Statistical properties of the system of two falling balls

Péter Bálint, Gábor Borbély, and András Némedy Varga
Department of Differential Equations, Institute of Mathematics,
Budapest University of Technology and Economics
Egry József u. 1, H-1111 Budapest, Hungary
pet,borbely,nemedy@math.bme.hu

November 28, 2011

Abstract

We consider the motion of two point masses along a vertical half-line that are subject to constant gravitational force, and collide elastically with each other and the floor. This model was introduced by Wojtkowski who established hyperbolicity and ergodicity in case the lower ball is heavier. Here we investigate the dynamics in discrete time and prove that, for an open set of the external parameter (the mass ratio of the two balls) the system mixes polynomially (modulo logarithmic factors, correlations decay as $\mathcal{O}(1/n^2)$) and satisfies the Central Limit Theorem.

Contents

Lead paragraph

Dynamical systems with intermittency – when the evolution alternates between chaotic and regular patterns – are popular models for describing a wide range of physical phenomena, and thus have been in the focus of active investigation for several decades ([23], [21], [20]) . Of particular relevance are systems in which the chaotic component is strong enough to ensure ergodicity with respect to a natural invariant measure, nonetheless, there is a significant regular component which makes the dynamics only non-uniformly hyperbolic. This often results in slower – polynomial, in contrast to exponential – rates in the decay of correlations. Another question of high importance is that of statistical limit laws, in particular, whether a natural class of observables exhibits standard or non-standard limit theorems (in other words, normal vs. anomalous diffusion).

Despite of their significance, the class of systems for which detailed, and mathematically rigorously justified information is available on the above mechanism is quite limited. The best understood models are one dimensional maps with neutral fixed points ([21], [17], [9], [25], [31]). There is quite much known about rates of mixing and statistical limit laws in certain two dimensional billiard systems of physics origin, in particular stadia ([15], [3]), Lorentz gases with infinite horizon ([33], [11]) and dispersing billiards with cusps ([13], [2]).

In the present paper we investigate another billiard type model with intermittent behavior. The system of two falling balls, introduced by Wojtkowski in [34], describes the motion of two point particles of mass m_1 and m_2 that move along the vertical half-line, subject to constant gravitational force, and collide elastically with each other and the floor. We consider the case when the lower ball is heavier (i.e. $m_1 < m_2$) which corresponds to ergodic and hyperbolic dynamics; while the regular component of intermittency is related to arbitrary long series of bounces of the lower ball on the floor before hitting the upper ball. We present a detailed analysis of this model and prove that correlations decay, modulo logarithmic factors, as $\mathcal{O}(1/n^2)$. This rate is summable, accordingly, the central limit theorem is also proved; that is, the system exhibits normal diffusion.

1 Setting

1.1 Description of the dynamical system

Historical background. The system of falling balls can be regarded as a billiard. The table is one dimensional: a vertical half line bounded from below. In this line infinitesimally small balls move up and down under a constant gravity force – throughout the paper, g denotes the corresponding standard gravity acceleration. The balls bounce and collide totally elastically with each other and with the floor. Our system has only two particles of mass m_1 and m_2.

This system can be regarded both in continuous and in discrete time (see below for further details); the corresponding flow preserves a natural measure (the Liouville measure) which projects to an absolutely continuous invariant measure for the Poincaré map. As in the literature, ergodicity is considered with respect to this measure, while hyperbolicity means that all non-trivial Lyapunov exponents are non zero almost everywhere, again with respect to the natural measure. The model was first studied by Wojtkowski in [34], actually, in the more general context of n balls of masses $m_i, i = 1 \to n$. He proved hyperbolicity (i.e non-vanishing of all relevant Lyapunov exponents) if the masses strictly decrease up the line, i.e. $m_1 > m_2 > \ldots > m_n$. In [32] Simányi established hyperbolicity under a weaker assumption: if the masses decrease non-strictly up the line, but there are at least two different masses. Hence hyperbolicity is well understood, ergodicity is, however, still an open problem for three or more particles.

The system of two falling balls is known to be ergodic if $m_1 > m_2$, i.e. when the lower ball is heavier ([22]). If $m_1 = m_2$, the balls exchange their velocity at each collision which makes the system integrable;

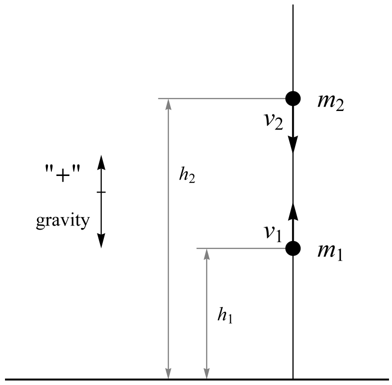


Figure 1: The system of two falling balls

while an elliptic periodic orbit can be observed if $m_1 < m_2$ (see [34]).

The dynamical system. From now on we restrict our analysis to the case of two balls with ergodic dynamics, i.e. $m_1 > m_2$. Let us express the state of the system with the usual physical quantities: h_1, h_2 are the height of the lower and the upper ball, v_1, v_2 are the velocities. We neglect the air resistance therefore the total energy of the system $J := \frac{1}{2}m_1v_1^2 + m_1gh_1 + \frac{1}{2}m_2v_2^2 + m_2gh_2$ is an integral of motion. This motivates introducing the phase space:

$$\tilde{\mathcal{M}} := \left\{(h_1, v_1, h_2, v_2) \in \mathbb{R}^4 | 0 < h_1 < h_2, \frac{1}{2}m_1v_1^2 + m_1gh_1 + \frac{1}{2}m_2v_2^2 + m_2gh_2 = J\right\}$$

The dynamics S^t act on $\tilde{\mathcal{M}}$ in continuous time and preserves the (normalized) Lebesgue measure (the Liouville measure) $\tilde{\mu}$ on this three dimensional Riemannian manifold.

Following [34] we introduce the outgoing Poincaré section to this flow corresponding to the moments when the lower ball hits the floor:

$$\mathcal{M} = \left\{(h_1, v_1, h_2, v_2) \in \tilde{\mathcal{M}} | h_1 = 0, v_1 > 0\right\}$$

This way we obtain a discrete time map of a two dimensional phase space $\hat{T} : \mathcal{M} \mapsto \mathcal{M}$.

Instead of the usual moments we use the following coordinates of $\mathcal{M}$ (also from [34]):

$$h := \frac{1}{2}m_1v_1^2 \qquad \text{the total energy of the lower ball, since } h_1 = 0$$

$$z := v_2 - v_1$$

These coordinates seem to be suitable because they make our formulas simpler. It is also interesting to see that these quantities are invariants during the inter-collisional motion. Now the phase space is:

$$\mathcal{M} := \left\{(h, z) \in \mathbb{R}^2 | \, (0 < h < J) \wedge \left(J - h > \frac{1}{2}m_2{v_2}^2\right)\right\}$$

where v_2 can be expressed from our coordinates: $v_2 = v_1 + z = \sqrt{\frac{2h}{m_1}} + z$.

Piecewise smoothness. $\hat{T}$ is only piecewise continuous; there are two possible scenarios; the two balls either collide or not before the lower one returns to the floor. Let

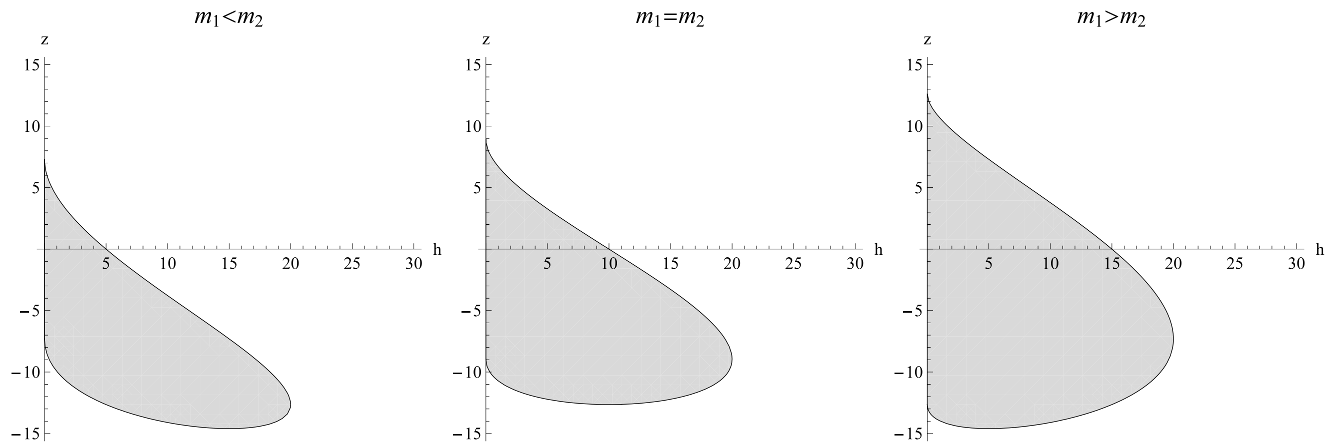


Figure 2: The phase space for different masses with $J = 20$

$$\mathcal{M} \supseteq \mathcal{M}_1 = \{\text{configurations in which the balls will collide}\}$$
$$\mathcal{M} \supseteq \mathcal{M}_2 = \{\text{configurations in which the balls won't collide}\}$$
$$(h', z') = \hat{T}(h, z) = \begin{cases} F_1(h, z) & \text{if } (h, z) \in \mathcal{M}_1 \\ F_2(h, z) & \text{if } (h, z) \in \mathcal{M}_2 \end{cases}$$

The action of $\hat{T}$ is denoted by prime

To determine the sets $\mathcal{M}_i$ and the smooth maps $F_i : \mathcal{M}_i \mapsto \mathcal{M}$, pretend for a while that the two balls move independently. To fall back to the floor the lower ball would take $t = 2\frac{v_1}{g}$ time, meanwhile the upper ball would reach the height $h_t = h_0 + v_2 t - \frac{1}{2}gt^2$ where h_0 is the starting height, which can be calculated from the potential energy of the upper ball: $h_0 = \frac{J-h-\frac{1}{2}m_2 v_2^2}{gm_2}$. After the substitutions:

$$h_t = \frac{m_2 m_1 z \left(2\sqrt{\frac{2h}{m_1}} - z\right) - 2h\,(m_1 + m_2) + 2Jm_1}{2gm_1 m_2} \tag{1.1}$$

Whether the balls collide or not is determined by the sign of h_t. Now the sets can be characterized:

$$\mathcal{M}_1 := \{(h, z) \in \mathcal{M} | h_t < 0\}; \qquad \mathcal{M}_2 := \{(h, z) \in \mathcal{M} | h_t > 0\}.$$

Note that the boundary case $h_t = 0$ is an event of zero Lebesgue measure, and no matter how we define the map $\hat{T}$ in this non-typical case, it does not effect the statistical properties of the system.

