2}) lies entirely to the right of the arc (r_j, r_{2i-j}). Hence, at most one other forward point w_l (namely w_{j+1}), is such that of $r_j \le r_l \le r_{2i-j}$.

compact subset of $\mathcal{M}$ (this is not hard to establish): there q is bounded. Thus, we only need to prove *(h)* for z_{-n_k} lying far on the right of $\mathcal{M}_0$; either above the graph of $\pi - h$, or below $\mathcal{S}^{2-}$.

In the former case, the backward orbit of z_{-n_k} is depicted in Fig. 9 and corresponds to a billiard trajectory going (back in time) towards the cusp and coming back, as shown in Fig. 11. The caption of that figure explains that there can be at most one value of n (precisely $n = n_{k+1} - 1$) for which $r_{-n} > r_{-n_k}$ and $1/q(z_{-n}) = \sin\varphi_{-n} \le \sin\varphi_{-n_k} = 1/q(z_{-n_k})$. For this n, it is not hard to get convinced that $\sin\varphi_{-n} \sim \sin\varphi_{-n_k}$. As for the other values of n for which $\sin\varphi_{-n} < \sin\varphi_{-n_k}$ (to avoid confusion we point out that there are none, in Fig. 9), z_{-n} must lie to the left of z_{-n_k}, and above $\mathcal{S}^{2-}$ by construction. Hence, from the asymptotics $g(r) \gg h(r)$,

$$\sin\varphi_{-n} \sim \varphi_{-n} \ge g(r_{-n}) \ge g(r_{-n_k}) \gg h(r_{-n_k}) \ge \varphi_{-n_k} \sim \sin\varphi_{-n_k}. \tag{6.33}$$

On the other hand, there is nothing to prove for the case in which z_{-n_k} lies below $\mathcal{S}^{2-}$. In fact, the past billiard trajectory of that point crosses the y-axis in Q_4: therefore z_{-n_k-1} lies above $\mathcal{S}^{2+}$, hence in $\mathcal{M}_0$.

7 Absolute continuity

The purpose of this section is to establish the absolute continuity of the LSUMs w.r.t. the invariant measure μ. Later on we will specify precisely what this means.

We need to introduce yet another cross-section for the billiard flow: the cross-section induced by countably many *transparent walls* $G_n := \{X_n\} \times [0, f(X_n)]$, as depicted in Fig. 12. We choose X_n ($n > 1$) such that $f(X_n) = n^{-3}$, and consider only line elements based in one of the G_n's. The phase space is $\mathcal{M}_1 := \bigsqcup_{n\geq 1} \left(\mathcal{M}^{r,n} \sqcup \mathcal{M}^{l,n}\right)$, with $\sqcup$ denoting the disjoint union. $\mathcal{M}^{l,n}$ is defined as $(0, n^{-3}) \times (0, \pi)$ and its points indicated by (r, φ); the position variable r is the y-coordinate of the point in G_n, and the direction variable φ is the counterclockwise angle ($\leq \pi$) between the velocity vector and the y-direction. Thus, line elements in $\mathcal{M}^{l,n}$ point left, whence the notation. $\mathcal{M}^{r,n}$ is formally defined in the same way, but r equals n^{-3} minus the y-coordinate of the point, and φ is the counterclockwise angle between the unit vector and the negative y-direction. Line elements in $\mathcal{M}^{r,n}$ point right (Fig. 12).

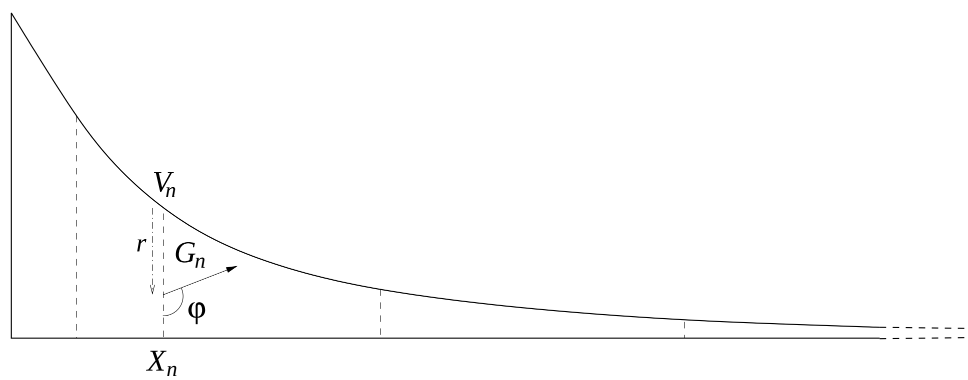

Figure 12: The definition of G_n, V_n, and $\mathcal{M}^{r,n}$.

The Poincaré map, which we name T_1, is defined on all points of $\mathcal{M}_1$ that would not result in a tangency or the hitting of a vertex V_n (see Fig. 12 for the definition of V_n). We call the set of these excluded points $\mathcal{R}$ and, for $i = l, r$, denote $\mathcal{R}^{i,n} := \mathcal{R} \cap \mathcal{M}^{i,n}$. A sketch of these two sets is given in Fig. 13, and an explanation follows momentarily, after an important remark.

Remark 7.1 For a point $z \in \mathcal{M}_1 \setminus \mathcal{R}$ whose billiard trajectory hits $\mathcal{U}$ k times before crossing the next transparent wall (say G_m), $D(T_1)_z$ equals the product of the differentials DT_{z_i} corresponding to the rebounds on $\mathcal{U}$ (at suitable z_i, $i = 0, ..., k-1$), times $-DT_{z_k}$, where DT_{z_k} would correspond to a *rebound* on G_m—as opposed to a crossing. This is so because the transparent wall can be regarded as a "double bouncer", i.e., causing two instantaneous collisions, one from and one towards

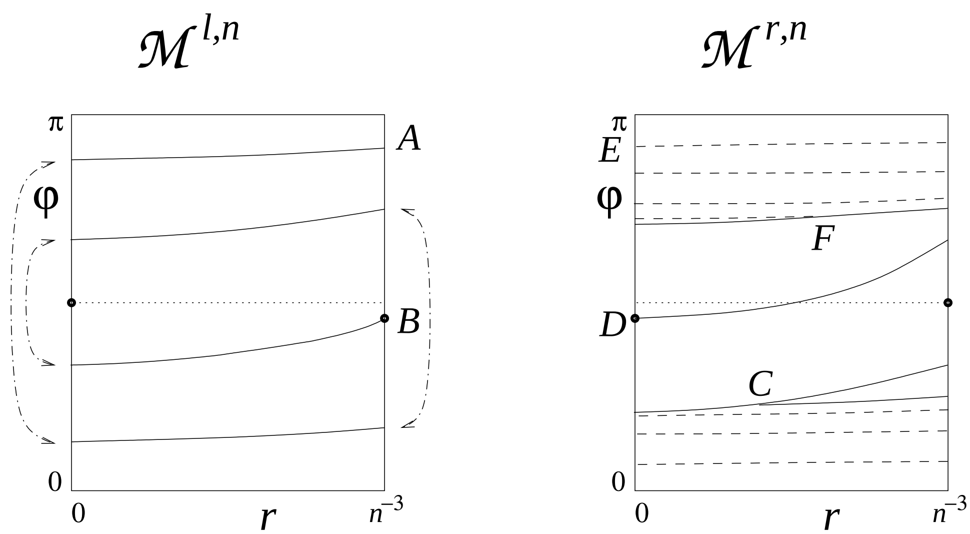

Figure 13: The singularity lines $\mathcal{R}^{l,n} \in \mathcal{M}^{l,n}$ and $\mathcal{R}^{r,n} \in \mathcal{M}^{r,n}$. The bullets represent the fixed points for the identifications in the vertical segments of $\partial\mathcal{M}^{i,n}$, $i = l, r$. These identifications are made explicit, for some points, in the left picture. Notice that the distance from B (or D) to the line $\varphi = \pi/2$ is β_n.

G_m. The transfer matrix between them is of the form (3.5) with $\varphi_1 = \pi - \varphi$, and $k = k_1 = \tau = 0$; that is, minus the identity.

It is not hard to figure out that $\mathcal{R}^{i,n}$ is a collection of curves, each being the graph of an increasing function of r. Denote by $\beta_n := \arctan(|f'(X_n)|)$ the angle between the horizontal direction and the tangent line to $\mathcal{U}$ in V_n. In Fig. 13, in the left picture, $\mathcal{R}^{l,n}$ can be regarded as a continuous curve, once we identify $(0, \varphi) \leftrightarrow (0, \pi - \varphi)$ and $(n^{-3}, \varphi) \leftrightarrow (n^{-3}, \pi - 2\beta_n - \varphi)$; which are the proper identifications for line elements based in $(X_n, 0)$ and V_n, respectively. This continuous curve, from point A to point B, emcompasses all initial conditions in $\mathcal{M}^{l,n}$ that end up in V_{n-1}, or hit $\mathcal{U}$ tangencially.

As concerns $\mathcal{M}^{r,n}$, the suitable identifications are $(0, \varphi) \leftrightarrow (0, \pi - 2\beta_n - \varphi)$ and $(n^{-3}, \varphi) \leftrightarrow (n^{-3}, \pi - \varphi)$. There are two possible sources of singularity in this case: Initial conditions that end up in V_{n+1} or hit $\mathcal{U}$ tangentially (they correspond to the solid curve running from C to D); and initial conditions that move off to the right, come back and hit V_n (dashed curve from E to F). One can recognize that the self-intersection in C corresponds to the billiard trajectory of Fig. 14. Furthermore, F corresponds to a trajectory that hits V_{n+1} almost vertically and then continues its motion to the left to hit V_n.

What was discussed above should convince one that the number of lines in $\mathcal{R}^{i,n}$ is related to the maximum number of rebounds against $\mathcal{U}$, for points in G_n. Let us call this latter integer M_n. For the rest of this section, we use the symbols $\ll, \gg, \sim$ for the asymptotics $n \to +\infty$.

Lemma 7.2 *Assuming* (A5), $M_n \ll n^{\theta_0}$, *with* $\theta_0 := 6\xi - 1$.

PROOF. For a material point starting in G_n, the maximum number of $\mathcal{U}$-rebounds before crossing one of the walls G_{n-1}, G_n, or G_{n+1}, is evidently given by the trajectory shown in Fig. 14.

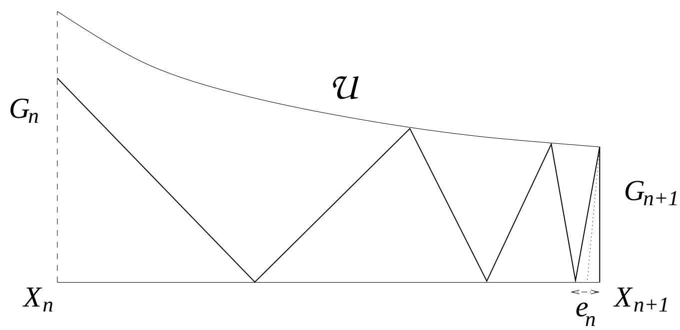

Figure 14: Among the trajectories that start in G_n, this one achieves the maximum number of rebounds against $\mathcal{U}$, before intersecting another transparent wall. The depicted polyline is run over twice, its velocity being reversed at the point $(X_{n+1}, 0)$.

Define e_n as in the picture, that is,

$$e_n := f(X_{n+1}) \tan(2\beta_{n+1}) \sim f(X_{n+1}) \, |f'(X_{n+1})| = (n+1)^{-3} \, |f'(X_{n+1})|. \tag{7.1}$$

One sees that

$$M_n \ll \frac{X_{n+1} - X_n}{e_n}, \tag{7.2}$$

the implicit constant being at most two, or so (the trajectory is run over twice). Now, $f'(X_{n+1}) = f'(f^{-1}((n+1)^{-3})) = 1/(f^{-1})'((n+1)^{-3})$. For the sake of the notation, let us name $g := f^{-1}$ and $t_n := n^{-3}$. Then (7.1) becomes

$$e_n \sim \frac{t_{n+1}}{|g'(t_{n+1})|}. \tag{7.3}$$

Also,

$$X_{n+1} - X_n = g(t_{n+1}) - g(t_n) \leq |g'(t_{n+1})| \, (t_n - t_{n+1}), \tag{7.4}$$

having used the monotonicity of f' (hence g'). Now, (7.3) and (7.4) turn (7.2) into

$$M_n \ll |g'(t_{n+1})|^2 \, \frac{t_n - t_{n+1}}{t_{n+1}} \sim |g'(t_{n+1})|^2 \, \frac{1}{n}. \tag{7.5}$$

But from (A5)

$$|g'(t_{n+1})| = \frac{1}{|f'(f^{-1}(t_{n+1}))|} \ll \frac{1}{[f(f^{-1}(t_{n+1}))]^\theta} \sim t_n^{-\theta} = n^{3\theta}, \tag{7.6}$$

which, together with (7.5), finishes the proof of the lemma. Q.E.D.

Coming back to the number of singularity lines in $\mathcal{M}^{i,n}$, we can now be more precise and see that $\mathcal{R}^{l,n}$ has about $M_{n-1}/2$ lines, whereas $\mathcal{R}^{r,n}$ has $M_n + M_{n+1}/2$ (in Fig. 13, the dashed lines plus the solid lines).

Furthermore, we are in the position to show that the stable and unstable manifolds of T in $\mathcal{M}$ can be carried over to $\mathcal{M}_1$. We do this for the LUMs only, the other case being of course just the same. Let us consider T_2, the Poincaré map corresponding to the cross-section $\mathcal{M}_2 := \mathcal{M} \sqcup \mathcal{M}_1$. For $z \in \mathcal{M}_1$, set k to be the smallest positive integer for which $T_2^k z \in \mathcal{M}$; then define $\widetilde{W}^u(z) := T_2^{-k}\, W^u(T_2^k z)$. Although $\widetilde{W}^u(z)$ is contracting in the past, in general it will not be a C^1 curve, due to the singularities $\mathcal{R}$. We must therefore "prune" it in such a way that all conditions of Definition 6.1 remain valid.