As for the maps F_i, it is easier to express F_2, when the balls do not collide: the individual energies are conserved, so $h' = h$, and consequently ${v_1}' = v_1$. Since the balls accelerate equally ${v_2}' = v_2 - 2v_1$ hence $z' = {v_2}' - {v_1}' = (v_2 - 2v_1) - v_1 = z - 2v_1$. This way:

$$\begin{pmatrix} h' \\ z' \end{pmatrix} = F_2(h, z) = \begin{pmatrix} h \\ z - 2v_1 \end{pmatrix} = \begin{pmatrix} h \\ z - 2\sqrt{\frac{2h}{m_1}} \end{pmatrix}$$

To determine F_1 we express, in terms of h and z: the time when the particles reach the same height (and also the height itself), count the velocities in that moment, apply the rules of an elastic collision and calculate the additional time needed for the lower particle to hit the floor again. The values of h' and z'

can be determined from the new velocities and kinetic energies. The result of this calculation is shown in the formula (1.2). For simplicity from now on we make the assumption: $m_1 + m_2 = 1$ because these parameters only scale the system; let, furthermore, $J = \frac{1}{2}, m_1 = m$ and $m_2 = 1 - m$ where $\frac{1}{2} < m < 1$ (ergodic case).

$$\begin{pmatrix} h' \\ z' \end{pmatrix} = F_1(h,z) = \begin{pmatrix} m\left(1 + \left(1 - 3m + 2m^2\right) z^2\right) - h \\ -\sqrt{8}\sqrt{-\frac{h}{m} + 1 + \left(1 - 3m + 2m^2\right) z^2} - z \end{pmatrix}. \tag{1.2}$$

Non-uniform hyperbolicity. The Jacobian of the maps can be calculated:

$$\mathrm{D}\, F_1(h,z) = \begin{pmatrix} -1 & 2m\alpha z \\ \frac{\sqrt{2}}{m\sqrt{1-\frac{h}{m}+\alpha z^2}} & -1 - \frac{2\sqrt{2}\alpha z}{\sqrt{1-\frac{h}{m}+\alpha z^2}} \end{pmatrix}$$

$$\mathrm{D}\, F_2(h,z) = \begin{pmatrix} 1 & 0 \\ -\sqrt{\frac{2}{hm}} & 1 \end{pmatrix}$$

where $\alpha := (1-m)((1-m)-m) = 1 - 3m + 2m^2$. The sign of α is determined by the mass ratio of the balls: if $m_1 > m_2$ then $\alpha < 0$, if $m_1 = m_2$, then $\alpha = 0$, otherwise $\alpha > 0$. We are interested in the case of negative α because this corresponds to ergodic dynamics.

As both matrices have determinant 1, the normalized *Lebesgue measure*, to be denoted by $\hat{\mu} := \frac{\lambda}{\lambda(\mathcal{M})}$ is preserved by the map $\hat{T}$. By ergodicity ([22]) $\hat{\mu}$ coincides with the natural measure, induced by the Liouville measure $\tilde{\mu}$ of the flow S^t.

There exists a constant, invariant (unstable) cone field of our system. Note that our phase space is embedded in $\mathbb{R}^2$ hence the tangent bundle is trivial; for any $x \in \mathcal{M}$, $\mathcal{T}_x\mathcal{M}$ can be identified with $\mathbb{R}^2$. Let $\mathcal{C}^u_x := \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \cdot v_2 \leq 0\}$, that is, the union of the lower, right quarter and the upper, left quarter of the plane.

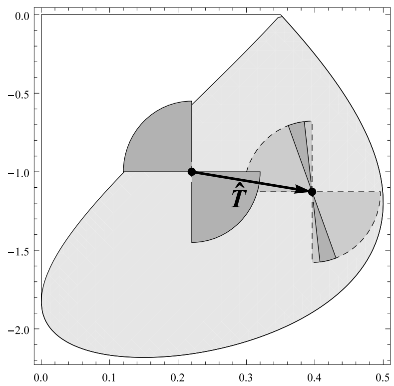


Figure 3: The invariance of the cone field under F_1 $(\mathcal{M}_1)$

Let $\underline{v} = (v_1, v_2) \in \mathcal{C}^u_x$ be a vector in the cone. The invariance of the cone field was proven in [34], but it also can be checked easily that the following vectors have components of opposite sign (hence belong to $\mathcal{C}^u_x$):

$$\mathrm{D}\, F_1 \cdot \underline{v} = \begin{pmatrix} 2mv_2 z\alpha - v_1 \\ \frac{\sqrt{2}(v_1 - 2mv_2 z\alpha)}{m\sqrt{-\frac{h}{m}+z^2\alpha+1}} - v_2 \end{pmatrix} \qquad \mathrm{D}\, F_2 \cdot \underline{v} = \begin{pmatrix} v_1 \\ v_2 - \frac{\sqrt{2}v_1}{\sqrt{hm}} \end{pmatrix}$$

On the other hand, the cones are not mapped strictly inside themselves, the vertical direction is fixed under F_2.

$$\begin{pmatrix} 1 & 0 \\ -\sqrt{\frac{2}{hm}} & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \tag{1.3}$$

Further details about hyperbolicity are given in section 3.3.

The first return map and the singularity stripes. The important consequence of (1.3) is that the map $\hat{T}$ is only non-uniformly hyperbolic. To obtain uniformly hyperbolic dynamics we consider the first return map to the set $\mathcal{M}_1$. Physically this means that we wait while the lower ball bounces on the floor until it collides with the upper ball. We introduce the following notations:

$n_* : \mathcal{M}_1 \mapsto \mathbb{N}$

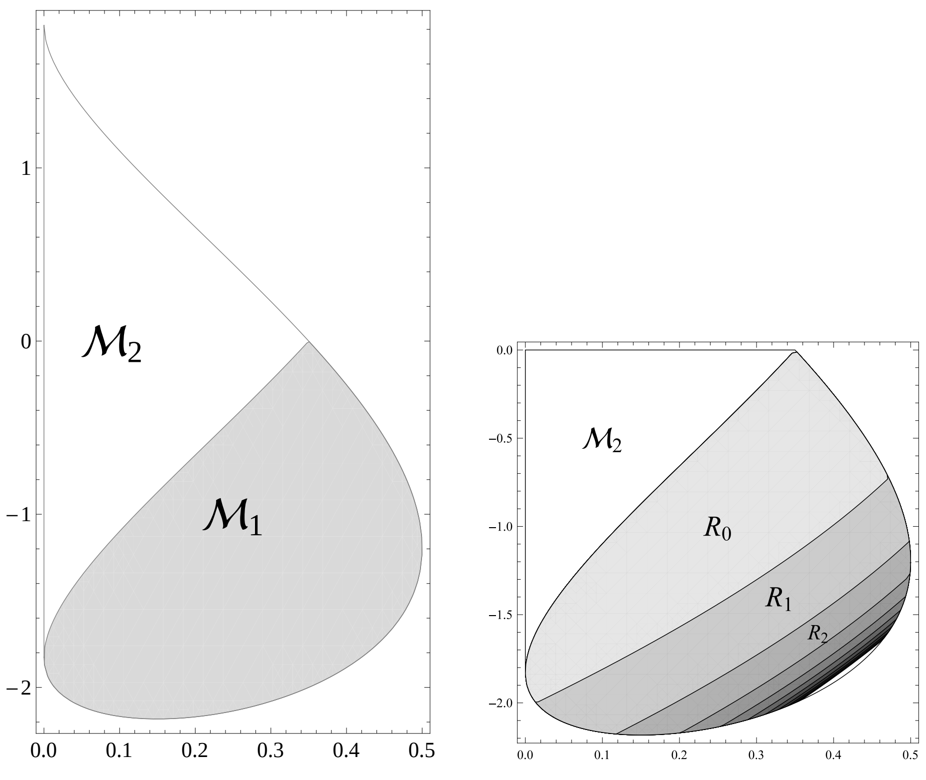
$n_x := \min\{n \in \mathbb{N} | F_2{}^n (F_1(x)) \in \mathcal{M}_1\}$ — the number of iterations needed to return

$R_n := \{x \in \mathcal{M}_1 | n_x = n\}$ — the sets where the recurrence time is constant

$T : \mathcal{M}_1 \mapsto \mathcal{M}_1$

$Tx := F_2{}^{n_x} (F_1(x))$

The invariant measure for T is the normalized Lebesgue measure on $\mathcal{M}_1$ (a conditional probability), denoted by μ.



(a) The sets $\mathcal{M}_1$ and $\mathcal{M}_2$

(b) The sets R_n of constant first return time

Figure 4: $m = 0.7$

Starting from the set R_n, the lower ball collides with the upper ball, then hits the floor $n + 1$ times until it collides again with the upper one. This phenomenon hastens the decay of correlations in the new system and also causes bigger rates of expansion in the Lyapunov exponents.

1.2 Statement of results and structure of proofs

Results. In this paper we consider decay of correlations and statistical limit laws for the flow $(\tilde{M}, S^t, \tilde{\mu})$ and the map $(\mathcal{M}, \hat{T}, \hat{\mu})$. More precisely, both the phase space $\mathcal{M}$ and the dynamics $\hat{T}$ acting on it depend on the parameter $m \in (0,1)$ (the mass ratio of the two balls): we prove that there exists an open set $I \subset (\frac{1}{2}, 1)$ such that for $m \in I$ correlations decay polynomially and the central limit theorem applies (see Theorems 1.1, 1.2, 1.3 below).

An observable $f : \mathcal{M} \to \mathbb{R}$ is Hölder continuous if there are constants $\gamma \in (0,1)$ and $C > 0$ such that $|f(x) - f(y)| \leq C d(x,y)^\gamma$, where $d(x,y)$ is the Riemannian distance on $\mathcal{M}$; furthermore, f has zero mean if $\int_{\mathcal{M}} f(x) d\hat{\mu}(x) = 0$. As the maps $\hat{T}$ and T are only piecewise continuous, it is natural to consider a slightly larger class of functions, for which discontinuities are allowed as long as they coincide with those of the dynamics. An observable $f : \mathcal{M} \to \mathbb{R}$ is *uniformly piecewise Hölder continuous* if it is Hölder continuous when restricted to the subsets $\mathcal{M}_1$ and $\mathcal{M}_2$. Piecewise Hölder continuous functions on the flow phase space $\tilde{\mathcal{M}}$ can be defined analogously. Little more care is needed when considering observables for the first return map as there are infinitely many pieces; $g : \mathcal{M}_1 \to \mathbb{R}$ is uniformly piecewise Hölder continuous if $g|_{R_k}$ is Hölder continuous with the same $\gamma \in (0,1)$ and $C > 0$ for any $k \geq 0$. The class of uniformly piecewise Hölder continuous observables on $\mathcal{M}$ without or with zero mean will be denoted by $\mathcal{H}(\mathcal{M})$ and $\mathcal{H}_0(\mathcal{M})$, respectively; the notations $\mathcal{H}(\tilde{\mathcal{M}}), \mathcal{H}_0(\tilde{\mathcal{M}}), \mathcal{H}(\mathcal{M}_1)$ $\mathcal{H}_0(\mathcal{M}_1)$ should be interpreted analogously.

Our first result is on the decay of correlations for (uniformly piecewise) Hölder continuous observables. Given $f, g \in \mathcal{H}_0(\mathcal{M})$, for $n \in \mathbb{N}$ the n-time correlation of f and g is defined as $Corr_n(f,g) = \int_{\mathcal{M}} f(x) g(\hat{T}^n x) d\hat{\mu}(x)$.