To this end, notice that, by Lemma A.3 of the Appendix $\mu\left(\mathcal{R}_{\{\varepsilon\}}\right)$ goes to zero like a power in ε (the notation $\mathcal{R}_{\{\varepsilon\}}$ is also introduced there). Furthermore, setting $q(r,\varphi) = 1/\sin\varphi$ as in Section 6, it is evident that $\int_{\mathcal{M}_1} q\, d\mu < \infty$. By the standard arguments that we are familiar with, by now, these two facts imply respectively that the backward images of a.e. $z \in \mathcal{M}_1$, via T_1, approach $\mathcal{R}$ only polynomially, and the deformation constant between the Riemannian distance and the $\|\cdot\|$-distance also grows polynomially. Since $T_1^{-k}\,\widetilde{W}^u(z)$ is contracting exponentially w.r.t. $\|\cdot\|$ (by Lemma A.1, because $\mu(\mathcal{M}_1) < \infty$), it follows that it can fail to be smooth only for a finite number of k's. Hence it can be pruned in such a way that Definition 6.1, *(b)*, holds, without the risk of reducing its length to zero. (The reader that finds this argument too sketchy can check that Theorem 6.5 applies to $(\mathcal{M}_2, T_2, \mu)$.)

Remark 7.3 At this point, it might be worthy to discuss why we needed to introduce the new Poincaré section $\mathcal{M}_1$. Its main asset is simply that $\mu(\mathcal{M}_1) < \infty$, which guarantees exponential contraction. One might reply that we had exponential contraction already, w.r.t. the return times to $\mathcal{M}_0$. However, using that fuzzy cross-section, it is not clear how many T-iterations ($\mathcal{U}$-rebounds) can occur between two returns to $\mathcal{M}_0$; whereas $\mathcal{M}_1$ was specifically designed to ensure that the number of T-iterations between two returns grows at most polynomially, as we go back in time (we will check this below). So, why did we not use the simpler dynamical system $(\mathcal{M}_1, T_1, \mu)$ from the beginning, and avoid the cumbersome machinery of the previous sections? The answer is that for the proof of the tail bound (see Lemma 8.7), we need that the ε-neighborhoods of the singularity set have measure of order ε or better: $\mathcal{S}_{[\varepsilon]}$ satisfies this; for $\mathcal{R}_{\{\varepsilon\}}$ that is not clear.

In the rest of the section we will prove that the stable and unstable foliations verify the forthcoming definition. The term *foliation* is used in a rather sloppy way here. What we mean, of course, is the type of object we have been dealing with, so far: a collection of short leaves for a.e. point. *Measurable foliation* would be a more precise name, but we will not delve into these questions.

Definition 7.4 *A ν-dimensional foliation $\mathcal{W}$ in $\mathbb{R}^N$ is said to be* absolutely continuous *with respect to the measure μ if the following happens: Given any cylinder C endowed with an axis Θ (a cylinder is a set that, in some orthogonal frame, looks like $A \times \mathbb{R}^{N-\nu}$, with A a Borel set in $\mathbb{R}^\nu$; and an axis is any $\{x\} \times \mathbb{R}^{N-\nu}$, with $x \in A$), and any union L of leaves $W(z) \in \mathcal{W}$, transversal to C (i.e., to all axes of C) and exceeding it (i.e., $\partial W(z) \in \mathbb{R}^N \setminus C$), then we have*

$$\mu(L \cap C) = 0 \qquad \Longrightarrow \qquad \mathrm{Leb}_\Theta(L \cap \Theta) = 0;$$

where Leb_Θ *is the $(N-\nu)$-dimensional Lebesgue measure on Θ.*

More precisely, the above states that the transversal measure on $\mathcal{W}$ defined by Leb on Θ is absolutely continuous with respect to μ.

In our case, the situation can be simplified. First of all, we can use the 2-dimensional Lebesgue measure instead of μ, since they are *equivalent* (absolutely continuous w.r.t. each other). Second, consider the foliation $\mathcal{V}$ given by parallel straight lines strictly contained in the constant cone field $\mathcal{C}^-$ (this is no loss of generality). With a slight abuse of notation, we denote by ℓ not just the length given by the ordinary (Riemannian) distance d, but also the 1-dimensional Lebesgue measure on any smooth curve—e.g., on $\Theta \in \mathcal{V}$. Also, it is obvious that ℓ on a straight line orthogonal to $\mathcal{V}$ is the appropriate transverse measure for $\mathcal{V}$.

Now pick a $\Theta \in \mathcal{V}$ and a segment $I \subset \Theta$ such that, for ℓ-almost every $z \in I$, $W^u(z)$ exists (this can be achieved for ℓ-almost every Θ, by Fubini's Theorem). Call $W^u_r(z)$ the right part (say) of $W^u(z)$, w.r.t. Θ, and L the union of all the $W^u_r(z)$ based in I. If we can show that $\exists d_0 > 0$ such that, for ℓ-a.a. $\Theta' \in \mathcal{V}$ to the right of Θ, with $d(\Theta', \Theta) \le d_0$, $\ell(\Theta' \cap L) > 0$, then we have proved that the unstable foliation verifies Definition 7.4. (Simply by using Fubini and integrating Θ over $\mathcal{V}$.) Needless to say, the case of the stable foliation is completely analogous.

Theorem 7.5 *Assuming* (A1)-(A5)*, the stable and unstable foliations in $\mathcal{M}_1$ are absolutely continuous w.r.t. μ.*

Proof. Except for its final part (where we use Lemma 7.2 and the *ad hoc* construction of $\mathcal{M}_1$) this proof is very standard. We present it completely, however, because it is hard to derive it as a rigorous corollary of any of the theorems available in the literature. Archetypal results include [G, §4], [KS, Part II], and [PS, §4].

Set $A \subset \mathcal{M}_1$ to be a full-measure set of points satisfying some properties which will be unveiled during the course of the proof. To begin with, every $z \in A$ has a LUM.

Using the notation introduced above, take a $\Theta \in \mathcal{V}$ and a $I \in \Theta$ such that, for ℓ-a.e. $z \in I$, $z \in A$. Fix any such z and name it z_0. Denoting by L_ε the union of all $W_r^u(z)$, with $z \in I$ and $\ell(W_r^u(z)) \geq \varepsilon$, then $\exists \varepsilon_0 > 0$ such that $z_0 \in I \cap L_{\varepsilon_0}$ and $\ell(I \cap L_{\varepsilon_0}) > 0$. Fix $d_0 := const\, \varepsilon_0$; for a suitable choice of *const* we are assured that any $W_r^u(z)$ longer than ε_0 intersects any Θ' to the right of Θ, with $d(\Theta', \Theta) \leq d_0$ (remember that, by the construction of $\mathcal{V}$, the angle between the LUMs and the direction of $\mathcal{V}$ is bounded from below). Thus, fixed such a Θ', we define the *holonomy map* h on $I \cap L_\varepsilon$ so that $h(z)$ equals the unique point (by transversality) in $W_r^u(z) \cap \Theta'$. To simplify the notation, set $w_0 := h(z_0)$. See Fig. 15.

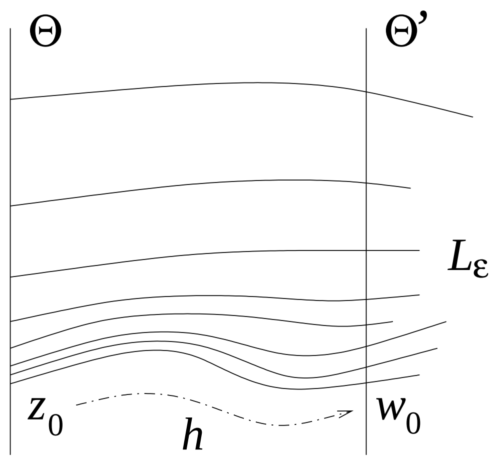

Figure 15: The construction of the holonomy map h.

Without loss of generality (i.e., possibly modifying A by a null-measure set), z_0 is a density point of $I \cap L_\varepsilon$, via the Lebesgue Density Theorem. Thus z_0 is also an accumulation point of $I \cap L_\varepsilon$, and it makes sense to speak of the Jacobian of the map h there, that is,

$$Jh_{z_0} := \lim_{z \to z_0} \frac{|h(z) - h(z_0)|}{|z - z_0|}. \tag{7.7}$$

The proof of Theorem 7.5 consists precisely in showing that the above limit exists and is positive.

The usual trick one employs is to pull back h to the holonomy map h_n between $T_1^{-n}\Theta$ and $T_1^{-n}\Theta'$, as sketched in Fig. 16. Set $z_{-n} := T_1^{-n} z_0$ and $w_{-n} := T_1^{-n} w_0$. Then, by definition (7.7), $\forall n \in \mathbb{N}$,

$$Jh_{z_0} = J(T_1^{-n})_{z_0}\, J(h_n)_{z_{-n}}\, J(T_1^n)_{w_{-n}} = J(h_n)_{z_{-n}} \prod_{k=0}^{n-1} \frac{J(T_1^{-1})_{z_{-k}}}{J(T_1^{-1})_{w_{-k}}}. \tag{7.8}$$

Introducing u_{-k}, a unit tangent vector to $T_1^{-k}\Theta$ at z_{-k}, and v_{-k}, a unit tangent vector to $T_1^{-k}\Theta'$ at w_{-k}, one sees that the quantities

$$J(T_1^{-1})_{z_{-k}} = \left| D(T_1^{-1})_{z_{-k}} u_{-k} \right|, \qquad J(T_1^{-1})_{w_{-k}} = \left| D(T_1^{-1})_{z_{-k}} v_{-k} \right| \tag{7.9}$$

are actual derivatives, since, by definition of A, $W^u(z_0)$ (which contains z_0 and w_0) belongs in $\mathcal{M}_1 \setminus \mathcal{R}^-_\infty$. Therefore, $\forall k$, a certain neighborhood of $W^u(z_{-k})$ is contained in a connected component of $\mathcal{M}_1 \setminus \mathcal{R}^-$, where T_1^{-1} is smooth.

At this point, one would like to show that the product in (7.8) converges to a finite number, and $\lim_{n\to+\infty} J(h_n)_{z_{-n}} = 1$, since $T_1^{-n}\Theta$ and $T_1^{-n}\Theta'$ get closer and closer to each other as n grows. The argument, however, is a bit more complicated. Indeed we will prove those two facts not for the Jacobians, but for suitable approximations; precisely, quantities like the following:

$$R(h_n)(z_{-n}, z'_{-n}) := \frac{\ell(\mathrm{arc}(w_{-n}, w'_{-n}))}{\ell(\mathrm{arc}(z_{-n}, z'_{-n}))}, \tag{7.10}$$

where $w'_{-n} := h_n(z'_{-n})$ and $\mathrm{arc}(z, z')$ denotes the arc segment on the appropriate backward image of Θ or Θ' (in this case, $T_1^{-n}\Theta$ and $T_1^{-n}\Theta'$).

Recall the definition of "goodness" (6.7), used in the proof of Theorem 6.2, and keep in mind that, for $(\mathcal{M}_1, T_1, \mu)$, we can use the ordinary distance, as opposed to the unstable distance—because $\mu\left(\mathcal{R}_{\{\varepsilon\}}\right) \ll \varepsilon^\alpha$, for some $\alpha > 0$. Now, fix a $z'_0 \in I \cap L_\varepsilon$, which we think of as close to z_0, and consider the curvilinear quadrilateral P specified by z_0, w_0, $w'_0 := h(z'_0)$, as illustrated in Fig. 16. If λ denotes the same constant as in (6.6), set $\lambda_1 < \lambda$ and $C > 0$ such that $B(z_{-n}, C\, e^{-\lambda_1 n})$ is good $\forall n \geq 0$, and $W^u_r(z_0)$ is strictly contained in $B(z_0, C)$. (For this we might have to cut $W^u_r(z_0)$, but this is no loss of generality: one will simply take smaller ε_0 and d_0.) Furthermore, let n be the minimum integer such that $P_n := T_1^{-n} P$ is not contained in $B(z_{-n}, C\, e^{-\lambda_1 n})$. (This is possible since $\mu(P_n)$ is constant in n, while $\mu\left(B(z_{-n}, C\, e^{-\lambda_1 n})\right)$ vanishes—although maybe not monotonically, since $\mu \neq$ Leb.)

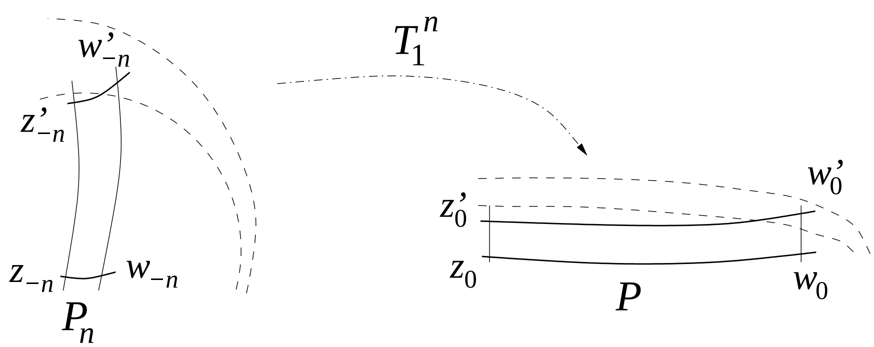

Figure 16: The stretching of the curvilinear quadrilateral P backwards in time. In the left picture, the inner dashed curve represents (part of) $\partial B(z_{-n}, C_2\, e^{-\lambda_1 n})$, and the outer dashed curve represents $T_1^{-1}\partial B(z_{-n+1}, C_2\, e^{-\lambda_1(n-1)})$. The dashed curves n the right picture are the T_1^n-iterates of these curves.