Theorem 1.1. *There exists an open set $I \subset (\frac{1}{2}, 1)$ such that for $m \in I$ the map $(\mathcal{M}, \hat{T}, \hat{\mu})$ has polynomial decay of correlations. More precisely, for any $f, g \in \mathcal{H}_0(\mathcal{M})$ there exists a positive constant $C = C(f,g)$ such that*

$$|Corr_n(f,g)| \leq C \frac{(\log n)^3}{n^2}.$$

Our next result is on the central limit theorem for Birkhoff sums of (uniformly piecewise) Hölder continuous observables. For $f \in \mathcal{H}_0(\mathcal{M})$, let $S_n f(x) = f(x) + f(\hat{T}x) + ... + f(\hat{T}^{n-1}x)$ denote the Birkhoff sum of f. By the Birkhoff ergodic theorem and ergodicity of $\hat{T}$, $\lim_{n\to\infty} \frac{S_n f(x)}{n} = 0$ for $\hat{\mu}$-a.e. $x \in \mathcal{M}$.

Theorem 1.2. *There exists an open set $I \subset (\frac{1}{2}, 1)$ such that for $m \in I$ the map $(\mathcal{M}, \hat{T}, \hat{\mu})$ satisfies the central limit theorem. That is, for $f \in \mathcal{H}_0(\mathcal{M})$, there exists $\sigma_f \geq 0$ such that the sequence $\frac{S_n f(x)}{\sqrt{n}}$ converges in distribution to $\mathcal{N}(0, \sigma_f)$, the normal law with zero mean and variance σ_f. Here σ_f can be calculated from the autocorrelations of f and we have $\sigma_f = 0$ if and only if $f = g - g \circ \hat{T}$ for some square integrable function g (see eg. [12], Chapter 7).*

In order to handle the issue of the central limit theorem for the flow $(\tilde{\mathcal{M}}, S^t, \tilde{\mu})$, we consider two observables of particular interest, the first return times of the flow to the Poincaré sections $\mathcal{M}$ and $\mathcal{M}_1$, respectively:

- for $x \in \mathcal{M}$, let $\hat{\tau}(x) = \min\{t > 0 | S^t x \in \mathcal{M}\}$, and let $\bar{\hat{\tau}} = \int_{\mathcal{M}} \hat{\tau}(x) d\hat{\mu}(x)$;
- for $x \in \mathcal{M}_1$, let $\tau(x) = \min\{t > 0 | S^t x \in \mathcal{M}_1\}$, and let $\bar{\tau} = \int_{\mathcal{M}_1} \tau(x) d\mu(x)$.

Given any flow observable $F : \tilde{\mathcal{M}} \to \mathbb{R}$, there is a canonically associated map observable $f : \mathcal{M} \to \mathbb{R}$ defined by the formula $f(x) = \int_0^{\hat{\tau}(x)} F(S^t x) dt$; that is, the values of f are obtained by integrating the values of F up to the first return.

Now consider an observable $F \in \mathcal{H}_0(\tilde{\mathcal{M}})$ and $T > 0$, and let $S_T F(x) = \int_0^T F(S^t x)dt$; by ergodicity, $\lim_{T\to\infty} \frac{S_T F(x)}{T} = 0$ for $\tilde{\mu}$-a.e. $x \in \tilde{\mathcal{M}}$.

Theorem 1.3. *There exists an open set $I \subset (\frac{1}{2}, 1)$ such that for $m \in I$ the flow $(\tilde{\mathcal{M}}, S^t, \tilde{\mu})$ satisfies the central limit theorem. That is, for $F \in \mathcal{H}_0(\tilde{\mathcal{M}})$, there exists $\sigma_F \geq 0$ such that $\frac{S_T F(x)}{\sqrt{T}}$ converges in distribution to $\mathcal{N}(0, \sigma_F)$. Moreover, for the associated map observable we have $f \in \mathcal{H}_0(\mathcal{M})$, hence Theorem 1.2 applies, and we have $\sigma_F = \sigma_f / \bar{\hat{\tau}}$.*

Remark 1.4. *As exposition relies on modeling $\hat{T}$ with a Young tower and S^t as a suspension flow on it (see below for details), we automatically establish all the results that are available in this setting. In particular, both the map and the flow exhibit normal diffusion in the sense of the almost sure invariance principle which can be regarded as a stronger version of the central limit theorem (see [27] for details). Results on large deviations ([25]) and convergence of moments ([30]) also apply.*

Structure of proofs. An efficient way of investigating decay of correlations and statistical limit laws in hyperbolic systems is by the method of Young towers. By Young tower we mean the scheme developed in [35]: roughly speaking, the existence of a positive Lebesgue measure set Δ with hyperbolic product structure (the base of the tower), and a partition $\Delta = \cup_i \Delta_i$ such that the points of Δ_i return, in an appropriate sense, to Δ simultaneously. As proved in [35] and [36], the return time statistics (with respect to Lebesgue measure) of the Δ_i essentially determine the rate of mixing of the map (see below). Throughout, we will use the usual terminology: we will say that a hyperbolic map $\mathcal{T} : M \to M$ *can be modeled by a Young tower with exponential(polynomial) tails*, whenever there exists $\Delta \subset M$ with certain (exponential, polynomial...) return time statistics. For further details on the general method of Young towers we refer to the original references [35] and [36]. Actually, we do not work directly with Young towers in the present paper, instead, we refer to [15]. In this paper (see also [10] for an earlier version) a set of assumptions is given, designed specifically for two dimensional uniformly hyperbolic maps with singularities; whenever a dynamical system satisfies the assumptions, it admits a Young tower with exponential tails. For the readers convenience, we collect these assumptions in the Appendix. One of the two main ingredients in the proof of Theorems 1.1 and 1.2 is the following proposition.

Proposition 1.5. *For $m \in I$, the first return map $(\mathcal{M}_1, T, \mu)$ satisfies assumptions A0-A7 from the Appendix. Consequently:*

- *by [15], $(\mathcal{M}_1, T, \mu)$ can be modeled by a Young tower with exponential tails, hence*
- *by [35], $(\mathcal{M}_1, T, \mu)$ has exponential decay of correlations, and satisfies the central limit theorem.*

However, we are mainly interested in the statistical properties of the original map $(\mathcal{M}, \hat{T}, \hat{\mu})$. At this point we refer one more time to [15], more precisely, to the following proposition (see also [24] for an earlier version).

Proposition 1.6 (Theorem 4. from [15]). *Given a general hyperbolic map $\mathcal{T} : M \to M$ preserving an ergodic absolutely continuous invariant measure ν, and a set M_1 of positive ν measure, let $\mathcal{T}_1 : M_1 \to M_1$ denote the first return map, ν_1 the induced invariant measure, and $R : M_1 \to \mathbb{N}$ the first return time. Introduce, furthermore, the sets $\mathcal{R}_n = \{x \in M_1 : R(x) = n\}$ of constant first return time in M_1. Assume that*

- *the first return map $(M_1, \mathcal{T}_1, \nu_1)$ can be modeled by a Young tower with exponential tails;*
- *there exist constants $b > 0$ and $C > 0$ such that $\nu_1(\mathcal{R}_n) \leq Cn^{-b-2}$.*

Then

- *the original map* $(M, \mathcal{T}, \nu)$ *can be modeled by a Young tower with polynomial tails; more precisely, the tail distribution is bounded by* $C(\log n)^{b+1} \cdot n^{-b}$*;*

- *consequently, by [36] (see also [9] and [28])* $(M, \mathcal{T}, \nu)$ *has polynomial decay of correlations, more precisely,* $|Corr_n(f,g)| \leq C(\log n)^{b+1} \cdot n^{-b}$*; furthermore, if* $b > 1$*,* $(M, \mathcal{T}, \nu)$ *satisfies the central limit theorem.*

Remark 1.7. *The exposition of [36] (in contrast to [35]) restricts to expanding (and thus non-invertible) dynamical systems, hence literally it does not cover our case. Nonetheless, it is possible to extend the results of [36] to hyperbolic (invertible) systems modeled by Young towers with polynomial tails ([18], [28]).*

In order to apply Proposition 1.6, we prove the following Proposition, which is the second main ingredient in the proof of Theorems 1.1 and 1.2.

Proposition 1.8. *There exists a constant* $C > 0$ *(which depends on* m*) such that:*

$$\mu(R_n) \leq C \cdot \frac{1}{n^4}.$$

Summarizing, Propositions 1.5, 1.6 and 1.8 together imply Theorems 1.1 and 1.2.

As for Theorem 1.3, we recall the notion of a suspension flow, see eg. [29]. Given a map $(M, \mathcal{T}, \nu)$ (the base transformation) and a ν-integrable function $r : M \to \mathbb{R}^+$ (the roof function), the phase space of the associated suspension flow is $\tilde{M} = \{(x,s) \in M \times \mathbb{R}^+ : 0 \leq s \leq r(x)\}/\sim$, where $\sim$ is the equivalence relation $(x, r(x)) \sim (\mathcal{T}x, 0)$. The flow $\mathcal{S}^t$ acts on $\tilde{M}$ by $\mathcal{S}^t(x,s) = (x, s+t)$ computed modulo identifications, and has an invariant measure $d\tilde{\nu} = d\nu \times \frac{d\lambda}{\bar{r}}$, where λ is the restriction of the one dimensional Lebesgue measure and $\bar{r} = \int_M r(x) d\nu(x)$.

To prove Theorem 1.3, our main reference is the following result from the literature:

Proposition 1.9 ([27],[29])**.** *Consider a suspension flow* $(\tilde{M}, \mathcal{S}^t, \tilde{\nu})$ *with*

- *base transformation* $(M, \mathcal{T}, \nu)$ *that can be modeled by a Young tower with exponential tails;*

- *roof function* $r : M \to \mathbb{R}^+$ *that is uniformly piecewise Hölder continuous.*

Then $(\tilde{M}, \mathcal{S}^t, \tilde{\nu})$ *satisfies the central limit theorem (and some stronger limit theorems, see Remark 1.4).*

In our case, the flow $(\tilde{M}, S^t, \tilde{\mu})$ can be represented as a suspension flow in two different ways:

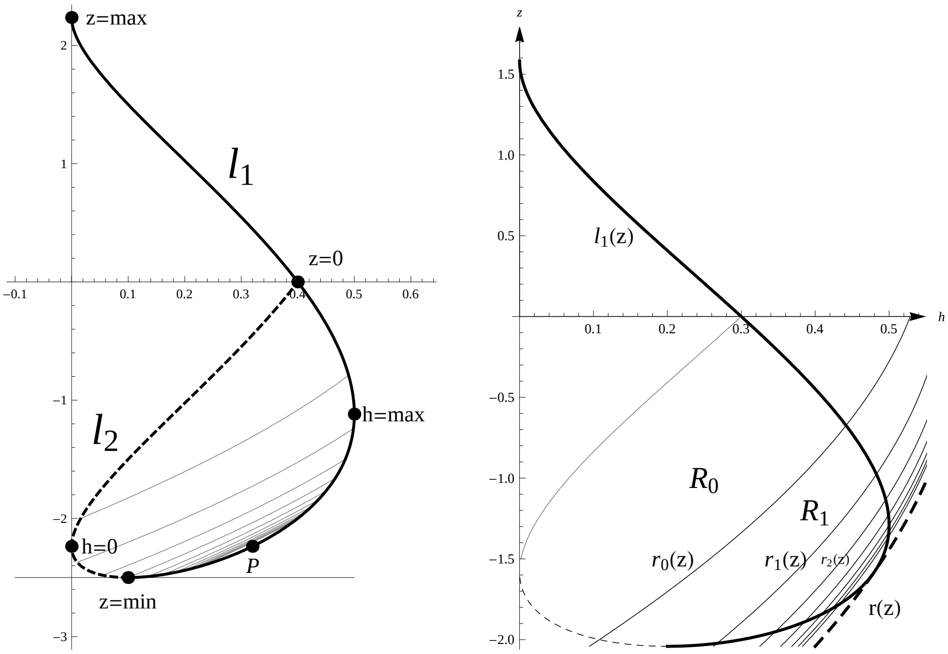
- with base transformation $(\mathcal{M}, \hat{T}, \hat{\mu})$ and roof function $\hat{\tau}$,

- with base transformation $(\mathcal{M}_1, T, \mu)$ and roof function τ.