By construction, all points in P_n are good, relative to z_0. Therefore, for all $z \in I \cap L_\varepsilon$, $W^u_r(z)$ contracts at a rate faster than $e^{-\lambda' n}$, for any $\lambda' \in (\lambda_1, \lambda)$, see

(6.16). But the diameter of P_n is larger than $C\, e^{-\lambda_1 n}$. These two facts imply that, no matter how deformed P_n becomes, it will look more and more stretched along the stable direction, as $n \to +\infty$. Moreover, its two long opposite sides will have approximately the same length, as Fig. 16 tries to show. In fact not only is the distance between z_{-n} and z'_{-n} about the same as the distance between w_{-n} and w'_{-n}, but also the two joining curves will be rather "parallel", its tangent lines having to belong to the cone field $\mathcal{C}_n$, which must be narrow for n big—at least for almost every z_0.

This demonstrates that, as $z'_0 \to z_0$ (hence $n \to +\infty$), $R(h_n)(z_{-n}, z'_{-n}) \to 1$. We will have proved Theorem 7.5 when we are able to show that

$$\lim_{n\to+\infty} \prod_{k=0}^{n-1} \frac{R(T_1^{-1})(z_{-k}, z'_{-k})}{R(T_1^{-1})(w_{-k}, w'_{-k})} > 0. \tag{7.11}$$

By the Lagrange Mean Value Theorem, $R(T_1^{-1})(z_{-k}, z'_{-k}) = J(T_1^{-1})_{\bar{z}_{-k}}$, for some $\bar{z}_{-k} \in \operatorname{arc}(z_{-k}, z'_{-k})$, and analogously for $R(T_1^{-1})(w_{-k}, w'_{-k})$. Then, a sufficient condition for (7.11) is

$$\lim_{n\to+\infty} \sum_{k=0}^{n-1} \left| \log J(T_1^{-1})_{\bar{z}_{-k}} - \log J(T_1^{-1})_{\bar{w}_{-k}} \right| < \infty. \tag{7.12}$$

Note that we do not use the notation $\sum_{k=0}^{\infty}$ because the terms of the sum depend on n too, through $\bar{z}_{-k}$ and $\bar{w}_{-k}$. We can apply the Mean Value Theorem to each term above, to obtain

$$|\bar{w}_{-k} - \bar{z}_{-k}| \left| \frac{\partial}{\partial b_k} \log J(T_1^{-1})_{\widetilde{z}_{-k}} \right|, \tag{7.13}$$

with $\widetilde{z}_{-k}$ lying on the segment between $\bar{z}_{-k}$ and $\bar{w}_{-k}$, and b_k being a unit vector in the direction of such segment.

Here is where the specific design of $\mathcal{M}_1$ comes into play. For every point $z \in \mathcal{M}_1$, denote by $M(z)$ the number of $\mathcal{U}$-rebounds (i.e., T-iterations, or, to be more precise, T_2-iterations) before the point first returns to $\mathcal{M}_1$. By Lemma 7.2, there exists a $C_1 > 0$ such that

$$\left\{ z \in \mathcal{M}_1 \;\middle|\; M(z) \geq C_1\, k^{\theta_0} \right\} \subseteq \bigsqcup_{n=k}^{\infty} \left(\mathcal{M}^{r,n} \sqcup \mathcal{M}^{l,n} \right), \tag{7.14}$$

whose measure is of order k^{-2}. The associated series converges, therefore we can apply the usual Borel-Cantelli argument to conclude that, for all z in a full-measure subset of $\mathcal{M}_1$ (which contains A, without loss of generality), there exists a $C_2 = C_2(z)$ for which $M(z_{-k}) = M(T_1^{-k} z) \leq C_2\, k^{\theta_0}$.

Now let us notice that for any $w \in P_k$ (in particular for $\widetilde{z}_{-k}$), the number of $\mathcal{U}$-rebounds must be the same as for z_{-k}. In fact, if $M(w) \neq M(z_{-k})$ then w and

z_{-k} are separated by an $\mathcal{R}$-singularity line, which is not possible by the construction of P_k. Therefore $M(\widetilde{z}_{-k}) \le C_2\, k^{\theta_0}$. Decomposing $D(T_1^{-1})_{\widetilde{z}_{-k}}$ into the product of $M(\widetilde{z}_{-k})$ differentials of T_2^{-1}, turns (7.13) into

$$|\bar{w}_{-k} - \bar{z}_{-k}| \left| \sum_{i=0}^{M(\widetilde{z}_{-k})-1} \frac{\partial}{\partial b_k} \log J(T_2^{-1})_{\widetilde{z}_{-k,i}} \right|, \tag{7.15}$$

with $\widetilde{z}_{-k,i} = T_2^{-i} \widetilde{z}_{-k}$, for $i = 0, ..., M(\widetilde{z}_{-k}) - 1$. The strategy is now rather clear: Since $|\bar{w}_{-k} - \bar{z}_{-k}| \ll e^{-\lambda' k}$ and the number of terms in the sum is a power-law in k, it suffices to check that each such term is bounded by a power of k.

First of all, setting $b_{k,i} := D(T_2^{-i})_{\widetilde{z}_{-k}} b_k / |D(T_2^{-i})_{\widetilde{z}_{-k}} b_k|$, one observes that

$$\left| \frac{\partial}{\partial b_k} (\,\cdot\,) \right| \le \frac{1}{\sin \widetilde{\varphi}_{-k,i}} \left| \frac{\partial}{\partial b_{k,i}} (\,\cdot\,) \right|. \tag{7.16}$$

In fact, quite generally, b_k lies in the unstable cone of $\widetilde{z}_k$ (and, anyway, this can be made happen by suitably adjusting the definition of $\bar{w}_{-k}$ so that is lies on $W_r^u(\bar{z}_{-k})$). Hence, $\| D(T_2^{-i})_{\widetilde{z}_{-k}} b_k \| \le \| b_k \|$ and $|D(T_2^{-i})_{\widetilde{z}_{-k}} b_k| \le 1/ \sin \widetilde{\varphi}_{-k,i}\, |b_k|$.

We are reduced therefore to consider directional derivatives of $\log J(T_2^{-1})$. By Remark 7.1 we can consider $D(T_2^{-1})$ as the differential of a regular billiard map. Therefore we can apply Lemma A.4 of the Appendix to $\widetilde{z}_{-k,i} =: (\widetilde{r}_{-k,i}, \widetilde{\varphi}_{-k,i})$ and find a constant C_3 such that

$$\left| \frac{\partial}{\partial b_{k,i}} \log J(T_2^{-1})_{\widetilde{z}_{-k,i}} \right| \le \frac{C_3}{\sin^2 \widetilde{\varphi}_{-k,i}\, \sin^4 \widetilde{\varphi}_{-k,i+1}}, \tag{7.17}$$

with the understanding that, when $i = M(\widetilde{z}_{-k}) - 1$, $\widetilde{\varphi}_{-k,i+1}$ really means the second coordinate of $T_2^{-1} \widetilde{z}_{-k,i}$. But this point belongs to P_{k+1} and one can see that the forthcoming arguments are not invalidated.

Remark 7.6 There are two issues to clarify in order to use Lemma A.4. First, the lemma applies to billiards with finite horizon. This problem is easily circumvented. Say, for instance, that we have a T_2-iteration that corresponds to a segment of trajectory going from G_n to G_{n-1} (this is the longest free path, for points in $\mathcal{M}^{l,n}$). We can always divide it into M_n segments of approximately the same length, by imagining as many transparent walls between G_{n-1} and G_n. It is easy to see that this length is less than a quantity that depends only on the shape of the billiard.

The second issue is to control $(\partial u_{-k,i} / \partial b_{k,i})(\widetilde{z}_{-k,i})$ uniformly in k and i. We only spend a few words on this, which is a standard argument in Pesin's theory. The fact is that the $\widetilde{z}_{-k,i}$'s belong to a sequence of good sets with respect to the same z_0—more precisely, neighborhoods of $\{z_{-k,i}\}$. Within this sequence one enjoys some local form of uniform hyperbolicity. Therefore, to the (backward) images of the (stable) direction field $\mathcal{V}$ one applies the line of reasoning of [KH, Thm. 6.2.8, Step 5]: the iterates of $\Theta, \Theta' \in \mathcal{V}$ (or rather, appropriately short pieces thereof) approach uniformly the stable direction *faster* than they get close to each other. See also [PS, 3.10-3.14].

In view of (7.16)-(7.17) it remains to show that $q(\widetilde{z}_{-k,i}) = 1/\sin\widetilde{\varphi}_{-k,i}$ grows like a power of k. This property is easily checked for $q(z_{-k,i})$. Indeed, for the subsequence of $\{z_{-k,i}\}$ corresponding to the returns to $\mathcal{M}$, we have already proved it in Section 6—see in particular Theorem 6.5, *(f)* and *(h)*. As regards the returns to $\mathcal{M}_1$, one uses the same arguments, given that also $\int_{\mathcal{M}_1} q\, d\mu < \infty$.

At this point, we cannot argue that the same property must hold for $\widetilde{z}_{-k,i}$, being this point sufficiently close to $z_{-k,i}$ as to be good throughout the backward orbit of z_0. In other words, we cannot use Theorem 6.5, *(g)*, because we are now dealing with T_2-iterations and, although the Riemannian distance between $\widetilde{z}_{-k,i}$ and $z_{-k,i}$ vanishes exponentially, the unstable distance might be infinity (that is, $\widetilde{z}_{-k,i}$ does not belong to any $B^u(z_{-k,i}, \rho)$, $\rho > 0$). However, one can reason as follows: Since $\widetilde{z}_{-k,i} \in P_{k,i} := T_2^{-i} P_k$, whose sides shrink exponentially in k, then, at least for large k, it must belong to the curvilinear triangle $F_{k,i}$ defined like this: take the two sides of $P_{k,i}$ that intersect in $z_{-k,i}$ and prolong them arbitrarily, in the stable and unstable direction, respectively, until their lengths are $d^s(z_{-k,i}, \partial\mathcal{M})/2$ and $d^u(z_{-k,i}, \partial\mathcal{M})/2$. (These latter quantities are bigger than a negative power of k.) Finally, connect the two resulting vertices by a segment. Since $q(z)$ is a function of one variable only, its maximum on $F_{k,i}$ is always achieved by a point on the curvilinear sides. Then, a double application of Theorem 6.5, *(g)*, for both unstable and stable balls, shows that the value of q along the two curvilinear sides is comparable to $q(z_{-k,i})$. But this has the right rate of growth, as we have recalled. Q.E.D.

We conclude this section by noting that, due to the a.e. smoothness of T_2, the absolute continuity that we have established for $\mathcal{M}_1$ is immediately proved for $\mathcal{M}$ as well.

8 Ergodicity

For infinite-measure dynamical systems some formulations of the ergodicity property fail to be equivalent—some fail to be physically significant altogether. We must therefore declare which one we will use from this time forward. It turns out that Boltzmann's very definition is convenient for our purposes.

Definition 8.1 *A dynamical system $(\mathcal{M}, T, \mu)$ is called* ergodic *when, $\forall \psi \in L^1(\mathcal{M}, \mu)$, the time average*

$$\psi^*(z) := \lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\psi \circ T^k)(z)$$

is constant for μ-a.e. $z \in \mathcal{M}$.

This is admittedly a rather weak notion of ergodicity: the system can be decomposed into two invariant subsets of infinite measure and still be ergodic. Although we are able to prove constance of the time average for a larger class of functions (see Remark 8.10), the result that definitely settles the question of ergodicity is

Proposition 8.11. This claims that a certain *finite* cross-section of the billiard flow (corresponding to the returns onto the leftmost part of the boundary) is ergodic.

The main ingredient for all these results is the *local ergodicity theorem*, which we present here in a fairly general rendition suited to infinite systems (Theorem 8.5). Our version is a generalization of the one formulated by Liverani and Wojtkowski in [LW]. We have chosen it because it utilizes invariant cones. Since much of its proof is of a local nature, and is unaffected by the finiteness, or lack thereof, of the invariant measure, we will not duplicate that part, referring the reader to [LW, §§8-12]. The only lemma that needs modification is the so-called *tail bound* [LW, §13], which we restate and prove later in this section.

Theorem 8.5, like all similar results, requires several technical conditions. In order to state them we need another definition.

Definition 8.2 *A compact subset A of $\mathbb{R}^N$ is called* regular *if it is a finite union of pieces Σ_i of codimension-one submanifolds, such that:*

(a) Σ_i is the closure of its interior (in the topology of the submanifold);

(b) The pieces overlap at most on their boundaries, i.e., $\Sigma_i \cap \Sigma_j \subset \partial\Sigma_i \cap \partial\Sigma_j$;

(c) $\partial\Sigma_i$ is a finite union of compact subsets of codimension-two submanifolds.

A is called locally regular *if, $\forall z \in \mathbb{R}^N$, there is a neighborhood U of z such that $A \cap U$ is regular.*

For the sake of the format, we list the conditions before giving the statement of the theorem. It is the goal of Section 9 to provide a class of billiards that verify (C1)-(C8).

(C1) Phase space. The phase space $\mathcal{M}$ is an open, connected subset of $\mathbb{R}^{2\nu}$, with $\partial\mathcal{M}$ regular. $\mathbb{R}^{2\nu}$ is endowed with a symplectic form which is assumed to be equivalent to the standard one; i.e., the symplectic volume element $d\mu$ is assumed to be absolutely continuous w.r.t. the standard volume element dLeb, and viceversa.

(C2) Map. The map T is invertible and recurrent. T is not defined on the singularity set $\mathcal{S}^+$; and T^{-1} is not defined on $\mathcal{S}^-$. (Morally $\mathcal{S}^\pm = T^{\mp 1}\partial\mathcal{M}$, in the sense that one can construct ill-behaved extensions of T and T^{-1} for which that holds.) On $\mathcal{M} \setminus \mathcal{S}^+$, T preserves the symplectic form mentioned in (C1).

Remark 8.3 Some of the restrictions formulated above are not really essential. First, the choice of a linear space was only made to simplify the estimation of the measure of the tubular neighborhoods of certain regular sets (see below). A 2ν-dimensional symplectic manifold, embedded in $\mathbb{R}^N$, connected, and with bounded geometry, would have worked as well. Second, requiring the dynamical system to be Hamiltonian (that is, a symplectomorphism where defined) has the effect that certain properties are symmetric for time reversal (e.g., (C3) or [LW, §8]). Assuming those properties to hold for both directions of time would yield the same result. See also Remark 4.2.