In view of Proposition 1.9, to obtain Theorem 1.3, it is enough to prove the following Proposition.

Proposition 1.10. *Both of the roof functions* $\hat{\tau} : \mathcal{M} \to \mathbb{R}^+$ *and* $\tau : \mathcal{M}_1 \to \mathbb{R}^+$ *are uniformly piecewise Hölder continuous.*

Remark 1.11. *Actually, Proposition 1.9 applies in case the base transformation can be modeled by a Young tower with summable tails. Hence, to obtain Theorem 1.3, it would be enough to establish the piecewise Hölder continuity of* $\hat{\tau} : \mathcal{M} \to \mathbb{R}^+$*. However, we choose to include the uniform piecewise Hölder continuity of* $\tau : \mathcal{M}_1 \to \mathbb{R}^+$ *as this could be useful for further work (see section 5).*



(a) $m = 0.8$ (b) $m = 0.6$

Figure 5: l_1 connects the points $z = \min$ and $z = \max$, l_2 connects the points $z = \min$ and $z = 0$

The rest of the paper is organized as follows. In section 2 – after summarizing some system-specific notations – we prove Proposition 1.8. The proof of Proposition 1.5 is contained in sections 3 and 4. More precisely, in section 3, a detailed geometric analysis of the first return map $(\mathcal{M}_1, T, \mu)$ is presented, with special emphasis on the dynamics of unstable curves. As a consequence of this geometric analysis, we prove that $(\mathcal{M}_1, T, \mu)$ satisfies assumptions A0-A6 from the Appendix. A further consequence is that bounds on the Hölder regularity of the roof function are obtained, that is, we prove Proposition 1.10 (section 3.7).

Assumption A7, the so called growth lemma, requires special care, it is proved in section 4. This completes the proof of all our results. It is important to point out that the growth lemma (Assumption A7) is the only point where we need to put restrictions on the external parameter m. We close the paper in section 5 where the status of the restrictions on m are discussed, along with the description of some open questions.

2 Analysis of the first return sets

In this section we derive the asymptotic measure (area) of the sets R_n i.e. prove Proposition 1.8.

2.1 Notations concerning the geometry of the phase space

Let us collect first some system-specific notations for further reference. See Figure 5(a).

$$P := \left(2m(1-m), \frac{-1}{\sqrt{1-m}}\right) \in \overline{\mathcal{M}_1} \qquad \text{the accumulation point of the sets } R_n \tag{2.1}$$

$$\left(\frac{1-m}{2}, \frac{-1}{\sqrt{m(1-m)}}\right) \in \overline{\mathcal{M}_1} \qquad z = \min \tag{2.2}$$

$$\left(0, \frac{-1}{\sqrt{1-m}}\right) \in \overline{\mathcal{M}_1} \qquad h = 0 \tag{2.3}$$

$$\left(\frac{1}{2}, \frac{-1}{\sqrt{m}}\right) \in \overline{\mathcal{M}_1} \qquad h = \max \tag{2.4}$$

$$\left(0, \frac{1}{\sqrt{1-m}}\right) \in \overline{\mathcal{M}_1} \qquad z = \max \tag{2.5}$$

$$\alpha = 1 - 3m + 2m^2 \qquad \text{negative in the ergodic case} \tag{2.6}$$

$$F = 1 - \frac{h}{m} + z^2\alpha \qquad \text{for the brevity of the formulas} \tag{2.7}$$

All these formulas immediately follow once we know the equations for the curves $h = l_1(z)$ and $h = l_2(z)$ (right and left boundary curves of the phase space) and $h = r_n(z)$ (the curve that separates R_n from R_{n+1}). The relevant formulas (2.8)-(2.10) will be derived in section 2.2:

$$l_1(z) = \frac{1}{2}m\left(1 - 2(1-m)z\sqrt{1-(1-m)mz^2} + z^2\alpha\right) \tag{2.8}$$

$$l_2(z) = \frac{1}{2}m\left(1 + 2(1-m)z\sqrt{1-(1-m)mz^2} + z^2\alpha\right) \tag{2.9}$$

$$r_n(z) = m + mz^2\alpha - m\left(\frac{(1-m)(2n+3)z - \sqrt{m^2z^2 - m\left(4n^2+12n+z^2+8\right) + (2n+3)^2}}{\sqrt{2}\left(4m\left(n^2+3n+2\right) - (2n+3)^2\right)}\right)^2 \tag{2.10}$$

Convention 2.1. *Throughout the paper we consider m as a fixed parameter and use the subscript m only if we want to emphasize dependence on m. In our analysis, curves in $\mathcal{M}_1$ that arise as graphs of smooth functions $h = f(z)$ often appear. Slightly abusing terminology, we will refer to such a curve as "the curve $f(z)$"; also, the case when the function $f(z)$ is (strictly) decreasing/increasing will be referred to as a strictly decreasing/increasing curve.*

2.2 Bounding Functions

First let us derive formulas (2.8)-(2.10).[1] The curves that define the boundary of the whole $\mathcal{M}$ can be derived from the equation:

$$\frac{1}{2} - h = \frac{1}{2}m_2v_2^2 = \frac{1}{2}(1-m)\left(\sqrt{\frac{2h}{m}} + z\right)^2$$

solving which for h defines a function

$$h = l(z) = \frac{1}{2}m\left(1 \pm 2(1-m)z\sqrt{1-(1-m)mz^2} + z^2\alpha\right).$$

[1]For symbolic manipulations – simplifying and evaluating our formulas – we used the software Wolfram Mathematica throughout the arguments of the paper.

Since $z < 0$ for $(h,z) \in \mathcal{M}_1$, the "$-$" sign represents the greater quantity from the two options thus we should use "$-$" to get the curve that bounds $\mathcal{M}$ from the right (l_1). For l_2 choose the "$+$" sign. See Figure 5(a) or 5(b).

To express the curves separating the sets R_n and R_{n+1} we calculate

$$(h'(h,z), z'(h,z)) = {F_2}^n(F_1(h,z))$$

and test whether (h', z') is on the boundary curve between $\mathcal{M}_1$ and $\mathcal{M}_2$, that is the curve for which $h_t = 0$ (recall formula (1.1)). This separates the two possible scenarios – whether F_2^n or F_2^{n+1} should be applied to obtain T.

$$\frac{(1-m)mz'\left(2\sqrt{\frac{2h'}{m}} - z'\right) - 2h' + m}{2gm(1-m)} = 0 \iff (1-m)mz'\left(2\sqrt{\frac{2h'}{m}} - z'\right) - 2h' + m = 0.$$

After the substitutions and some simplifications the equation to solve takes the following form.

$$2\left(1 - \frac{h}{m} + z^2\alpha\right) - 1 +$$
$$(1-m)\left(2\sqrt{2}(n+1)\sqrt{1 - \frac{h}{m} + z^2\alpha} + z\right)\left(\sqrt{2}(2n+4)\sqrt{1 - \frac{h}{m} + z^2\alpha} + z\right) = 0$$

We use the F notation defined in (2.7) and solve the (quadratic) equation for $\sqrt{F}$.

$$2F - 1 + (1-m)\left(2\sqrt{2}(n+1)\sqrt{F} + z\right)\left(\sqrt{2}(2n+4)\sqrt{F} + z\right) = 0$$
$$\Downarrow$$
$$\sqrt{F} = \frac{(1-m)(2n+3)z \pm \sqrt{m^2z^2 - m\left(4n^2 + 12n + z^2 + 8\right) + (2n+3)^2}}{\sqrt{2}\left(4m\left(n^2 + 3n + 2\right) - (2n+3)^2\right)}$$

We choose the "$-$" sign, in order to have a positive solution. One can show that the other solution is negative. Now we can solve the following equation for h to eliminate F.

$$\overbrace{1 - \frac{h}{m} + z^2\alpha}^{F} = \left(\frac{(1-m)(2n+3)z - \sqrt{m^2z^2 - m\left(4n^2 + 12n + z^2 + 8\right) + (2n+3)^2}}{\sqrt{2}\left(4m\left(n^2 + 3n + 2\right) - (2n+3)^2\right)}\right)^2$$

This will give exactly (2.10).

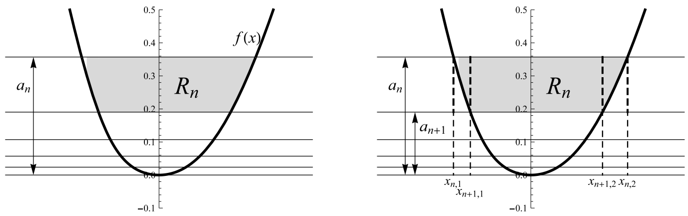
Theoretically knowing these formulas would allow us to calculate the exact value of $\mu(R_n)$. But the integrals seem impossible to calculate, therefore our estimation is based on the asymptotic behaviour of $r_n(z)$.

2.3 A Simplified Model

As the Figures 5(b) and 4(b) suggest, there is a limit where the functions r_n tend to.

$$r(z) := r_\infty(z) = \lim_{n\to\infty} r_n(z) = m\left(1 + \alpha z^2\right)$$

The figures also suggest that the sets R_n are shaped like parallel stripes and these stripes accumulate on $r(z)$. We can make a strongly simplified model of the stripes: take a sequence of parallel and horizontal



(a) (b)

Figure 6: The stripes R_n in the simplified model

lines crossed by the graph of a specific function $f(x)$. The height of the n^{th} line is a_n. $f(x)$ simulates $l_1(z)$, the a_n (constant functions) are analogous to $r_n(z)$, while the origin corresponds to the accumulation point P. First we will obtain an upper bound on the areas in this simple case, and consider the relevance of this model in section 2.5.

Notice that the tail bound for $\mu(R_n)$ is strongly determined by the order of the first non-vanishing derivative of f in 0. The higher the degree of the tangency, the less rapidly the areas of R_n decrease. To maintain the concept of Figure 6(a) we make some restrictions on f. Suppose that f is continuous, monotone decreasing for $x < 0$ and monotone increasing for $x > 0$ (consider only a small neighbourhood of 0). Also suppose that the following limit exists.

$$\lim_{x \to 0} \frac{f(x)}{x^k} = D \tag{2.11}$$

for some even number k and $D > 0$. This property implies that the first k derivatives of f exist in 0 and the value of the first non-vanishing derivative is $k! \cdot D$. Another consequence of these properties is that $f(0) = 0$ and f has a local minimum in 0, therefore the equation $f(x) = a_n$ has exactly two solutions (for a sufficiently large n): $x_{n,1} < 0 < x_{n,2}$.