(C3) Cone bundle. There exists a cone bundle $\mathcal{C}^+ = \mathcal{C}$, continuous where the map is defined, and *eventually strictly invariant* for the map. This means that, for almost all z, there is an $n = n(z)$ such that $DT_z^n \mathcal{C}(z) \subset int\mathcal{C}(T^n z)$. (C1) guarantees that the same happens for $\mathcal{C}^-$, the cone bundle comprised by the closures of the cones $\mathbb{R}^{2\nu} \setminus \mathcal{C}^+$.

(C4) Local regularity of singularity sets. Denoting, as we have done in the past, $\mathcal{S}_n^\pm := \bigcup_{i=0}^{n-1} T^{\mp 1}\mathcal{S}^\pm$, we suppose that $\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ are locally regular for all n.

(C5) Measure of tubular neighborhoods. On $\mathcal{S}^-$ there is a finite measure π_-, such that, for every closed subset A of $\mathcal{S}^-$ (in the topology of $\mathcal{S}^-$),

$$\mu(A_{[\varepsilon]}) \leq \pi_-(A)\,\varepsilon + o(\varepsilon), \qquad \text{as } \varepsilon \longrightarrow 0^+.$$

Here, as introduced in (5.1), $A_{[\varepsilon]}$ denotes the tubular neighborhood of A, of radius ε, in the unstable distance. π_- must be absolutely continuous w.r.t. $\mu_{\mathcal{S}^-}$, the measure induced on $\mathcal{S}^-$ by μ and the ordinary distance. An analogous condition stands for $\mathcal{S}^+$ and the stable distance.

(C6) Proper alignment of singularity sets. For a codimension-one subspace of a linear symplectic space, the characteristic line is defined as the skew-orthogonal complement. (In our case, in two dimensions, the subspace and the characteristic line are the same thing.) We assume that the tangent space to $\mathcal{S}^+$ at any point $z \in \mathcal{S}^+$ has a characteristic line contained in $\mathcal{C}^-(z)$. The reversed condition holds for $\mathcal{S}^-$.

(C7) Non-contraction property around the singularity sets. For every $z_0 \in \mathcal{M}$, there exist a neighborhood U_0 of z_0, an $\varepsilon_0 > 0$, and a $K > 0$, such that, every time $w \in \mathcal{S}^-_{[\varepsilon_0]}$ and $T^n w \in U_0$, with $n > 0$, then

$$|DT_w^n v| \geq K|v|,$$

for $v \in \mathcal{C}_1(w)$. Here $|\cdot|$ is the modulus of a vector in $\mathbb{R}^{2\nu}$ (i.e., the appropriate Riemannian norm). Also, the definition of $\mathcal{C}_1(w)$ is the same as in (4.2). The analogous condition stands for the time-reversed dynamical system.

At this point, there is one last condition to formulate, the so-called *Sinai–Chernov Ansatz*. This can be given in two versions: a very general one which, however, can be cumbersome to check in some instances; and a more specific one for systems (like semi-dispersing billiards) that possess an increasing norm for unstable vectors. For the sake of completeness we give both, although in the remainder we will only work with the second version, more natural in our framework.

For the first case we need to use the notion of *expansion coefficient* $\sigma_{\mathcal{C}}(L)$ for a linear symplectic map L w.r.t. an invariant cone $\mathcal{C}$. We skip its rather lengthy definition which, however, can be found in [LW, §§4-6]. (After all, we never use this object in the present work.)

(C8)$_1$ Sinai–Chernov Ansatz for the expansion coefficient. For π_--a.e. $z \in \mathcal{S}^-$,

$$\lim_{n\to\infty} \sigma_{\mathcal{C}}(DT^n_z) = +\infty.$$

Once again, the analogous condition stands for the time-reversed dynamical system.

Before presenting the alternative version of the Sinai–Chernov Ansatz, we need to better specify what we need by increasing norm, mainly to simplify definitions and proofs.

Definition 8.4 *The dynamical system defined above is said to have an* increasing norm *for unstable vectors, $\|\cdot\|$, if this norm satisfies Theorem 6.2, (c) and (d), the latter with $H = \emptyset$, and is locally equivalent to the Riemannian norm $|\cdot|$.*

(C8)$_2$ Sinai–Chernov Ansatz for the increasing norm. For π_--a.e. $z \in \mathcal{S}^-$,

$$\lim_{n\to+\infty} \inf_{0\neq v\in\mathcal{C}_1(z)} \frac{\|DT^n_z v\|}{\|v\|} = +\infty.$$

Of course, the analogous condition holds for $\mathcal{S}^+$ and T^{-1}.

We are now in position to state the local ergodicity theorem.

Theorem 8.5 *Consider a dynamical system $(\mathcal{M}, T, \mu)$, endowed with a hyperbolic structure (LSUMs at a.e. point, absolutely continuous w.r.t. μ). Suppose, furthermore, that this system satisfies* (C1)-(C8)$_i$ *($i = 1$ or 2). Then, for any z_0 that possesses a semiorbit (i.e., $z_0 \in \mathcal{M} \setminus \mathcal{S}^+_\infty$, or $z_0 \in \mathcal{M} \setminus \mathcal{S}^-_\infty$), there is a neighborhood U of z_0 belonging to one ergodic component of T.*

Remark 8.6 The statement can be strengthened to include points that only possess a finite orbit. Using the expansion coefficient, Liverani and Wojtkowski [LW, §7], prove it for z such that $\sigma(DT^n_z) > 3$, for some n, positive or negative.

For the purpose of stating and proving the tail bound lemma that, together with the results in [LW, §§8-12], will grant Theorem 8.5, we need to recall just one fact from the arguments that we omit. Since it is crucial that the local stable and unstable manifolds be as large as possible, one decides to prolong $W^{s(u)}(z)$, as defined in Section 6, as much as it is compatible with the requirements of Definition 6.1 (see [LW, Thm. 9.7]). This implies that the boundary of the new, say, unstable manifolds are made up of points of $\partial\mathcal{M}$ or $T^i\mathcal{S}^-$, for some $i \geq 0$ (no smoothness is possible across the singularity!). For simplicity, we keep denoting any such "grown" LUM with the same symbol $W^u(z)$, and say that it is *cut* by $\partial\mathcal{M}$ or $T^i\mathcal{S}^-$. Furthermore, looking back at the proof of Theorem 6.2, we define the *radius* of $W^u(z)$ to be the inf of $\ell(\gamma)$ over all smooth curves $\gamma \subset W^u(z)$ that join z with $\partial W^u(z)$.

Lemma 8.7 *For every $z_0 \in \mathcal{M}$, there is a neighborhood U of z_0, and a $\delta_0 > 0$ such that, $\forall \eta > 0$, $\exists M$ that verifies*

$$\mu\left(\left\{ z \in U \;\middle|\; W^u(z) \text{ has radius} < \delta \text{ because it is cut by } \bigcup_{i=M+1}^{\infty} T^i\mathcal{S}^- \right\}\right) \leq \eta\delta,$$

for every $\delta \leq \delta_0$.

Remark 8.8 As anticipated, we prove this lemma only under condition $(\mathrm{C8})_2$. Using $(\mathrm{C8})_1$ would only amount to minor changes in the proof which, anyway, can be reconstructed with the aid of [LW, §13].

Proof of Lemma 8.7. For the sake of the notation, we will henceforth drop the super/subscripts from $\mathcal{S}^-$ and π_-. For a linear map $L : \mathrm{T}\mathcal{M}_z \longrightarrow \mathrm{T}\mathcal{M}_w$ leaving the cone bundle $\mathcal{C}$ inviariant, let us denote

$$\sigma_*(L) := \inf_{0 \neq v \in \mathcal{C}(z)} \frac{\|Lv\|}{|v|}. \tag{8.1}$$

Since the increasing norm and the ordinary norm are locally equivalent, it follows from $(\mathrm{C8})_2$ that $\lim_{n\to\infty} \sigma_*(DT_z^n) = +\infty$, for π-a.e. $z \in \mathcal{S}$.

Fix $h > 0$, to be thought of as a small parameter. Since π is finite, one can choose $E_1 \subset \mathcal{S}$ such that $\pi(E_1) \leq h$ and $\mathcal{S} \setminus E_1$ is bounded. From (C4), for any compact subset B of $\mathcal{M}$, $\mathcal{S} \cap B$ is also compact. Hence, without loss of generality, we can also assume that $\mathcal{S} \setminus E_1$ is compact and $\pi(\overline{E_1}) \leq h$, where the bar denotes, here and in the sequel, the closure in $\mathcal{S}$. Now, for every (large) parameter $s > 0$, there is an $M = M(h, s)$ such that

$$E_2 := \left\{ z \in \mathcal{S} \;\middle|\; \sigma_*(DT_z^M) \leq s + 1 \right\} \tag{8.2}$$

has measure $\pi(E_2) \leq h$. Notice that E_2 is closed. As the reader has apprehended, we are throwing away points of $\mathcal{S}$ that somehow might give us complications. For reasons that will be clear later, we want the remaining set to be compact. Eliminating the points in E_2 might have destroyed this property, which we would like to recover now, through another purge. The map T^M is discontinuous at $\mathcal{S}_M^+$ which, by proper alignment, intersects $\mathcal{S} = \mathcal{S}^-$ only in pieces of codimension-two manifolds.

Therefore the discontinuity set of $T^M\big|_{\mathcal{S}}$ is a set of $\mu_{\mathcal{S}}$-measure(hence π-measure) zero. Let us remove a neighborhood E_3 of this set (in $\mathcal{S}$), such that $\pi(\overline{E_3}) \le h$.

The above choice guarantees the compactness of

$$\mathcal{S}^s := \mathcal{S} \setminus \bigcup_{j=1}^{3} E_j = \left\{ z \in \mathcal{S} \setminus (E_1 \cup E_3) \;\middle|\; \sigma_*(DT_z^M) \le s+1 \right\} \tag{8.3}$$

because of the continuity of $\sigma_*(DT_z^M)$ in the above domain. Moreover, $\sigma_*(DT_z^M)$ is also continuous in a certain neighborhood of $\mathcal{S}^s$ in $\mathcal{M}$. This neighborhood necessarily contains $\mathcal{S}^s_{[c]}$, for c small enough.

At this point, take U_0 and ε_0 as in (C7). Since the increasing norm and the Riemannian norm are locally equivalent, $\exists C_1 > 0$ such that, $\forall w \in U_0$,

$$\| \cdot \|_w \le C_1 | \cdot |_w. \tag{8.4}$$

Now take U, a smaller neighborhood of z_0, such that $U_{[C_1\delta_0]} \subseteq U_0$. (Should this condition lead to trouble—in the sense that $B^u(z_0, C_1\delta_0)$ already exceeds U_0—we can always take a smaller δ_0 with no damage for the proof.)

The main idea behind the tail bound is to split $Y(\delta, M)$, the set that appears in the statement of the lemma, into pieces whose measure is easy to estimate. We proceed as follows. For $z \in Y(\delta, M)$, denote $m(z)$ the smallest $i \ge M+1$ such that $W^u(z)$ is cut by $T^i\mathcal{S}$. Also set

$$k(z) := \# \left\{ i = 1, ..., m(z) - M \;\middle|\; T^{-i}z \in U \right\}. \tag{8.5}$$

Why the returns to U is this stretch of orbit are important, we will see later. Let us introduce $Y_m^k := \{ z \in Y(\delta, M) \mid m(z) = m,\, k(z) = k \}$. We claim that

$$T^{-m}Y_m^k \cap T^{-m'}Y_{m'}^k = \emptyset, \text{ for } m \ne m'. \tag{8.6}$$

In fact, suppose not and assume, say, that $m < m'$. If $w \in T^{-m}Y_m^k \cap T^{-m'}Y_{m'}^k$ then, for $z := T^m w$ and $z' := T^{m'} w$, we would have $k(z') > k(z)$, which is absurd. From the claim and the invariance of μ we obtain that, for a fixed $k \in \mathbb{N}$,

$$\mu\left(\bigcup_{m>M} Y_m^k \right) \le \sum_{m>M} \mu\left(Y_m^k\right) = \sum_{m>M} \mu\left(T^{-m}Y_m^k\right) = \mu\left(\bigcup_{m>M} T^{-m}Y_m^k \right). \tag{8.7}$$

By Definition 8.4 (noting in particular that $H = \emptyset$), we see that there exist a $\rho < 1$ such that, $\forall w \in U_0$ and $v \in \mathcal{C}(w)$,

$$\|DT_w^{-n}v\|_{T^{-n}w} \le \rho^j \|v\|_w, \tag{8.8}$$

if j is the number of returns to U_0 of the piece of orbit $\{T^{-i}w\}_{i=1}^n$. The above holds *uniformly* in $w \in U_0$. In fact, assuming that U_0 stays away from $\partial\mathcal{M}$, the amount of contraction (for unstable vectors and relative to the increasing norm) at every return is bounded below by $\inf_{w \in U_0} \kappa(w)^{-1} =: \rho$ (see Theorem 6.2, *(c)*).

Fix $z \in Y_m^k$ ($k \in \mathbb{N}$, $m \geq M+1$); there is by definition a smooth curve γ that connects z to a point $z' \in \partial W^u(z) \cap T^m \mathcal{S}$, and such that $\ell(\gamma) < \delta$. We observe that

$$\ell_{\|}(T^{-n}\gamma) < C_1 \rho^k \delta, \tag{8.9}$$

n being the largest $i \leq m - M$ such that $T^{-i}z \in U$. In fact, by the definition of U, if $T^{-i}z$ returns k times to U, $T^{-i}\gamma$ returns k times to U_0, and one can apply (8.4) and (8.8) to the points of γ. Now, $T^{-m}z' \in \mathcal{S}$ and we have two cases.