With these notations:

$$(x_{n+1,2} - x_{n+1,1})(a_n - a_{n+1}) \le \mu(R_n) \le (x_{n,2} - x_{n,1})(a_n - a_{n+1})$$

Now let us estimate $x_{n,2} - x_{n,1}$ from above.

$$a_n = f(x_{n,1}) = (x_{n,1})^k \frac{f(x_{n,1})}{(x_{n,1})^k}$$

Now by (2.11) there exist constants (depending on f) c_1, c_2 and an $N_0 \in \mathbb{N}$ such that $0 < c_1 \le D \le c_2$ and $c_1 \le \frac{f(x_{n,1})}{(x_{n,1})^k} \le c_2$ holds for $n > N_0$ (notice that $x_{n,1} \to 0$ as $n \to \infty$). Consequently

$$(x_{n,1})^k c_1 \le a_n \le (x_{n,1})^k c_2 \quad \Longrightarrow \quad -x_{n,1} = |x_{n,1}| \le \sqrt[k]{\frac{a_n}{c_1}}.$$

In the same way we can estimate $x_{n,2}$. There exist other constants d_1, d_2 and an $M_0 \in \mathbb{N}$ such that $0 < d_1 \le D \le d_2$ and $d_1 \le \frac{f(x_{n,2})}{(x_{n,2})^k} \le d_2$ holds for $n > M_0$, so that:

$$(x_{n,2})^k d_1 \le a_n \le (x_{n,2})^k d_2 \quad \Longrightarrow \quad x_{n,2} = |x_{n,2}| \le \sqrt[k]{\frac{a_n}{d_1}}.$$

Consequently, for $n > \max\{N_0, M_0\}$, we get

$$\mu(R_n) \le (x_{n,2} - x_{n,1})(a_n - a_{n+1}) \le \left(\sqrt[k]{\frac{1}{d_1}} + \sqrt[k]{\frac{1}{c_1}} \right) \sqrt[k]{a_n}\, (a_n - a_{n+1}).$$

In particular, if we let $a_n = \frac{1}{n^\alpha}$ for some $\alpha > 0$, we obtain

$$\mu(R_n) = \mathcal{O}\left(\frac{1}{n^{\frac{\alpha}{k}}}\frac{1}{n^{\alpha+1}}\right) = \boxed{\mathcal{O}\left(\frac{1}{n^{\alpha+1+\frac{\alpha}{k}}}\right)}. \tag{2.12}$$

Notice that the bigger the k the slower the $\mu(R_n)$ tends to 0. Along similar lines, it is also possible to obtain a lower bound estimation of the same order of magnitude, which could be useful for further purposes (see eg. [3]).

2.4 Straightening the Stripes

In this section a map is constructed which distorts Figure 5(a) into Figure 6(a). To ensure the relevance of the estimate on $\mu(R_n)$ obtained in section 2.3, it is to be shown that the Jacobian of this map is uniformly bounded away from both 0 and ∞.

Let π denote the transformation, to be constructed below, that straightens the graphs of r_n. π is expected to have the following properties (for motivation see Figure 6(a) or 6(b)):

$$\pi : \mathcal{R}_m \mapsto \mathbb{R}^2, \quad \text{where} \quad \mathcal{R}_m = \left\{(h,z) \in \mathbb{R}^2 | z_{min} \le z \le 0 \text{ and } r_{\nu_0(m)}(z) \le h \le r(z)\right\},$$
$$\pi\begin{pmatrix} r_n(z) \\ z \end{pmatrix} = \begin{pmatrix} a_n \\ z \end{pmatrix}$$

The parameter ν_0 will be described later.

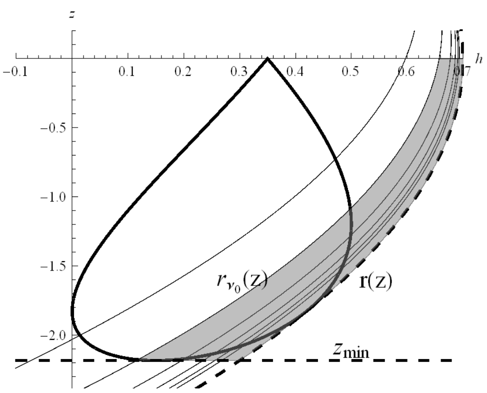


Figure 7: The region $\mathcal{R}_m$

In order to determine a_n let us calculate the following limit.

$$\lim_{n\to\infty} n^2 \left(r(z) - r_n(z)\right) = \frac{m\left(-(1-m)z + \sqrt{1-m}\right)^2}{8(1-m)^2}$$

We assumed that $1 > m > 0$ and $z < 0$ during the calculations. Suggested by this fact we make the choice $a_n = \frac{1}{n^2}$; this particular choice will guarantee that the transformation does not distort the area in a degenerate way, see section 2.5.

One can see that π transforms only the h coordinate and the action of π depends only on n. Pretend for a while that the parameter n is continuous, as the formula of r_n allows to substitute n with any positive number. This kind of generalisation implies that the curves $\{r_\nu(z)\}_{\nu\in[0,\infty)}$ cover the set $\bigcup_{n\in\mathbb{N}} R_n$. The domain of π is a subset of the region $\{(h,z) \in \mathbb{R}^2 | z_{\min} \le z \le 0, r_0(z) \le h < r(z)\}$ (see Figure 5(b)).

Even though the graphs $r_\nu(z)$ cover the above mentioned region, they do not foliate that. To see this note that, for example, $r_1(-6) = r_2(-6)$, if $m > \frac{5}{6}$. It can be checked that the curves $r_\nu(z)$ do foliate the region $\mathcal{R}_m$ for ν_0 large enough (the value of ν_0 depends on m).

Thus any point $x \in \mathcal{R}_m$ can be represented with two coordinates: z, the coordinate we have already used, and the above introduced ν which labels the graph on which x lies. To determine π we introduce a function which determines the ν coordinate.

$$\nu : \mathcal{R}_m \mapsto [0, \infty] \quad \text{so that} \quad r_{\nu(h,z)}(z) = h \quad \text{for all } (h, z) \in \mathcal{R}_m$$

With this notation π can be formalized easily:

$$\pi \begin{pmatrix} h \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{\nu(h,z)^2} \\ z \end{pmatrix} \tag{2.13}$$

Notice that π is undefined if $\nu(h, z) = 0 \Leftrightarrow r_0(z) = h$, but this is excluded by choosing ν_0 positive. Also one can define π if $h = r(z) \Leftrightarrow \nu(h, z) = \infty$ in a continuous way: $\pi(r(z), z) := (0, z)$.

$$\pi(h, z) = \begin{cases} (0, z) & \text{if } \nu(h, z) = \infty \\ \left(\frac{1}{\nu(h,z)^2}, z\right) & \text{otherwise} \end{cases}$$

We can express the function ν explicitly by solving the following equation (derived from (2.10)) for ν.

$$r_\nu(z) = h$$

$$\Updownarrow$$

$$\sqrt{1 - \frac{h}{m} + \alpha z^2} = \frac{(1-m)(2\nu+3)z - \sqrt{m^2 z^2 - m\left(4\nu^2 + 12\nu + z^2 + 8\right) + (2\nu+3)^2}}{\sqrt{2}\left(4m\left(\nu^2 + 3\nu + 2\right) - (2\nu+3)^2\right)}$$

This equation has a single real, positive solution, which is exactly the function $\nu(h, z)$:

$$\nu(h, z) = \frac{\sqrt{2}\left(\sqrt{(1-m)\,(1-2Fm)} - (1-m)z\right) - 6\sqrt{F}(1-m)}{4\sqrt{F}(1-m)}$$

where F is again a short notation for $1 - \frac{h}{m} + z^2\alpha$.

The transformation π distorts Figure 5(b) into Figure 8.

2.5 Overview

In section 2.4 we constructed a transformation which maps the sets R_n into the stripes of the simplified model. Now we have to prove the non-degeneracy of the Jacobian in order to make sure that the Lebesgue measures of the sets R_n are distorted only by constant. We also have to verify that the function $\pi(l_1(z))$ satisfies the conditions in section 2.3, and substitute the quantities k and a_n into formula (2.12).

We do not have to prove non-degeneracy in the whole set $\mathcal{R}_m$, only in a neighbourhood of P since we are interested in the asymptotic behaviour of $\mu\left(R_n\right)$.

Let $J_{h,z}$ denote the determinant of the Jacobian of π, derived from formula (2.13).

$$J_{h,z} = \text{Det}\begin{pmatrix} \left(\frac{1}{\nu(h,z)^2}\right)'_h & * \\ 0 & 1 \end{pmatrix} = \left(\frac{1}{\nu(h,z)^2}\right)'_h$$

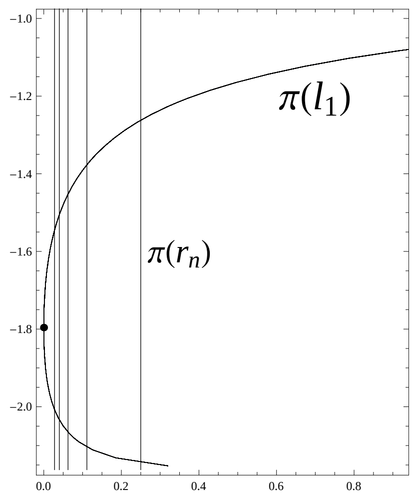


Figure 8: The image of the region $\mathcal{R}_m$ by the transformation π

The formula of $\left(\frac{1}{\nu(h,z)^2}\right)'_h$ is too long to copy here, nonetheless what is relevant for us is the following limit.

$$\lim_{h\to r(z)}\left(\frac{1}{\nu(h,z)^2}\right)'_h = \frac{8(1-m)^2}{m\left(-(1-m)z+\sqrt{1-m}\right)^2}$$

This limit is non-zero and the denominator is zero only if $z=\frac{1}{\sqrt{1-m}}$, but this point is outside of the phase space, since $(h,z)\in\mathcal{M}_1\Rightarrow z<0$. Therefore the non-degeneracy holds in a small neighbourhood of the point P.

To determine k we calculate the limit in formula (2.11). z_0 denotes the z coordinate of P.

$$\lim_{z\to z_0}\pi(l_1(z))(z-z_0)^{-2}=1-m$$

Hence $k=2$ and we can finally prove Proposition 1.8.

Proof. We calculated that the simplified model is relevant and the constants in formula (2.12) are $k=2,\alpha=2$. Consequently

$$\mu\left(R_n\right)\le C\frac{1}{n^{\alpha+1+\alpha/k}}=C\frac{1}{n^{2+1+2/2}}=C\frac{1}{n^4}.$$

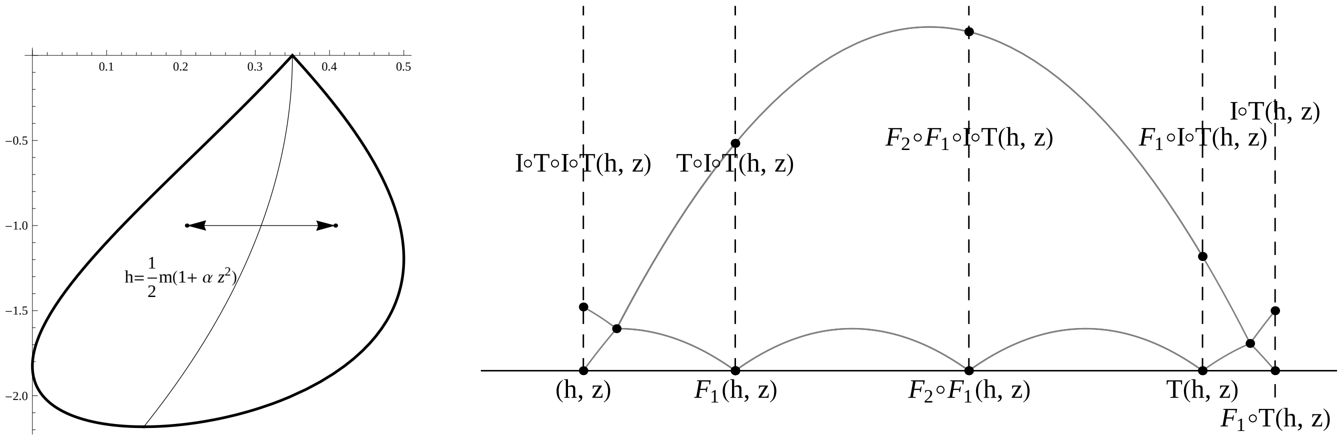
□

3 Regularity properties of the first return map

In the present section a detailed geometrical analysis is given of the first return map $(\mathcal{M}_1,T,\mu)$. Special emphasis is put on the dynamics of unstable curves.