CASE 1: $T^{-m}z' \in \bigcup_{j=1}^3 E_j$. Here we first use (8.9), then switch to the Riemannian length, and finally employ (C7) to estimate the maximum (Riemannian) expansion during the time $i = n+1, ..., m$. The net result is

$$\ell(T^{-m}\gamma) < \frac{C_2}{K} \rho^k \delta, \tag{8.10}$$

for some $C_2 > 0$. The above is an *a priori* estimate that can be proved correct (by contradiction, for instance) once we choose $\delta_0 \leq \varepsilon_0 / C_2$, so that $C_2 \rho^k \delta_0 \leq \varepsilon_0$, $\forall k \leq 0$. Therefore, setting $C_3 := C_2/K$,

$$T^{-m}z \in T^{-m}\gamma \subset \left(\bigcup_{j=1}^3 \overline{E_j} \right)_{[C_3 \rho^k \delta]}, \tag{8.11}$$

whose measure, by (C5) and the estimates above, does not exceed, say, $2\pi(\cup_j \overline{E_j}) C_3 \rho^k \delta \leq 6hC_3\rho^k\delta$ (possibily taking a smaller δ_0).

CASE 2: $T^{-m}z' \in \mathcal{S}^s$. In this case it is not hard to see that

$$\ell(T^{-m}\gamma) < \frac{C_1}{s} \rho^k \delta. \tag{8.12}$$

Once again, this is an *a priori* estimate and everything works rigorously provided $C_1\delta_0 \leq c$, where c was introduced before. Therefore, in analogy with (8.11),

$$T^{-m}z \in \mathcal{S}^s_{[C_1 s^{-1} \rho^k \delta]}. \tag{8.13}$$

Using again (C5) and estimating $\pi(\mathcal{S}^s)$ from above by $\pi(\mathcal{S})$, we conclude that the measure of the above set is less than or equal to $2\pi(\mathcal{S})C_1 s^{-1}\rho^k\delta$, for δ_0 small enough.

The previous two estimates, together with (8.6)-(8.7), yield

$$\mu\left(\bigcup_{m>M} T^{-m} Y_m^k \right) \leq \left[6\,h\,C_3 + \frac{2\,\pi(\mathcal{S})\,C_1}{s} \right] \rho^k \delta, \tag{8.14}$$

whence

$$\mu\left(Y(\delta, M)\right) \leq \left[6\,h\,C_3 + \frac{2\,\pi(\mathcal{S})\,C_1}{s} \right] \frac{\delta}{1-\rho}. \tag{8.15}$$

For a given $\eta > 0$, the coefficent of δ above can be made smaller than η, if h and s are chosen suitably small and large, respectively. These, in turn, determine how big M must be for the statement of Lemma 8.7 to hold true. Q.E.D.

The most valuable consequence of Theorem 8.5 is, of course, the *global* ergodicity of some of our systems.

Proposition 8.9 *If the billiard map T introduced in Section 3 is locally ergodic, as in Theorem 8.5, then it is ergodic, in the sense of Definition 8.1.*

PROOF. We do this in two steps. First, we show that only a countable number of points in $\mathcal{M} = (0, +\infty) \times (0, \pi)$ can fail to verify Theorem 8.5, that is, to be in the interior of an ergodic component. Second, we observe that removing these points leaves $\mathcal{M}$ connected, nevertheless. Ergo, there is only one ergodic component.

As concerns the first assertion, we evaluate the cardinality of $\mathcal{S}^+_\infty \cap \mathcal{S}^-_\infty$: the sets $\mathcal{S}^+_\infty$ and $\mathcal{S}^-_\infty$ are countable unions of smooth curves, respectively increasing and decreasing (in fact, $\mathcal{S}^-_\infty$ is even a mirror image of $\mathcal{S}^+_\infty$). Therefore there can be at most one point of intersection for each pair of increasing–decreasing curves. This means, at most countably many points.

The proof of the second assertion is the contents of Lemma A.5 in the Appendix. Q.E.D.

Remark 8.10 As already anticipated, one can verify the statement of Definition 8.1 for a larger class of *observables* than just $L^1(\mathcal{M}, \mu)$. In fact, let us recall Hopf's main argument (cf. [LW, Prop. 2.3]): Once one has proved that, in a certain neighborhood, almost any two points are connected by an alternating sequence of *typical* stable and unstable manifolds that intersect at *typical* points, the constance of ψ^* is automatic for all uniformly continuous ψ whose backward and forward time averages exist and coincide almost everywhere. Let us call this space $\mathcal{X}$. Then, by density (cf. [LW, Cor. 2.4]), the statement is extended to $\overline{\mathcal{X}}$, the completion of $\mathcal{X}$ in the L^1 metric. For instance, all functions of the type $\psi + const$, with $\psi \in L^1(\mathcal{M}, \mu)$, belong in $\overline{\mathcal{X}}$. Finally, it should be noticed that the requirement on the the almost-everywhere equality of the two time averages is not a concern for recurrent systems (Lemma A.6 of the Appendix).

Let us denote by $\mathcal{M}_3$ the region of $\mathcal{M}$ that lies above $\mathcal{S}^{2+}$ (Fig. 5). By the definition of $\mathcal{S}^{2+}$ it should be easy to realize that the points in $\mathcal{M}_3$ are precisely the line elements whose trajectory in Q_4 cross the y-axis before the next rebound (Fig. 4). Therefore, if we call T_3 the return map onto $\mathcal{M}_3$, this map is isomorphic (because of the invariance of μ) to the first return map onto the vertical part of ∂Q.

Proposition 8.11 *For billiards that verify Theorem 8.5, $(\mathcal{M}_3, T_3, \mu)$ is ergodic.*

Proof. This proposition is derived from the proof of Theorem 8.5 (once again, a porism—see Corollary 5.2) much in the same way as Simányi derives (under some assumptions) the ergodicity of the Lorentz gas from the ergodicity of the Sinai billiard [Si, Prop. 1.3].

By hypothesis and by Proposition 8.9, almost every two points $z', z'' \in \mathcal{M}_3$ are connected by a finite alternating sequence of typical stable and unstable manifolds (of T!), say, $W^s(z_0), W^u(z_1), ..., W^u(z_m)$, with $z_0 := z'$ and $z_m := z''$. These manifolds intersect at typical points. (See also Remark 8.10.) A typical point is by definition a point of A, where A is an arbitrary full-measure set of $\mathcal{M}_3$; a typical LSUM is a LSUM whose almost every point is typical.

Now, take a function ψ compactly supported in $\mathcal{M}_3$ and continuous (thus, uniformly continuous). For any $z \in \mathcal{M}_3$ for which the following exists, denote by

$$\psi^{\pm}(z) := \lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\psi \circ T_3^{\pm k})(z) \tag{8.16}$$

the forward and backward averages of ψ w.r.t. T_3. By Birkhoff's Theorem and Lemma A.6, the set

$$A := \left\{ z \in \mathcal{M}_3 \,\middle|\, \exists \psi^+(z), \psi^-(z) \text{ and } \psi^+(z) = \psi^-(z) =: \psi^*(z) \right\} \tag{8.17}$$

has full-measure in $\mathcal{M}_3$. We use this to define typicality. Ergodicity for T_3 will be achieved when we show that $\psi^+(z') = \psi^-(z'')$ for a.a. $z', z'' \in \mathcal{M}_3$. By virtue of the above we only need prove that, given z, $\psi^+(w)$ is constant for all typical $w \in W^s(z)$, and $\psi^-(w)$ is constant for all typical $w \in W^u(z)$.

Let us consider only the case $w \in W^s(z)$, the other being naturally analogous. We claim that, for $k > 0$, $T_3^k z = T^{n_k} z$ and $T_3^k w = T^{n_k} w$. In other words, the return times to $\mathcal{M}_3$ are the same for z and w. If this were not true, then we would have a k such that, say, $T^{n_k} z \in \mathcal{M}_3$ and $T^{n_k} w \notin \mathcal{M}_3$. Thus $T^{n_k} W^s(z)$ would intersect $\mathcal{S}^{2+}$ in an interior point, which is absurd. (To be honest, we have previously proved only that the interior of any $T^n W^s(z)$ cannot intersect $\mathcal{S}^-$, but it is very easy to modify the proof of Theorem 6.2 to make sure that it will not intersect $\mathcal{S}^+$, either.)

Therefore, as $k \to +\infty$, $\left| T_3^k w - T_3^k z \right| \to 0$ and, since ψ is uniformly continuous,

$$\left| \psi(T_3^k w) - \psi(T_3^k z) \right| \to 0. \tag{8.18}$$

Whence $\psi^+(w) = \psi^+(z)$. Q.E.D.

9 Ergodic tables

In view of Propositions 8.9 and 8.11, we devote the last part of this work to checking that the functions

$$f(x) = Cx^{-p}, \quad C, p > 0, \tag{E1}$$

give rise to systems that verify Theorem 8.5—thus yielding examples of ergodic billiards with a non-compact cusp. Moreover, some of these tables have an infinite area, which is especially nice, for the reasons we have outlined in the introduction.

To start with, we will fix $C := 1$, since that constant never plays a role in our computations. Conditions (C1) through (C6) from Section 8 are rather easy to establish. (For (C4) one might notice that $\mathcal{S}^{\pm}$ is composed of three smooth curves, two of which are unbounded. Thus, $\mathcal{S}_n^{\pm}$ will comprise a finite number of smooth curves, some of which failing to be compact only because of their unboundedness. Hence, *local* regularity is guaranteed.)

As concerns $(C8)_2$, we derive from (6.22) and (4.6)-(4.7) that

$$\inf_{0 \neq v \in \mathcal{C}(z)} \frac{\| DT_z v \|}{\| v \|} \geq \min \left\{ \left(1 + \frac{k\tau}{\sin \varphi} \right), \left(1 + \frac{k_1 \tau}{\sin \varphi_1} \right) \right\}, \tag{9.1}$$

with the usual notation $z = (r, \varphi)$, etc. It is implicitly written in (A4) that $f'' > 0$, so the curvature of $\mathcal{U}$ is always positive; and continuous, by the assumptions on the differentiability of f. Therefore $k\tau$ and $k_1\tau$ are bounded below when the point is sufficiently far away from V and from the cusp at infinity. But Proposition 3.2 ensures that, for *every* orbit, this happens infinitely many times.

It remains to verify the non-contraction property (C7). This is quite often the hardest property to check for a billiard system (e.g., [LW, §14]). As a matter of fact, our formulation of (C7) is so cumbersome precisely because we are only able to prove non-contraction in its weakest useful form. Only in this part of the paper do we need more properties of f than just (A1)-(A5).

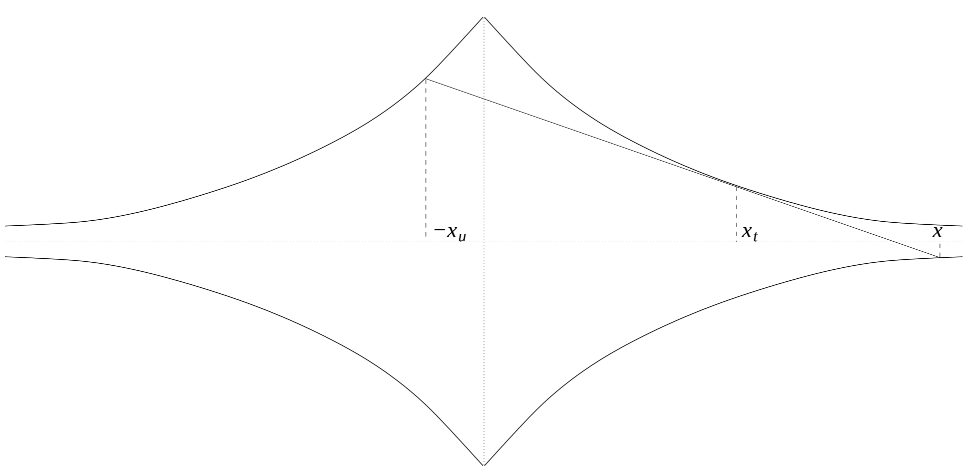

Figure 17: The definition of x_u.

Let us start by illustrating a feature of f that is very similar to (A2). Looking at Fig. 17 and recalling the definition of $x_t = x_t(x)$, consider the tangent line to ∂Q_4

at $(x_t, f(x_t))$. Then denote by $-x_u = -x_u(x) < 0$ the abscissa of the point at which this line intersects ∂Q_4 in the second quadrant. We verify that

$$|f'(x_u)| \ll |f'(x)|. \tag{9.2}$$

In fact, in analogy with (3.1), x_u is uniquely determined by

$$\frac{f(x) + f(x_u)}{x + x_u} = -f'(x_t). \tag{9.3}$$

Using the fact that, for our f, $x_t/x = const$ (cf. Section 3), we can write an equation similar to (3.3) and conclude that $x_u/x = const$, too; whence (9.2).

At this point, let us introduce r_T, a large number to be determined later. Denoting $\mathcal{M}_T := \{(r, \varphi) \in \mathcal{M} \mid r < r_T\} = (0, r_T) \times (0, \pi)$, we single out the line elements relative to Q_T, a certain truncated billiard whose four-fold copy, $Q_{4,T}$, appears in Fig. 19. Given the shape of $\mathcal{S}^-$, it is obvious that there is an ε_0 so small that $\mathcal{S}^-_{[\varepsilon_0]} \cap \mathcal{M}_T$ stays away from the line $\varphi = 0$ (see Fig. 18). Therefore, for $w \in \mathcal{S}^-_{[\varepsilon_0]} \cap \mathcal{M}_T$ and v an unstable vector based in w, one has

$$|DT^n_w v| \geq \min\left\{1, \frac{\sin\varphi}{\sin\varphi_n}\right\} |v| \geq C_1 |v|, \tag{9.4}$$

for some positive C_1 and all n. (We have changed notation since (9.1): here and for the rest of the section $(r, \varphi) := w$ and $(r_n, \varphi_n) := w_n := T^n w$.)