The main aim of our analysis is to prove assumptions A0-A6 from the Appendix. Actually, $(\mathcal{M}_1,T,\mu)$ automatically satisfies some of these assumptions:

- Assumption A0: $\mathcal{M}_1$ is a smooth and compact Riemannian manifold (the closure of a bounded open set in $\mathbb{R}^2$ with piecewise smooth boundary);
- Assumption A1: the map T is discontinuous precisely at $\mathcal{S}=\cup_{n=0}^{\infty}(r_n(z),z)$, which is a countable collection of smooth and compact curves, hence it is closed and has zero Lebesgue measure.



(a) I as a reflection on the phase space

(b) The physical interpretation of I

Figure 9: Two interpretations of the involution

- Assumption A3: the invariant measure μ is the normalized Lebesgue measure on $\mathcal{M}_1$, the ergodicity and mixing of which is established in [22].

As for the further conditions, after some technical preparation, assumptions A2, A4, A5 and A6 are proved in sections 3.3, 3.4, 3.5 and 3.6, respectively.

3.1 Involution

Definition 3.1. *Given a (piecewise) smooth map $\mathcal{T} : M \to M$, by involution (or time reflection) we mean a smooth map $\mathcal{I} : M \to M$ for which $\mathcal{I}^{-1} = \mathcal{I}$ and $\mathcal{T}^{-1} = \mathcal{I} \circ \mathcal{T} \circ \mathcal{I}$; in other words, $\mathcal{I}$ smoothly conjugates the forward to the backward dynamics.*

In billiard dynamics, there is a useful and natural involution which simply reverses the velocity of the billiard particle (see eg. [12]). It turns out to be useful to implement the involution to our setting, however, this should be done with care as we work with the first return map $(\mathcal{M}_1, T, \mu)$.

Proposition 3.2. *For the first return dynamics the map $I(h, z) = (m(1 + \alpha z^2) - h, z)$ is an involution, i.e. $T^{-1} = I^{-1} \circ T \circ I$ and $I = I^{-1}$ on the set $\mathcal{M}_1$.*

Proof. Let the action of I be the following: let F_1 act on the point (h, z) and then reverse the velocity of the upper particle. Notice that the action of F_1 - which is needed to ensure $I(x) \in \mathcal{M}_1$ - makes this involution unique, compared to the standard involution in billiards. By expressing the standard height and velocity coordinates with h and z one can check that this map has indeed the above formula, and the involution property also holds (the formula of T is also known). On the phase space the action of I is a reflection through the curve $\frac{1}{2}m\,(1 + \alpha z^2)$, along the vertical direction, therefore the property $I^{-1} = I$ is automatic. Beyond the formalism, the physical meaning of I (let the balls collide, wait for the lower ball to hit the floor and reverse the velocity of the upper ball) is more picturesque. See Figure 9(b). □

Notation 3.3. *Let $F : \mathcal{M} \mapsto \mathcal{M}$ be an arbitrary diffeomorphism of a 2 dimensional manifold and consider a curve $\gamma : [0, 1] \mapsto \mathcal{M}$. Also let $\gamma(t) = x$, some point lying on γ.*

$$\mathcal{J}_\gamma F x := \frac{\|\mathrm{D}_x F \cdot \gamma'(t)\|}{\|\gamma'(t)\|}$$

Analogously if v is a vector in the tangent bundle $\mathcal{T}_x\mathcal{M}$, then

$$\mathcal{J}_v Fx := \frac{\|\mathrm{D}_x F \cdot v\|}{\|v\|}$$

In particular we will use this notation for the first return map T, for which, by hyperbolicity (to be discussed in section 3.3), there exist local unstable manifolds of positive length for (Lebesgue) almost every $x \in \mathcal{M}_1$. When using the notation without a subscript, we always mean $\gamma = W^u(x)$, the unstable manifold through the point x:

$$\mathcal{J}Tx := \mathcal{J}_{W^u}Tx$$

we will refer to $\mathcal{J}Tx$ as the local expansion factor of T at x.

Lemma 3.4. *There are uniform constants $0 < C_1 < C_2$ such that for any smooth curve γ we have $C_1 \le \mathcal{J}_\gamma Ix \le C_2$ for all $x \in \mathcal{M}_1$.*

Proof. Let us take a curve γ and let $x = (h, z) = \gamma(t)$ and $\gamma'(t) = (\cos\phi, \sin\phi)$. The expansion of the involution, along γ, can be calculated.

$$\mathcal{J}_\gamma Ix = |\mathrm{D}_x I \cdot \gamma'(t)| = \sqrt{(2mz\alpha\sin(\phi) - \cos(\phi))^2 + \sin^2(\phi)}$$

This function is clearly uniformly bounded away both from 0 and ∞ (for a fixed m) since it is continuous and positive on the unit tangent bundle of $\overline{\mathcal{M}_1}$, which is apparently a compact set. □

The existence of the involution has numerous useful consequences, in particular the singularity set for the inverse map T^{-1} can be identified:

Corollary 3.5. *The map T^{-1} is discontinuous precisely at the curves $I(r_n(z), z)$.*

3.2 The asymptotics of F

The quantity $F = 1 - \frac{h}{m} + \alpha z^2$ has occurred in many formulas and we will need some estimations about it. There is only one point on the closure of the phase space where $F = 0$, this is the accumulation point of the singularities, see (2.1).

$$(h, z) \to P \Leftrightarrow n \to \infty \Leftrightarrow F \to 0$$

Let us consider $F|_{R_n}$, and search for extremum for a fixed n. Throughout the investigation, we will use the notations from section 2.1 and in addition, we introduce the notations X_n, and BX_n for the right and left end of r_n respectively (the lower right and left "corners" of R_n), and we also consider IX_n and IBX_n, their images by the involution (cf. Figure 10(b)).

Let us check first the gradient of F.

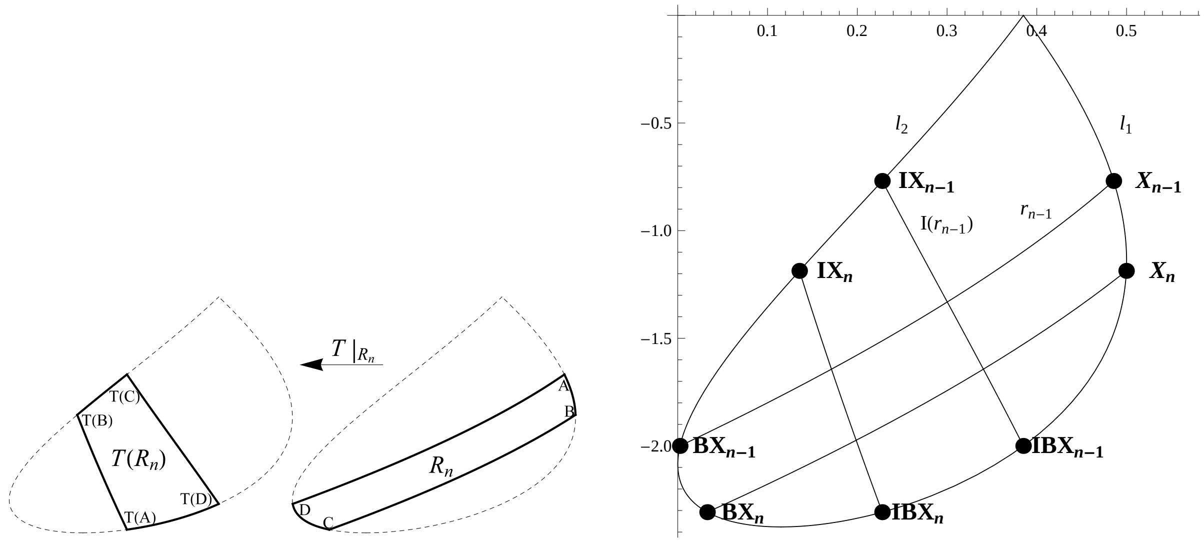
$$\partial_h F(h, z) = -\frac{1}{m} \neq 0$$
$$\partial_z F(h, z) = 2\alpha z.$$

It follows that there are no extrema in the interior of R_n.

Now let us consider the boundary curves r_n, l_1, l_2. We always describe these curves as graphs of functions $h = f(z)$ and we call such a curve (strictly) increasing/decreasing if the function $f(z)$ (strictly) increases/decreases in z.

Now notice that one can express F in the following way:

$$F(h, z) = \frac{1}{m}(I(h, z))_1. \tag{3.1}$$



(a) How the dynamics acts on the stripes

(b) The "corners" of the stripes R_n and IR_n

Figure 10:

That is essentially the first coordinate of the involution. This important remark allows us to investigate the monotonicity of F by geometric tools.

According to (3.1) and the fact that $I(l_1(z), z) = (l_2(z), z)$ we conclude that the function $F(l_2(z), z)$ strictly increases/decreases if and only if the curve $l_1(z)$ strictly increases/decreases; and reversely the monotonicity properties of $F(l_1(z), z)$ and $l_2(z)$ also coincide. Similarly the monotonicity of $F(r_n(z), z)$ is determined by the monotonicity of the curve $I(r_n)$. See Figure 10(b).

- It can be checked that $l_1(z)$ is increasing for $\frac{-1}{\sqrt{m(1-m)}} \leq z \leq \frac{-1}{\sqrt{m}}$, and decreasing for $\frac{-1}{\sqrt{m}} \leq z \leq 0$.
- $l_2(z)$ is decreasing for $\frac{-1}{\sqrt{m(1-m)}} \leq z \leq \frac{-1}{\sqrt{1-m}}$ and increasing for $\frac{-1}{\sqrt{1-m}} \leq z \leq 0$.
- The relevant segments of the curves $r_n(z)$ (i.e. which lie in $\mathcal{M}_1$, or equivalently, the segments $(BX_n)_2 \leq z \leq (X_n)_2$) are always strictly increasing.
- Let us denote the inverse singularities as $Ir_n(z) = I(r_n(z), z)_1$. The exact formula for them is

$$Ir_n(z) = \frac{m}{2} \frac{\left(-(1-m)(2n+3)z + \sqrt{m + (1-m)(2n+3)^2 - m(1-m)z^2}\right)^2}{(m + (1-m)(2n+3)^2)^2}$$

  The segments of these curves that lie in $\mathcal{M}_1$ (or equivalently the segments for $(IBX_n)_2 \leq z \leq (IX_n)_2$) are strictly decreasing as a function of z.

The above observations give a full description of the increases and decreases of F along ∂R_n for all values of n (see Fig. 11).