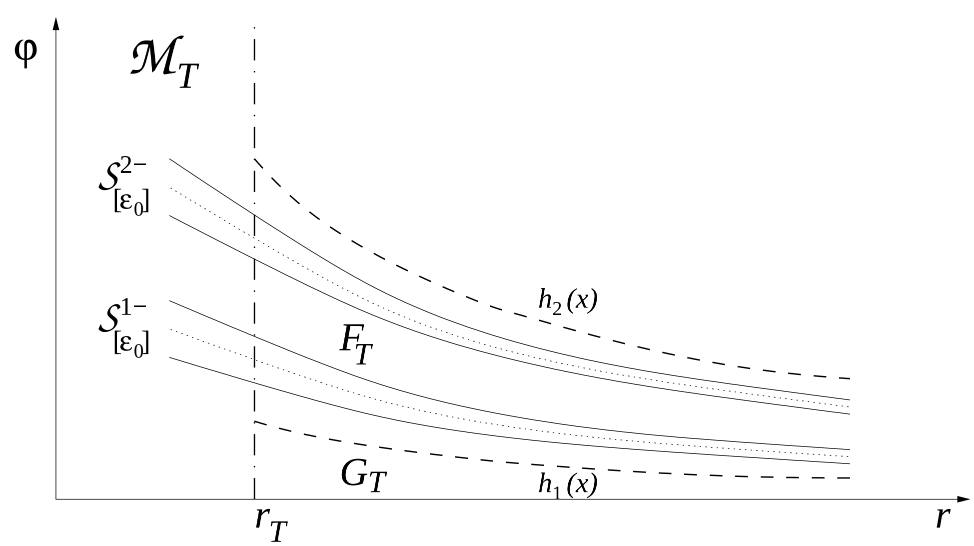

Figure 18: The definition of F_T and G_T.

It remains to consider the case $w \in \mathcal{S}^-_{[\varepsilon_0]} \setminus \mathcal{M}_T$. We will see later that it pays off to be more general and take $w \in F_T$, a set that we introduce now with the aid of Fig. 18. First of all, $F_T \subset \mathcal{M} \setminus \mathcal{M}_T$ and $F_T \supset \mathcal{S}^-_{[\varepsilon_0]} \setminus \mathcal{M}_T$. The leftmost part of ∂F_T belongs to the segment $r = r_T$. We know from Section 6 (see in particular Fig. 9, definition (6.27) and following paragraph) that $\mathcal{S}^{2-}$ and the upper and lower boundaries of $\mathcal{S}^{2-}_{[\varepsilon_0]}$ are the graphs of three functions that behave asymptotically like $|f'(x)|$. The same is true for $\mathcal{S}^{1-}$. This implies that it is possible to take the lower and upper boundaries of F_T to be the graphs of two functions h_j ($j = 1, 2$) such that

$$h_j(x) = K_j |f'(x)|. \tag{9.5}$$

On K_2 we will also impose some extra condition later on, other than its graph lying above $\mathcal{S}^{2-}_{[\varepsilon_0]}$. Finally, let us call G_T the region to the right of $r = r_T$ and below the graph of h_1.

At this point, we assume that r_T is big enough, so that $z_0 \in \mathcal{M}_T$. Then we take U_0 to be strictly contained in $\mathcal{M}_T$. Furthermore, without loss of generality, r_T is so large that every $w \in F_T$ points very much torwards the right, when regarded as a unit vector based somewhere on $\mathcal{U}$; e.g., we can make sure that the orbit of w makes a minimum number of rebounds to the right before starting to move left.

Here we are going to use the arguments expounded at the end of Section 6, about a trajectory moving towards the cusp at infinity and coming back—see in particular Fig. 11 and its caption. Fix a w as above and call m the number of rebounds the corresponding trajectory performs in the part of Q_4 that lies to the right of the truncated billiard; in other words, w_m is the last rebound before the material point either crosses the y-axis or hits the dispersing part of $\partial Q_{4,T}$ (or both). Fig. 19 shows some examples of w_m, together with w', which is the velocity vector right before the collision at r (more precisely, w' is a translation, along the billiard trajectory, of $w_{-1} := T^{-1}w$). As explained in Section 6, w_m and $-w'$ can be thought of as the (oriented) boundary of a dispersing beam of orbits originating in a point further right into the cusp (more or less the point where the trajectory starts moving left). Since the beam is dispersing, its focus (the intersection between the straight lines defined by w' and w_m) lies outside Q_4, as Fig. 19 illustrates.

Remark 9.1 One might point out that in Fig. 19 there is no need for r and r_m to lie on the same piece of ∂Q_4, that is, on the same copy of $\mathcal{U}$. This is absolutely true, and in fact we have assumed nothing of the sort: w_m should be based in the first-quadrant or in the fourth-quadrant copy of $\mathcal{U}$, depending on k being odd or even. The choice of Fig. 19 is purely illustrative and does not affect any of the forthcoming calculations.

We make a brief digression in order to derive an inequality that will be crucial in the remainder. Suppose we have a finite portion of an orbit (say $\{w_i\}_{i=0}^m$) and we want to estimate the amount of horizontal expansion for unstable vectors. We can do better than simply use (4.6) recursively, for $i = 0, ..., m$. In fact, a repeated

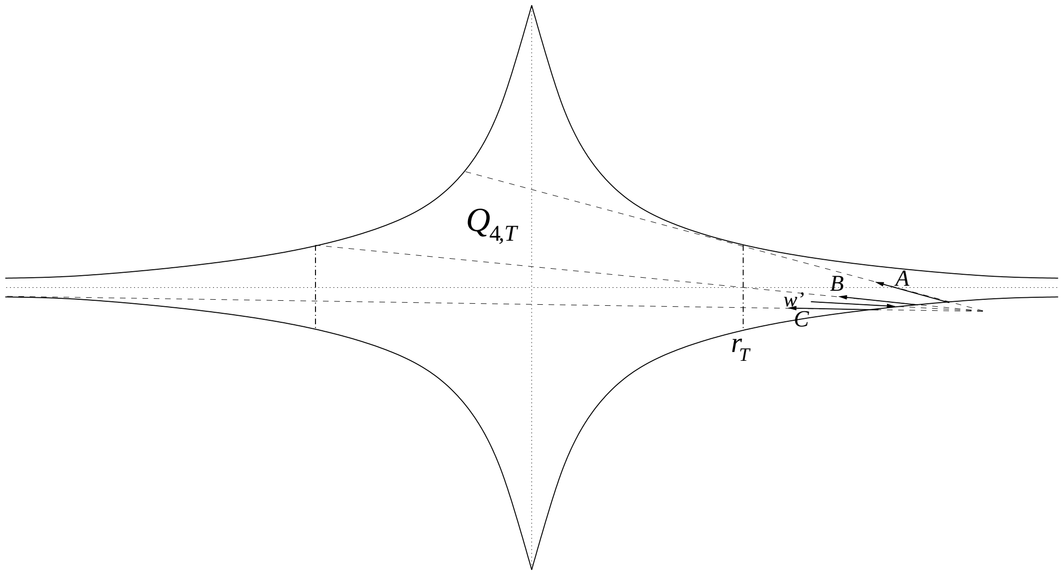

Figure 19: Given w', the velocity vector of a trajectory traveling towards the cusp, and necessarily coming back, we present several possibilities for w_m, the last rebound in the region to the right of $Q_{4,T}$. These possibilities are chosen from the same dispersing beam, i.e., they have a common focus.

application of Lemma A.7 of the Appendix, together with (4.6) itself, proves in particular that

$$(dr_{m+1})^2 \geq \left(\frac{\sin\varphi}{\sin\varphi_{m+1}} \right)^2 \left(1 + \frac{k}{\sin\varphi} \left(\sum_{i=0}^{m} \tau_i \right) \right)^2 dr^2. \tag{9.6}$$

In other words, what the above is saying is simply that the amount of horizontal expansion for unstable vectors can only decrease if one considers all rebounds (apart from the starting point $w = w_0$) to take place against flat pieces of the boundary. And this is, after all, obvious for semi-dispersing billiards.

Therefore, considering the finite segment of orbit $\{w_i\}_{i=0}^m$, if we take $v \in \mathcal{C}_1(w)$ and set $\bar\tau := \sum_{i=0}^m \tau_i$, we obtain

$$\frac{|DT_w^{m+1} v|}{|v|} \geq C_2 \left(\frac{\sin\varphi}{\sin\varphi_{m+1}} + \frac{k\bar\tau}{\sin\varphi_{m+1}} \right), \tag{9.7}$$

for some $C_2 \in (0,1)$. In fact, let us observe that, if $v =: (dr, d\varphi) \in \mathcal{C}_1(w)$, then $|dr| \geq C_2 |v|$, because $\mathcal{C}_1(w)$ becomes thinner and more horizontal as w stays in F and moves to the right—cf. (5.2). (That is, one can actually select C_2 arbitrarily close to 1, provided r_T is big enough.) The remaining part of estimate (9.7) is just (9.6).

It is easy to see that there are now only three cases, concerning w_{m+1}: $w_{m+1} \in F_T$, $w_{m+1} \in G_T$, and $w_{m+1} \in \mathcal{M}_T$.

Case 1: $w_{m+1} \in F_T$, that is, that the trajectory of w_m crosses the truncated billiard and hits the second or third-quadrant portions of ∂Q_4, with an incidence angle not so close to zero. An example is vector B of Fig. 19. We claim that the r.h.s. of (9.7) can be made bigger than 1 if r_T was previously selected to be large enough. In formula,

$$|DT_w^{m+1} v| > |v|. \tag{9.8}$$

Since $w \in F_T$, and due to (9.5),

$$\sin\varphi \gg |f'(x)|. \tag{9.9}$$

To estimate $\sin\varphi_{m+1}$ we consider the worst case. This occurs when the focus of the beam lies very close to ∂Q_4 (that is, the beam is as dispersing as it can be), and w_m describes a segment of trajectory tangent to $\mathcal{U}$. This case is labeled by A in Fig. 19. (Actually, vector A of Fig. 19 even overestimates the worst case, since its trajectory does not reach the region to the left of $Q_{4,T}$; however, there are in general segments of trajectory that reach the left region and are tangent to $\mathcal{U}$.) At the limit, when the focus of the beam lies on ∂Q_4, $r = r_m$. In this case, with the aid of Fig. 17,

$$\sin\varphi_{m+1} \sim |f'(x_t)| + |f'(x_u)| \ll |f'(x)|, \tag{9.10}$$

the last estimate coming from (A2) and (9.2). Using once again the correspondence $r \longleftrightarrow x$ defined by (5.7), we have $k(r) \sim f''(x)$. Therefore, putting everything together,

$$\frac{\sin\varphi}{\sin\varphi_{m+1}}\left(1 + \frac{k\bar{\tau}}{\sin\varphi}\right) \geq C_3 + C_4\,\frac{f''(x)}{|f'(x)|}\,\bar{\tau}, \tag{9.11}$$

where $\bar{\tau} = \bar{\tau}(w) = \bar{\tau}(x,\varphi)$. Since for the f's we are considering, $(f''/f')(x) = const/x$, the claim in (9.8) will be settled once we have proved the following lemma—which we do at the end of the section.

Lemma 9.2 *For f as in* (E1) *and r_T fixed,*

$$\lim_{x\to+\infty} \min_{(r(x),\varphi)\in F_T} \frac{\bar{\tau}(x,\varphi)}{x} = +\infty.$$

Case 2: $w_{m+1} \in G_T$. All the estimates that we have produced in Case 1 hold even more generously in this case, since $\sin\varphi_{m+1}$ is much smaller than the worst case (9.10). However, for a reason that is going to be clear momentarily, we want to avoid the case in which the angle of incidence of w_{m+1}, on the third-quadrant copy of $\mathcal{U}$ is too close to zero, e.g., vector C of Fig. 19.

We decide instead to look at w_{m+2}, the next iteration of T. First of all, let us verify that $w_{m+2} \in F_T$. One has that $|f'(x_{m+1})| \ll |f'(x_{m+2})|$; in fact, the largest value of $|f'(x_{m+1})|/|f'(x_{m+2})|$ is achieved when $x_{m+1} = x_t(x_{m+2})$, to which case we apply (A2). Also, by (3.6), if one wishes,

$$\varphi_{m+2} = \varphi_{m+1} + \arctan|f'(x_{m+1})| + \arctan|f'(x_{m+2})|. \tag{9.12}$$

But $\varphi_{m+1} \leq K_1|f'(x_{m+1})|$, because $w_{m+1} \in G_T$. Hence, if K_2 in (9.5) was chosen large enough, $\varphi_{m+2} \leq K_2|f'(x_{m+2})|$ and $w_{m+2} \in F_T$. Furthermore

$$\sin\varphi_{m+2} \ll |f'(x_{m+2})| < |f'(x)|, \tag{9.13}$$

as one sees that $x > x_{m+2}$. Now we use (9.7) with $m+2$ instead of $m+1$ and, exploiting also the other inequalities presented above, we conclude that

$$|DT_w^{m+2} v| > |v|. \tag{9.14}$$

Before moving on to Case 3, let us explain why we had decided to distinguish Case 2 from Case 1. The reason is because if we set w' to be w_{m+1} or w_{m+2}, depending on Case 1 or 2, then $w' \in F_T$. So one can consider the previous three cases on w' too; and so on, recursively. In particular we see that, as long as the trajectory keeps oscillating between the left and the right cusp and fails to bounce within $Q_{4,T}$,

$$|DT_w^{m_j+1} v| > |v|, \tag{9.15}$$

where the m_j are the analogues of m or $m+1$, at the future returns from the "excursions" in the cusp.

Case 3: $w_{m+1} \in \mathcal{M}_T$. Estimate (9.8) holds in this case as well; after all, we have derived it by considering situation A of Fig. 19, which is the worst possibility even when $w_{m+1} \in \mathcal{M}_T$.

As we will see later, we would like to have also

$$\sin\varphi_{m+1} \geq C_5, \tag{9.16}$$

for some $C_5 = C_5(r_T)$. This is in general not true, as Fig. 19 shows. We can have $w_m \notin \mathcal{M}_T$ and $\sin\varphi_{m+1}$ arbitrarily close to zero (in fact, the segment of trajectory originated by w_m can hit ∂Q_4 indefinitely close to tangentially). If this is the case, however, we can fix things at the next rebound, recycling the ideas used before. It is simple to verify that, chosen C_5 small enough, if $\sin\varphi_{m+1} > C_5$, then w_{m+1} is based in the first-quadrant part of $\partial Q_{4,T}$ (in fact, hitting the second or third-quadrant boundary would imply that $\varphi_{m+1} > \arctan|f'(x_T)| =: M$). Furthermore, the next rebound will occur in the second quadrant, with an angle of incidence φ_{m+2} bigger than, e.g., M.