Now we know that the maximum and minimum of F are in the upper left and lower right corners of the sets R_n, respectively. Hence we need to express the intersections $r_n(z) = l_1(z)$ and $r_n(z) = l_2(z)$. Trying to do this in a formal way seems too complicated, but one can derive the solutions (for a fixed n) by understanding the physical meaning of these points. Let us name the corners according to Figure 10(b). X_n corresponds to the configuration when (i) both masses are on the floor and (ii) after the lower one hits the floor $n-1$ times, they reach the ground again at the same time (see Figure 12). Due to (ii)

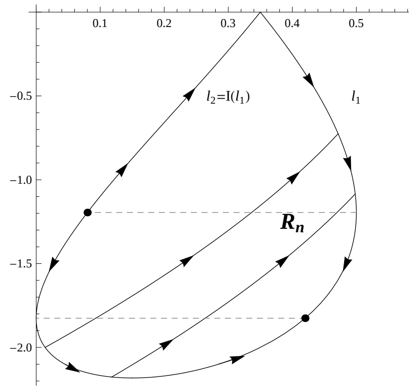


Figure 11: F decreases in the marked directions

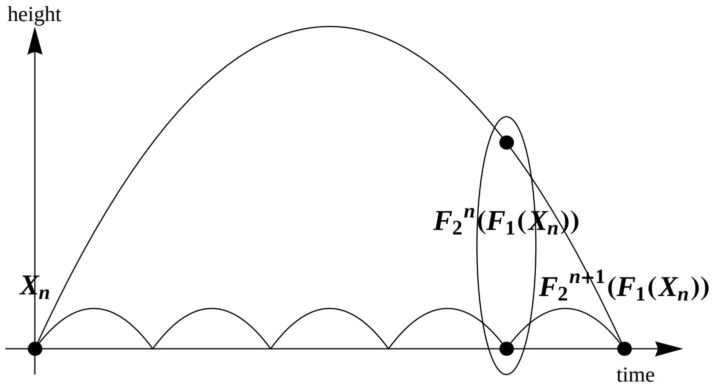


Figure 12: The physical interpretation of X_n

the point $F_2^n(F_1(X_n))$ is on the curve $h_t = 0$ (recall formula (1.1)). By the symmetry of the picture one can easily derive that, at the same time, $IX_n = F_2^n(F_1(X_n))$ which lies on the curve $z = -n\sqrt{\frac{2h}{m}}$. Hence solving the system of equations $h_t = 0$ and $z = -n\sqrt{\frac{2h}{m}}$ gives the coordinates of IX_n. Then simply the involution leads to X_n. By understanding how the dynamics act on the stripes (see Figure 10(a)) one can see that $T|_{R_n}(X_{n-1}) = IBX_n$ from which again the involution gives the coordinates of BX_n.

$$
\begin{aligned}
X_n &= \left(\frac{m(3+2n-2m(n+1))^2}{2(n+2)^2 - 2m(n+1)(n+3)}, -\frac{n+1}{\sqrt{(n+2)^2 - m(n+1)(n+3)}} \right) \\
BX_n &= \left(\frac{m(-2m(n+2)+2n+3)^2}{2(1-m)n(n+2)+2}, -\frac{n+2}{\sqrt{1-(m-1)n(n+2)}} \right) \\
IX_n &= \left(\frac{m}{2\left((n+2)^2 - m(n+1)(n+3)\right)}, -\frac{n+1}{\sqrt{(n+2)^2 - m(n+1)(n+3)}} \right) \\
IBX_n &= \left(\frac{m}{2(1-m)n(n+2)+2}, -\frac{n+2}{\sqrt{1-(m-1)n(n+2)}} \right)
\end{aligned}
\tag{3.2}
$$

Having these coordinates one can estimate F easily.

Lemma 3.6. $\exists\ 0 < c_1 < c_2 < \infty$ *such that for the quantity* $F(h,z) = 1 - \frac{h}{m} + \alpha z^2$ *the following holds:*

$$\frac{c_1}{n^2} \le F(h,z) \le \frac{c_2}{n^2} \qquad \forall (h,z) \in R_n$$

Proof. We have already showed that

$$F(X_n) \le F(h,z) \le F(BX_{n-1}) \qquad \forall (h,z) \in R_n$$

and we can calculate that

$$\lim_{n\to\infty} n^2 \cdot F(X_n) = \frac{1}{2-2m} = \lim_{n\to\infty} n^2 \cdot F(BX_{n-1}). \tag{3.3}$$

These limits ensure that the estimations hold. □

Lemma 3.7. $\exists\ 0 < c_3 < \infty$ *such that:*

$$|F(x) - F(y)| \le \frac{c_3}{n^3} \qquad \forall x, y \in R_n$$

Proof. If $x, y \in R_n$ then it is obvious that $|F(x) - F(y)| \le F(BX_{n-1}) - F(X_n)$. Let us calculate the limit

$$\lim_{n\to\infty} n^3 (F(BX_{n-1}) - F(X_n)) = \frac{2}{1-m}$$

The existence of this limit completes the proof. □

3.3 Hyperbolicity

In the present section we prove Assumption A2. The suitable unstable cone field $C^u_{x,1}$ will be obtained as a slight continuous modification of the constant cone field C^u_x, already introduced in section 1.1, while the stable cone field $C^s_{x,1}$ is obtained as the image of the unstable cone field by the involution.

Lemma 3.8. *There exists a slight continuous modification of the constant cone field* $C^u_x = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \cdot v_2 \le 0\}$ *that is a forward strictly invariant unstable cone field. There is a constant* $\Lambda = \Lambda(m) > 1$ *such that for all* $x \in \mathcal{M}_1$ *and* $\underline{v} \in C^u_x$*, the inequality* $\|D(T) \cdot \underline{v}\| \ge \Lambda \|\underline{v}\|$ *holds.*

Proof. We already know (recall section 1.1) that the constant cone field (that consists of the upper left and lower right quarter of the plane) is forward invariant, but it is not strictly invariant. The Jacobian of T at a point $(h,z) \in R_n$ maps the horizontal direction into a line with slope $\frac{-\sqrt{2}(n+1)}{m\sqrt{1-\frac{h}{m}+\alpha z^2}}$ and the vertical direction into a line with slope $\frac{-1}{2m\alpha z} - \frac{\sqrt{2}(n+1)}{m\sqrt{1-\frac{h}{m}+\alpha z^2}}$. Hence strict forward invariance of the cone field fails at $z = 0$ or $1 - \frac{h}{m} + \alpha z^2 = 0$ because either the horizontal or the vertical direction is mapped onto the vertical line. For each condition there exists only one point in $\bar{\mathcal{M}}_1$ satisfying its equation. They are $(\frac{m}{2}, 0)$ and $(2m(1-m), \frac{-1}{\sqrt{1-m}})$ respectively. At the point $(\frac{m}{2}, 0)$ the Jacobian of T is parabolic and it preserves the vertical direction, hence in a small neighbourhood of the point one has to modify the cone field in a continuous way by turning counterclockwise the vertical side a bit. At the other point the Jacobian is not even defined, but in the limit both the horizontal and the vertical directions are mapped onto the vertical line. Therefore to guarantee strict invariance in the neighbourhood of the limiting image of this point (i.e. in $(0, \frac{-1}{\sqrt{1-m}})$) one has to modify the cone field in a continuous way by turning clockwise the vertical side a bit. Unfortunately for $m = \frac{3}{4}$ the image of the point $(0, \frac{-1}{\sqrt{1-m}})$ is $(\frac{m}{2}, 0)$, so in this case the last modification should be done in such a way that the image of the modified cone at $(0, \frac{-1}{\sqrt{1-m}})$ is mapped by the dynamics strictly into the modified cone at $(\frac{m}{2}, 0)$. This can be done however, because the Jacobian of T at the point $(0, \frac{-1}{\sqrt{1-m}})$ is always hyperbolic.

Hence, the so defined cone field is now strictly forward invariant. General arguments (see [22]) ensure that the expansion is uniform in the following sense: there exist $c > 0$ and $\lambda > 1$ such that for any vector w in the cone we have $|DT_x^n v| \geq c\lambda^n |v|$. Nonetheless, in our case it can be verified that already the first iterate expands uniformly, that is, there exists some $\Lambda > 1$ such that $|DT_x v| \geq \Lambda |v|$. To see this we refer to Lemma 4.1 from section 4: any vector $v \in C_x^u$ is expanded by $D_x T$ at least as much as the vertical direction $(0,1)^T$. Now except for the point $(\frac{m}{2}, 0)$ the matrix of $D_x T$ is hyperbolic, hence it expands the vertical direction uniformly, while at $(\frac{m}{2}, 0)$ the tangent map is parabolic, hence it expands uniformly all directions except for the vertical. However, in the above procedure we have modified the cone field in a neighborhood of $(\frac{m}{2}, 0)$ in such a way that the vertical direction is excluded. □

Definition 3.9. *Let us define* $\{\mathcal{C}_x^s\}$ *as* $\mathcal{C}_{Ix}^s := I(\mathcal{C}_x^u)$ *for all* $x \in \mathcal{M}_1$ *(for brevity we denote the involution and its action on the tangent bundle by the same symbols* I*).*

Lemma 3.10. *The cone field* $\mathcal{C}_{Ix}^s$ *is backward invariant.*

Proof. We will show the backward invariance of the cone field $\mathcal{C}^s$, namely $T^{-1}(\mathcal{C}_{Ix}^s) \subseteq \mathcal{C}_{T^{-1}(Ix)}^s$ for every $x \in \mathcal{M}_1$. Note that I is onto, hence every $y \in \mathcal{M}_1$ is equal to Ix for some $x \in \mathcal{M}_1$.

$$\begin{aligned}
&\mathrm{D}_{Ix}\, T^{-1} \cdot \mathcal{C}_{Ix}^s = \\
&\mathrm{D}_{Ix}\, T^{-1} \cdot (\mathrm{D}_x\, I \cdot \mathcal{C}_x^u) = \mathrm{D}_{Tx}\, I \cdot \underbrace{\mathrm{D}_{I(Ix)}\, T}_{\mathrm{D}_x\, T} \cdot \underbrace{\mathrm{D}_{Ix}\, I \cdot \mathrm{D}_x\, I}_{\mathrm{Id}} \cdot \mathcal{C}_x^u = \\
&\mathrm{D}_{Tx}\, I \cdot \underbrace{\mathrm{D}_x\, T \cdot \mathcal{C}_x^u}_{\subseteq \mathcal{C}_{Tx}^u} \subseteq \mathrm{D}_{Tx}\, I \cdot \mathcal{C}_{Tx}^u = \\
&\mathcal{C}_{I(Tx)}^s = \mathcal{C}_{ITIIx}^s = \mathcal{C}_{T^{-1}(Ix)}^s.
\end{aligned}$$

□

Remark 3.11. *This argument is not specific to our system, it uses merely the forward invariance of* C_x^u *and the properties of the involution.*

Corollary 3.12. *Redefining the unstable cones as* $C_{x,1}^u := \overline{\left\{(v_1, v_2) \in \mathbb{R}^2 | \frac{v_2}{v_1} < \frac{-2}{\sqrt{m}}\right\}}$ *with the same continuous modifications as in Lemma 3.8 and set the stable cones as* $C_{Ix,1}^s := I(C_{x,1}^u)$*, the so constructed families of cones are strictly invariant and uniformly transversal.*