Finally we notice that, if we do decide to consider the next rebound, then $|DT_w^{m+2} v| > |v|$, as $\sin\varphi_{m+2}$ is smaller than $\sin\varphi_{m+1}$ would have been, had w_m been tangent. In either case, therefore, we are fine.

We are ready to verify (C7), at last. If, for some $n > 0$, w_n is to belong in U_0 (hence in $\mathcal{M}_T$), there must be a positive integer l such that $w_{m_l+1} \in \mathcal{M}_T$. Excluding the possibility that we have to consider w_{m_l+2} (this would not change much, as we have just seen), one can apply (9.15) and (9.16) for $m = m_l$, to obtain

$$\frac{|DT_w^n v|}{|v|} = \frac{|DT_w^{m_l+1} v|}{|v|} \frac{|DT_w^n v|}{|DT_w^{m_l+1} v|} > \min\left\{1, \frac{\sin\varphi_{m_l+1}}{\sin\varphi_n}\right\} \geq C_5, \tag{9.17}$$

which settles the non-contraction property.

It remains to give the proof that was held off earlier.

Proof of Lemma 9.2. First of all, it is clear that $\bar\tau(x,\varphi)$ is a decreasing function of φ, at least for φ small. Then, by definition of F_T,

$$\min_{(r(x),\varphi)\in F_T} \bar\tau(x,\varphi) = \bar\tau(x, h_2(x)) =: \bar\tau_m(x), \tag{9.18}$$

with the customary misuse of notation $h_2(x) = h_2(r(x))$. So we are reduced to studying the trajectory of $w := (r(x), h_2(x))$, for large values of x.

Using the same notation as before, we call $x_n = x(r_n)$ the abscissa of the n-th collision point, whose line element is $w_n = (r_n, \varphi_n)$. By construction, $x_0 = x$. Setting $\alpha_n := \arctan|f'(x_n)|$ and rephrasing (9.12) gives

$$\varphi_{n+1} = \varphi_n + \alpha_n + \alpha_{n+1}. \tag{9.19}$$

With a bit of geometry (taking perhaps a look at Fig. 8), we check that

$$\tau_n \sin(\varphi_n + \alpha_n) = f(x_{n+1}) + f(x_n). \tag{9.20}$$

$$\tau_n \cos(\varphi_n + \alpha_n) = x_{n+1} - x_n; \tag{9.21}$$

All these quantities ultimately depend on x.

Lemma 9.3 *If $f(x) = x^{-p}$, $p > 0$, there exists an increasing sequence $\{\xi_n\}$ such that, for fixed n,*

$$\lim_{x\to+\infty} \frac{x_n(x)}{x} = \xi_n.$$

Furthermore

$$\lim_{n\to+\infty} \xi_n = +\infty.$$

Proof. As concerns the first assertion, we will prove it by induction. For $n = 0$ there is nothing to prove, as $x_0 = x$ (whence $\xi_0 = 1$). So let us assume that the limit above exists for all $i \leq n$ and try to show that it exists for $n + 1$, too.

It will be convenient in the sequel to introduce the symbol $\simeq$. Its meaning refines that of $\sim$: by definition $f(x) \simeq g(x)$ states that $f(x)/g(x) \to 1$ as $x \to +\infty$.

Let us take the ratio of (9.20)-(9.21):

$$\tan(\varphi_n + \alpha_n) = \frac{f(x_{n+1}) + f(x_n)}{x_{n+1} - x_n}. \tag{9.22}$$

As n is fixed, $\varphi_n(x)$ and $\alpha_n(x)$ tend to zero, when $x \to +\infty$. Therefore

$$\tan(\varphi_n + \alpha_n) \simeq \varphi_n + \alpha_n \simeq h_2(x) + |f'(x)| + 2\sum_{i=1}^{n} |f'(x_i)|, \tag{9.23}$$

since, from (9.19), $\varphi_n = \varphi_0 + \alpha_0 + 2\sum_{i=1}^{n-1} \alpha_i + \alpha_n$, and $\varphi_0 = h_2(x)$. Once we plug $f(x) = x^{-p}$ in (9.22)-(9.23) we obtain

$$p\left(K_2 x^{-p-1} + x^{-p-1} + 2\sum_{i=1}^{n} x_i^{-p-1}\right) \simeq \frac{x_{n+1}^{-p} + x_n^{-p}}{x_{n+1} - x_n}, \tag{9.24}$$

having used (9.5), too. Divide both sides by x^{-p-1}:

$$p\left(K_2 + 1 + 2\sum_{i=1}^{n} \left(\frac{x_i}{x}\right)^{-p-1}\right) \simeq \frac{\left(\frac{x_{n+1}}{x}\right)^{-p} + \left(\frac{x_n}{x}\right)^{-p}}{\left(\frac{x_{n+1}}{x}\right) - \left(\frac{x_n}{x}\right)}. \tag{9.25}$$

Let us name $H(x)$ the above l.h.s., including the function $(1+o(x))$ that is implicitly meant by the $\simeq$ symbol. The induction hypothesis implies that that $H(x)$ has a limit, as $x \to +\infty$. This limit, denoted $\widetilde{H}$, is evidently positive. After a little algebra (9.25) becomes

$$\left(\frac{x_{n+1}}{x}\right)^{-p} = H(x)\left(\frac{x_{n+1}}{x}\right) - H(x)\left(\frac{x_n}{x}\right) - \left(\frac{x_n}{x}\right)^{-p}. \tag{9.26}$$

Since all the terms except (x_{n+1}/x) are known have a limit, and since $\widetilde{H} > 0$, it easy to see that the relation above, regarded as an equation in the variable (x_{n+1}/x), tends to an equation that has only one solution. This must be $\xi_{n+1} := \lim_{x\to+\infty}(x_{n+1}(x)/x)$.

As concerns the second assertion of the lemma, we proceed by contradiction. Suppose that $\xi_n \nearrow \widetilde{\xi} < +\infty$, as $n \to +\infty$. Applying the first assertion to (9.25) one obtains

$$p\left(K_2 + 1 + 2\sum_{i=1}^{n} \xi_i^{-p-1}\right) = \frac{\xi_{n+1}^{-p} + \xi_n^{-p}}{\xi_{n+1} - \xi_n}. \tag{9.27}$$

Then the l.h.s. of (9.27) grows asymptotically like n. But the numerator of the r.h.s. converges, implying that $\xi_{n+1} - \xi_n \sim n^{-1}$, which in turn contradicts the convergence of $\{\xi_n\}$. Q.E.D.

We are just a step away from the proof of Lemma 9.2. In fact, for any $M > 0$ we can fix n such that $\xi_n > M + 1$. Then

$$\frac{\bar{\tau}_m(x)}{x} = \frac{1}{x} \sum_{i=0}^{m(x)} \tau_i(x) \geq \frac{1}{x} \sum_{i=0}^{n} \tau_i(x) > \frac{x_n(x) - x}{x} > M, \tag{9.28}$$

for x large enough. This means precisely that $\bar{\tau}_m(x)/x \to +\infty$, as $x \to +\infty$, implying Lemma 9.2. Q.E.D.

10 Acknowledgments

I am indebted to G. Del Magno, N. Chernov and M. Lyubich for many enlightening discussions on the subject. I also wish to thank C. Liverani, R. Markarian, N. Simányi, Ya. G. Sinai, D. Szász and L.-S. Young for their useful inputs.

During the preparation of this paper I was invited to visit the Institute of Mathematics of the Technical University of Budapest, and the Max Planck Institute for Mathematics in the Sciences, in Leipzig. I would like to thank both institutions for their nice hospitality.

Partial financial support from Università di Bologna, through S. Graffi, is also gratefully acknowledged.

A Appendix: Scattered lemmas

Lemma A.1 *Let $(\mathcal{M}, T, \mu)$ be a dynamical system with $\mu(\mathcal{M}) < \infty$, and ψ a positive function on $\mathcal{M}$. Then its ergodic average ψ^* is positive almost everywhere.*

Proof. First of all, ψ^* exists almost everywhere by the first Birkhoff Theorem. Set $A := \{z \in \mathcal{M} \mid \psi^*(z) = 0\}$; this is an invariant set. If we assume that $\mu(A) > 0$, then we can apply the second Birkhoff Theorem to (A, T, μ). We obtain

$$0 = \int_A \psi^* \, d\mu = \int_A \psi \, d\mu > 0, \tag{A.1}$$

which is absurd. Q.E.D.

Lemma A.2 *Let $\{g_n\}_{n\in\mathbb{N}}$ be a family of positive functions defined on a compact set K_0. Assume that g_n is continuous on a compact K_n, with $K_n \supseteq K_{n+1}$. If, furthermore, $\forall x \in K_0$, $g_n(x) \searrow 0$, as $n \to \infty$, then, in the same limit,* $\max_{K_n} g_n \searrow$ *0.*

PROOF. Let $x_n \in K_n$ be defined by $\max_{K_n} g_n = g_n(x_n)$. One checks that $\max_{K_n} g_n = g_n(x_n) \geq g_n(x_{n+1}) \geq g_{n+1}(x_{n+1}) = \max_{K_{n+1}} g_{n+1}$; the first inequality comes from the definition of x_n and the fact that $x_{n+1} \in K_n$, whereas the second is a result of the monotonicity of $\{g_n\}$.

Passing perhaps to a subsequence, $x_m \longrightarrow \bar{x} \in \bigcap_n K_n$, as $m \to \infty$. Suppose now that the assertion is false, that is, $\exists \varepsilon > 0$ such that $g_n(x_n) \geq \varepsilon$, for all n.

Let us consider the pointwise convergence $g_n(\bar{x}) \searrow 0$. There exist an N for which $g_N(\bar{x}) \leq \varepsilon/3$. Since x_m eventually belongs to K_N, and g_N is continuous there, $\exists M = M(N)$ such that, $\forall m \geq M$, $g_N(x_m) \leq 2\varepsilon/3$.

Now, if $N \geq M$, we reach a contradiction since, from above, $g_N(x_N) \leq 2\varepsilon/3$. If $N < M$, we use again the monotonicity of the family and see that $g_M(x_M) \leq g_N(x_M) \leq 2\varepsilon/3$. Q.E.D.

Lemma A.3 *For some $\eta > 1$, let $\mathcal{M}^n := [0, n^{-\eta}] \times [0,1]$ and let $\mathcal{R}^n \subset \mathcal{M}^n$ be the union of M_n graphs of monotonic functions over $[0, n^{-\eta}]$. Define $\mathcal{R} := \bigsqcup_{n \in \mathbb{N}} \mathcal{R}^n$ and denote by $\mathcal{R}_{\{\varepsilon\}}$ the tubular neighborhood of $\mathcal{R}$ of radius ε (with respect to the ordinary distance). If M_n grows at most polynomially in n, then* Leb $\left(\mathcal{R}_{\{\varepsilon\}}\right)$ *decays polynomially in ε, as $\varepsilon \to 0^+$.*

PROOF. Without loss of generality, one can assume that $\mathcal{R}^n$ is made up of vertical segments in $\mathcal{M}^n$ (in fact, the graph of a monotonic function in $\mathcal{M}^n$ has length less than $1 + n^{-\eta}$). The worst situation, in the sense of least amount of overlap among the tubular neighborhoods, occurs when these segments are equispaced. So, let us consider this case. Suppose that $M_n \leq C_1 n^\rho$. Then the spacing between the segments $\mathcal{R}^n$ is at least $n^{-(\eta+\rho)}/C_1$. For a given $\varepsilon > 0$, let k be the maximum n such that $n^{-(\eta+\rho)}/C_1 \geq 2\varepsilon$. For $n > k$, the measure of $\mathcal{R}^k$ is estimated by the measure of the entire $\mathcal{M}^k$. Then

$$\mathrm{Leb}\left(\bigsqcup_{n=k+1}^{\infty} \mathcal{R}^n_{\{\varepsilon\}}\right) \leq \sum_{n=k+1}^{\infty} n^{-\eta} \leq C_2 (k+1)^{-\eta+1} \leq C_3\, \varepsilon^{\frac{\eta-1}{\eta+\rho}}. \tag{A.2}$$

For $n \leq k$, we are guaranteed that the ε-neighborhoods of the segments do not overlap, therefore

$$\mathrm{Leb}\left(\bigsqcup_{n=1}^{k} \mathcal{R}^n_{\{\varepsilon\}}\right) \leq \varepsilon\, C_1 \sum_{n=1}^{k} n^\rho \leq C_4\, \varepsilon\, k^{\rho+1} \leq C_5\, \varepsilon^{1-\frac{\rho+1}{\eta+\rho}} = C_5\, \varepsilon^{\frac{\eta-1}{\eta+\rho}}. \tag{A.3}$$

The last two estimates prove the statement. Q.E.D.

Lemma A.4 *Let T be the billiard map associated to a table Q with finite horizon, i.e., $\tau(z) \leq \tau_M$, $\forall z$. Assume that the curvature $k(r)$ and its derivative are bounded above by k_M and k'_M, respectively. Set $JT_z := |DT_z v(z)|$, the Jacobian of T relative* *to the smooth direction field $v \subset \mathrm{S}\mathcal{M}$ (here $\mathrm{S}\mathcal{M}$ is the unit tangent bundle of $\mathcal{M}$; hence $|v(z)| = 1$). Then, for $b \in \mathrm{S}\mathcal{M}_z$,*

$$\frac{\partial}{\partial b} \log JT_z \le \frac{C}{\sin^2 \varphi \, \sin^4 \varphi_1},$$

with the notation of Section 3. The constant C depends only on τ_M, k_M, k'_M and $|(\partial v/\partial b)(z)|$.

Proof. Looking back at (3.5), we set $F(z) := \sin\varphi_1 \, DT_z$, so that the matrix elements of F contain no annoying denominators. By hypothesis, then,

$$\|F(z)\| \le C_1 = C_1(\tau_M, k_M), \tag{A.4}$$

with $\|\cdot\|$ denoting (only in this proof!) the norm of a matrix as a linear operator. Also by (3.5), we know that $\det DT_z = \sin\varphi / \sin\varphi_1$, whence $\det F = \sin\varphi \, \sin\varphi_1$. Therefore F is invertible and F^{-1} only contains denominators of the form $\sin\varphi \, \sin\varphi_1$. It follows that

$$\|F^{-1}\| \le \frac{C_1}{\sin\varphi \, \sin\varphi_1}, \tag{A.5}$$

which, in turn, gives

$$|Fv| \ge \frac{\sin\varphi \, \sin\varphi_1}{C_1}, \tag{A.6}$$

as $|v| = 1$. For the sake of format, let us indicate the partial derivative w.r.t. b with the symbol ∂_b. Via elementary calculus we obtain

$$\partial_b \log JT_z = \frac{DT_z v(z) \cdot \partial_b \, DT_z v(z)}{|DT_z v(z)|^2} = \sin\varphi_1 \, \frac{Fv \cdot \partial_b \left(\dfrac{Fv}{\sin\varphi_1} \right)}{|Fv|^2}. \tag{A.7}$$

We focus on the most troublesome term in (A.7):

$$\partial_b \left(\frac{Fv}{\sin\varphi_1} \right) = \frac{(\partial_b F)v}{\sin\varphi_1} + \frac{F(\partial_b v)}{\sin\varphi_1} + Fv \, \partial_b \left(\frac{1}{\sin\varphi_1} \right) =: Y_1 + Y_2 + Y_3. \tag{A.8}$$

We start with Y_2 which can be easily bounded above by $C_2 / \sin\varphi_1$, where $C_2 = C_2(C_1, |\partial_b v|)$. As for Y_3, we observe that $\partial_b \varphi_1$ is simply the second component of the vector $DT_z b$, whose norm, by (A.4) and the fact that $|b| = 1$, does not exceed $C_1 / \sin\varphi_1$. Hence, working out the other terms, we end up with $|Y_3| \le C_1^2 / \sin^3 \varphi_1$. In order to estimate Y_1, we notice that $F(z)$ is indeed a polynomial in the variables

$$\sin\varphi, \ \sin\varphi_1, \ k = k(r), \ k_1 = k(r_1), \ \tau = dist_Q(r, r_1). \tag{A.9}$$

Here $dist_Q(r, r_1)$ is the distance on the table Q between the points represented by the coordinates r and r_1. (Incidentally, let us observe that this function is smooth, for $r \ne r_1$, with derivatives bounded by 1 in absolute value.)

A given matrix element of $\partial_b F$ is then a finite sum of products of the variables in (A.9) times the derivative of *one* of those variables. Since the functions $\sin, k$ and $dist_Q$ have bounded derivatives, the only singularity will occur when, by implicit differentiation, ∂_b hits r_1 or φ_1. This corresponds to either component of $DT_z b$, and the differentiation gives rise to a singularity of the type $1/\sin\varphi_1$. Therefore there is a $C_3 = C_3(\tau_M, k_M, k'_M)$ such that $|Y_1| \le C_3/\sin^2\varphi_1$.

Taking the worst case among the above estimates, we conclude that there exists a $C_4 = C_4(\tau_M, k_M, k'_M, |\partial_b v|)$ such that

$$\partial_b \left(\frac{Fv}{\sin\varphi_1} \right) \le \frac{C_4}{\sin^3\varphi_1}. \tag{A.10}$$

Plugging (A.4), (A.6) and (A.10) into (A.7) gives the assertion of the lemma. Q.E.D.

Lemma A.5 *If A is a countable subset of $\mathcal{M} := I \times J$, with I, J two non-degenerate intervals, then $\mathcal{M} \setminus A$ is path-connected.*

PROOF. Take z_1, z_2, two distinct elements of $\mathcal{M} \setminus A$, and let B denote the segment in $\mathcal{M}$ whose points are equidistant from z_1 and z_2. For $w \in B$, call γ_w the polyline connecting z_1 to w and w to z_2. $\{\gamma_w\}$ is an uncountable family of pairwise disjoint paths. Since only countably many of them can intersect A, it follows that there exist infinitely many paths connecting z_1 to z_2. Q.E.D.

Lemma A.6 *Let $(\mathcal{M}, T, \mu)$ be a recurrent dynamical system (with possibly $\mu(\mathcal{M}) = \infty$). For every measurable function ψ with the property that the forward and backward time averages,*

$$\psi^{\pm}(z) := \lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\psi \circ T^{\pm k})(z),$$

exist μ-a.e. in $\mathcal{M}$, one has

$$\psi^+(z) = \psi^-(z)$$

μ-a.e. in $\mathcal{M}$.

PROOF. Set

$$\psi_j^{\pm}(z) := \frac{1}{j} \sum_{k=0}^{j-1} (\psi \circ T^{\pm k})(z) \tag{A.11}$$

Also, for $\varepsilon > 0$ and $n \in \mathbb{N}$, denote

$$A^{\pm}_{\varepsilon,n} := \left\{ z \in \mathcal{M} \;\middle|\; \exists \psi^{\pm}(z);\ \left| \psi_j^{\pm}(z) - \psi^{\pm}(z) \right| \le \varepsilon,\ \forall j \ge n \right\}. \tag{A.12}$$

By hypothesis, the finite averages converge a.e. Therefore, as $n \to +\infty$, $A^{\pm}_{\varepsilon,n} \nearrow \mathcal{M}$ mod μ. Thus, almost every $z \in \mathcal{M}$ belongs to some set $A^+_{\varepsilon,n^+} \cap A^-_{\varepsilon,n^-}$. Also, a.e. z in this set is recurrent there. Take any such z; in particular $\exists n \geq \max\{n^+, n^-\}$ such that

$$T^n z \in A^-_{\varepsilon,n^-}. \tag{A.13}$$

Evidently, $\psi^+_n(z) = \psi^-_n(T^n z)$. This, together with (A.13) and the fact that $z \in A^+_{\varepsilon,n^+}$, yields

$$\left|\psi^+(z) - \psi^-(z)\right| = \left|\psi^+(z) - \psi^-(T^n z)\right| \leq 2\varepsilon, \tag{A.14}$$

having used the invariance of ψ^- as well. The above holds for a.a. $z \in \mathcal{M}$. If we take a countable vanishing sequence of ε's, we get the desired result. Q.E.D.

Lemma A.7 *Consider the finite segment of trajectory* $\{z_0 = z, z_1 = Tz, z_2 = T^2 z\}$ *of a billiard map* T*. Then, at the first order in* dz*, and up to a minus sign,* dz_2 *is given by a single differential a billiard map, with dynamical parameters:*
Free path: $\hat{\tau} = \tau_0 + \tau_1 + 2\dfrac{k_1}{\sin\varphi_1}\tau_0\tau_1$*;*
Curvature at initial point z_0*:* $\hat{k}_0 = k_0 + 2\sin\varphi_0 \dfrac{k_1}{\sin\varphi_1}\dfrac{\tau_1}{\hat{\tau}}$*;*
Curvature at final point z_2*:* $\hat{k}_2 = k_2 + 2\sin\varphi_2 \dfrac{k_1}{\sin\varphi_1}\dfrac{\tau_0}{\hat{\tau}}$*.*

In other words, $DT_{z_1}DT_{z_0} = -M$*, where* $M(\hat{\tau}, \hat{\varphi}_0, \hat{k}_0, \hat{\varphi}_2, \hat{k}_2)$ *is again a differential of the type* (3.5)*. If we take the same initial and final angles as for the actual segment of trajectory,* $\hat{\varphi}_0 = \varphi_0, \hat{\varphi}_2 = \varphi_2$*, then the other three parameters are fixed as above.*

PROOF. This is just a verification, which is made easier if we use on $\mathrm{T}\mathcal{M}$ the pair of variables $(\sin\varphi\, dr, d\varphi)$. Denoting by $\bar{D}T$ the differential in these new variables, we get from (3.5)

$$\bar{D}T_{z_i} = \begin{bmatrix} -1 - \eta_i\tau_i & \tau_i \\ \eta_i + \eta_{i+1} + \eta_i\eta_{i+1}\tau_i & -1 - \eta_{i+1}\tau_i \end{bmatrix}, \tag{A.15}$$

with $\eta_i := k_i / \sin\varphi_i$. Notice that $\det \bar{D}T_{z_i} = 1$. Hence, upon imposing

$$\bar{D}T_{z_1}\bar{D}T_{z_0} =: - \begin{bmatrix} -1 - \hat{\eta}_0\hat{\tau} & \hat{\tau} \\ \hat{\eta}_0 + \hat{\eta}_2 + \hat{\eta}_0\hat{\eta}_2\hat{\tau} & -1 - \hat{\eta}_2\hat{\tau} \end{bmatrix}, \tag{A.16}$$

and $\hat{k}_i = \hat{\eta}_i \sin\varphi_i$ $(i = 0, 2)$, we get the three parameters as in the statement of the lemma. Q.E.D.

References

[BCSzT] P. Bálint, N. I. Chernov, D. Szász and I. P. Tóth, *Geometry of multi-dimensional dispersing billiards*, preprint.

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

[BS] L. A. Bunimovich and Ya. G. Sinai, *On a fundamental theorem in the theory of dispersing billiards*, Math. USSR Sbornik **19**(3) (1973), 407-423.

[C1] N. I. Chernov, *On local ergodicity in hyperbolic systems with singularities*, Funct. Anal. Appl. **27**(1) (1993), 51-54.

[C2] N. I. Chernov, *Decay of correlations and dispersing billiards*, J. Statist. Phys. **94**(3-4) (1999), 513-556.

[CM] N. I. Chernov and R. Markarian, *Billiards with polynomial decay of correlations*, in progress.

[Co] J.-P. Conze, *Sur un critère de récurrence en dimension 2 pour les marches stationnaires, applications*, Ergodic Theory Dynam. Systems **19**(5) (1999), 1233-1245.

[CFS] I. P. Cornfeld, S. V. Fomin and Ya. G. Sinai, *Ergodic theory*, Springer-Verlag, New York, 1982.

[DDL1] M. Degli Esposti, G. Del Magno and M. Lenci, *An infinite step billiard*, Nonlinearity **11**(4) (1998), 991-1013.

[DDL2] M. Degli Esposti, G. Del Magno and M. Lenci, *Escape orbits and ergodicity in infinite step billiards*, Nonlinearity **13**(4) (2000), 1275-1292.

[G] G. Gallavotti, *Lectures on the billiard*, in: Dynamical systems, theory and applications, J. Moser (ed.), LNP 38, Springer-Verlag, Berlin-Heidelberg, 1975.

[GO] G. Gallavotti and D. S. Ornstein, *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83-101.

[KH] A. Katok and B. Hasselblatt, *Modern theory of dynamical systems*, Cambridge University Press, Cambridge, 1995.

[KS] A. Katok and J.-M. Strelcyn (in collaboration with F. Ledrappier and F. Przytycki), *Invariant manifolds, entropy and billiards; smooth maps with singularities*, LNM 1222, Springer-Verlag, Berlin-New York, 1986.

[K] J. L. King, *Billiards inside a cusp*, Math. Intelligencer **17**(1) (1995), 8-16.

[KSSz] A. Krámli, N. Simányi and D. Szász, *A "transversal" fundamental theorem for semi-dispersing billiards*, Comm. Math. Phys. **129**(3) (1990), 535-560.

[Le1] M. Lenci, *Escape orbits for non-compact flat billiards*, Chaos **6**(3) (1996), 428-431.

[Le2] M. Lenci, *More ergodic billiards with an infinite cusp*, preprint, downloadable from `www.arxiv.org/abs/nlin.CD/0201052`.

[L] A. M. Leontovich, *The existence of unbounded oscillating trajectories in a problem of billiards*, Soviet Math. Dokl. **3**(4) (1962), 1049-1052.

[LW] C. Liverani and M. Wojtkowski, *Ergodicity in Hamiltonian systems*, in: Dynamics Reported: Expositions in Dynamical Systems (N.S.), **4**, Springer-Verlag, Berlin, 1995.

[LM] A. Lopez and R. Markarian, *Open billiards: invariant and conditionally invariant probabilities on Cantor sets*, SIAM J. Appl. Math. **56**(2) (1996), 651-680.

[M1] R. Markarian, *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118**(1) (1988), 87-97.

[M2] R. Markarian, *Statistical properties of dynamical systems with singularities*, J. Statist. Phys. **80**(5-6) (1995), 1207-1239.

[P] Ya. B. Pesin, *Lyapunov characteristics exponents and smooth ergodic theory*, Russ. Math. Surveys **32**(4) (1977), 55-114.

[PS] C. Pugh and M. Shub, *Ergodic attractors*, Trans. Am. Math. Soc. **312**(1) (1989), 1-54.

[Sch] K. Schmidt, *On joint recurrence*, Comptes Rendu Acad. Sci. Paris Sér. I Math. **327**(9) (1998), 837-842.

[Si] N. Simányi, *Towards a proof of recurrence for the Lorentz process*, in: Dynamical Systems and Ergodic Theory, Banach Center Publications, Vol. 23, PWN, Polish Scientific Publishers, Warsaw, 1989.

[S1] Ya. G. Sinai, *On the foundation of the ergodic hypothesis for a dynamical system of statistical mechanics*, Soviet Math. Dokl. **4** (1963), 1818-1822.

[S2] Ya. G. Sinai, *Dynamical systems with elastic reflections*, Russ. Math. Surveys **25** (1970), 137-189.

[SC] Ya. G. Sinai and N. I. Chernov, *Ergodic properties of certain systems of two-dimensional discs and three-dimensional balls*, Russ. Math. Surveys **42**(3) (1987), 181-207.

[V] L. N. Vaserstein, *On systems with finite-range and/or repulsive interactions*, Comm. Math. Phys. **69**(1) (1979), 31-56.

[W] M. Wojtkowski, *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Comm. Math. Phys. **105** (1986), 391-414.

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