Proof. We already have a forward strictly invariant cone field, let us denote it as $\left\{C_{x,0}^u\right\}_{x \in \mathcal{M}_1}$. If the involution of these cones were transversal to the original ones then the statement would be true immediately. However the involution preserves (just reflects) the horizontal direction. It is a natural idea to turn the horizontal side of the unstable cones but keep their strict invariance to handle this issue. Note that this is how we defined the new cones: apart from the slight continuous modifications, $C_{x,1}^u$ is the domain between the line with slope $\frac{-2}{\sqrt{m}}$ and the y-axis. To prove strict invariance of this new cone field it is enough to show that the image of the two bounding lines of $C_{x,1}^u$ is in the interior of $C_{Tx,1}^u$ (except for the origin). This can be seen by investigating the one-step forward iterate of the original cone $C_{x,0}^u$. For a point $(h, z) \in R_n$ the image of the horizontal direction under the action of $D(T)$ has slope $\frac{-\sqrt{2}(n+1)}{m\sqrt{1-\frac{h}{m}+\alpha z^2}}$. The maximum of this over $\mathcal{M}_1$ is the same as the minimum of its absolute value. First we determine the maximum of the denominator. To achieve this recall (3.1), the connection between the quantity F and the involution. Since the right-most point of $\overline{\mathcal{M}_1}$ is $(\frac{1}{2}, \frac{-1}{\sqrt{m}})$, the maximum of the denominator is $m\sqrt{\frac{1}{2m}}$, and this is attained

at $I(\frac{1}{2}, \frac{-1}{\sqrt{m}}) = (\frac{(1-2m)^2}{2}, \frac{-1}{\sqrt{m}}) \in \partial\mathcal{M}_1$. This point always lies in $\overline{R_0}$ because the left endpoint of $r_0(z)$ is $(\frac{m}{2}(3-4m)^2, -2)$ and $-2 < \frac{-1}{\sqrt{m}}$ for $\frac{1}{2} \leq m \leq 1$. This means that in the numerator we have $n = 0$, which is the best possible lower bound and so the maximal slope of the image of the horizontal line is determined as $\frac{-2}{\sqrt{m}}$. From this it follows that the image of the side of $C^u_{x,1}$ with slope $\frac{-2}{\sqrt{m}}$ is mapped into the interior of $C^u_{Tx,1}$, because even the horizontal line is already mapped into $C^u_{Tx,1}$ (see Figure 13). The vertical edge of $C^u_{x,1}$ is the same as of $C^u_{x,0}$, so due to the strict invariance of the original cone field this side is always mapped into the interior of $C^u_{Tx,1}$.
Since the involution maps the line with slope $\frac{-2}{\sqrt{m}}$ into a line with strictly positive slope the statement is proved. □

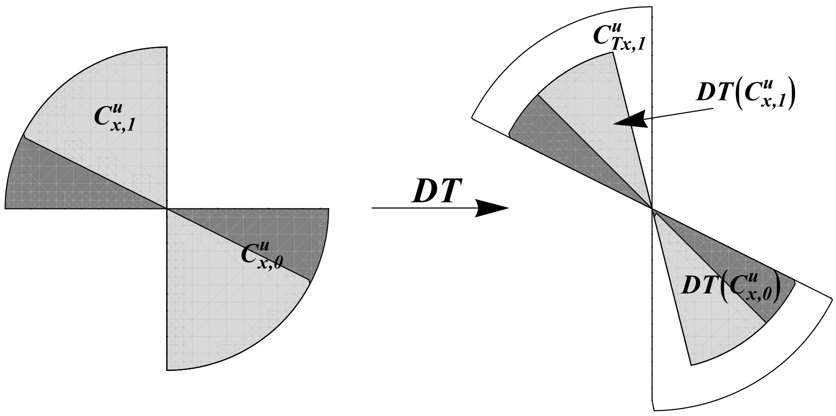


Figure 13: Strict forward invariance of the unstable cone field

We still need to check that the tangent lines of the singularities of the forward dynamics lie in the stable cone, and the tangent lines to the singularities of the backward dynamics lie in the unstable cone. Now T is singular precisely at the curves r_n which have positive slope, while T^{-1} is singular precisely at the curves Ir_n which have negative slope. This property extends to singularities of the higher iterates by the forward/backward invariance of the relevant cone fields.

3.4 Bounded curvature

In the present section we prove Assumption A4. Instead of unstable manifolds we consider first arbitrary unstable curves:

Definition 3.13. *A smooth curve $\gamma : [a, b] \to \mathcal{M}_1$ is an unstable curve (u-curve for short) if for all $t \in [a, b]$, $\gamma'(t)$, the tangent to γ at the point $\gamma(t)$, lies in the unstable cone $C^u_{x,1}$.*

If γ is an unstable curve, then for $n \geq 0$ the set $T^n\gamma$ consists of countably many smooth curves, which are all unstable by the invariance of the cone field $C^u_{x,1}$. These (maximal) smooth subcurves will be referred to as the components of $T^n\gamma$.

In particular, we would like to show that the dynamics act on unstable curves in a controllable way, even under infinitely many iterations. Let us investigate the effect of the dynamics on such a curve (and its curvature). Let γ be a parametrized unstable curve $\gamma(t) = (x(t), y(t))$, $\gamma'(t) \in \mathcal{C}^u_{(x(t),y(t))}$. The curvature of γ at a point (x, y) is

$$\kappa = \frac{x'y'' - x''y'}{(x'^2 + y'^2)^{\frac{3}{2}}} = \frac{\left\langle \gamma'^{\perp}, \gamma'' \right\rangle}{\|\gamma'\|^{\frac{3}{2}}}$$

where $(x,y)^\perp$ stands for a perpendicular vector $(-y,x)$. First we consider a general $\mathbb{R}^2 \mapsto \mathbb{R}^2$ map.

$$T(x,y) = (f_1(x,y), f_2(x,y))$$

With these notations one can calculate the curvature of $T(\gamma)$.

$$\frac{(-\nabla f_2 \cdot \gamma') \left(\nabla f_1 \cdot \gamma'' + \gamma'^\top \cdot H_{f_1} \cdot \gamma'\right) + (\nabla f_1 \cdot \gamma') \left(\nabla f_2 \cdot \gamma'' + \gamma'^\top \cdot H_{f_2} \cdot \gamma'\right)}{\|\mathrm{D}\,T \cdot \gamma'\|^{\frac{3}{2}}} \tag{3.4}$$

Here ∇ is the gradient vector and $\gamma'^\top \cdot H_{f_i} \cdot \gamma'$ is a quadratic form with the second derivative (Hesse) matrix.

Let us suppose, that γ is arc-length parametrized i.e. $\|\gamma'\| \equiv 1$ and $\kappa = \|\gamma''\|$. This means that we can use the following substitutions:

$$\begin{aligned} \gamma(t) &= (x,y) \\ \gamma(t)' &= (\cos\varphi, \sin\varphi) \\ \gamma(t)'' &= \kappa(-\sin\varphi, \cos\varphi). \end{aligned} \tag{3.5}$$

After substituting (3.5) into (3.4) one gets the curvature of $T(\gamma)$ at the point $T(x,y)$, which we denote by $G_T(x,y,\varphi,\kappa)$. Symbolically the various partial derivatives of T appear in G_T, however, this expression is *linear* in κ, hence can be represented as

$$G_T(x,y,\varphi,\kappa) = G_{1,T}(x,y,\varphi)\kappa + G_{2,T}(x,y,\varphi). \tag{3.6}$$

In order to investigate the regularity properties of unstable curves, below we derive estimates on these coefficient functions for our dynamics. The main goal is to find a $q < 1, K < \infty$ such that $|G_1| \le q, |G_2| \le K$ uniformly, which would be enough for our purposes (see Corollary 3.16). Note that we need to estimate the factors of Formula (3.6) for the first return map T, and not for the maps F_1 and F_2 separately (actually, there is no $a < 1$ such that $|G_{1,F_2}(x,y,\varphi)| \le a < 1$).

Lemma 3.14. *There exist two constants $a_1 < 1$ and $K < \infty$ such that*

$$|G_{1,T^2}| \le a_1 < 1 \text{ and } |G_{2,T^2}| \le 2K < \infty$$

where $T : \mathcal{M}_1 \mapsto \mathcal{M}_1$ is the first return map.

Proof. Let us start to establish the statement of the Lemma for the first return map T itself; the failure of this attempt will also make clear how to prove it for T^2. For $(h,z) \in R_n$ the first return map $T = F_2^n \circ F_1$ has the form

$$T(h,z)|_{R_n} = \begin{pmatrix} -h + m\,(1 + \alpha z^2) \\ -z - 2\sqrt{2}(n+1)\sqrt{F} \end{pmatrix} \tag{3.7}$$

where α and F are the system-specific expressions introduced in the former sections. Let us pretend as if n was a free variable in this expression; that is, its value could be chosen arbitrarily for any $(h,z) \in \mathcal{M}_1$. The exact formula of $G_{T|_{R_n}}$ is too lengthy to copy here, but we managed to reduce it by introducing the following quantity:

$$B := (n+1)\,(\cos(\varphi) - 2m\alpha z \sin(\varphi)) \tag{3.8}$$

Notice that the sign of B is negative for $\pi/2 < \varphi < \pi$ and $z < 0$ (these are the relevant values as we consider unstable curves on $\mathcal{M}_1$). Even on the closure, on the compact set $\mathcal{K} = \overline{\mathcal{M}_1} \times [\pi/2, \pi]$, $B = 0$ can only occur if $z = 0$ and $\varphi = \pi/2$. With this notation the curvature function has the following form:

$$G_{T|_{R_n}}(h,z,\varphi,\kappa) = \frac{\mathcal{A}\kappa + \mathcal{B}}{\mathcal{D}}$$

where

$$
\begin{aligned}
\mathcal{A} &= \left(Fm^2\right)^{\frac{3}{2}} \\
\mathcal{B} &= \frac{1}{2}m\Big(4m\alpha\sin^2(\varphi)\left(F^{\frac{3}{2}}m^2\sin(\varphi)-\sqrt{2}B((F-1)m+h)\right)+ \\
&\quad 4\sqrt{2}Bmz\alpha\sin(\varphi)\cos(\varphi)-\sqrt{2}B\cos^2(\varphi)\Big) \\
\mathcal{D} &= \Big(2B^2-\sqrt{8F}Bm\sin(\varphi)+Fm^2(1+4mz\alpha\sin(\varphi)(mz\alpha\sin(\varphi)-\cos(\varphi)))\Big)^{\frac{3}{2}} \\
&= (Fm^2+2B^2+\text{non-negative terms})^{\frac{3}{2}}.
\end{aligned}
$$

Now we have to show that there exist uniform constants k, ϵ and q such that $\mathcal{B} \le k < \infty$ and $\mathcal{D} \ge \epsilon > 0$ ($\frac{\mathcal{B}}{\mathcal{D}} \le \frac{k}{\epsilon} < \infty$) and $\frac{\mathcal{A}}{\mathcal{D}} \le q < 1$.

Since $\mathcal{B}$ has a continuous extension to the compact domain $\mathcal{K}$, the existence of $k < \infty$ for which $\mathcal{B} \le k$ is immediate.

Now let us investigate $\mathcal{D}$, which is apparently non-negative. To show *uniform* positivity, note that there is a Fm^2 and a $2B^2$ term in the expression. These two can not be 0 at the same time, since $B = 0 \Rightarrow z = 0$ and $z = 0 \Rightarrow F > 0$, hence $\mathcal{D} > 0$ on a compact set, which gives us the desired inequality (B and F apparently depend continuously on h, z and φ).

The last remaining expression to be checked has the following form: $\frac{\mathcal{A}}{\mathcal{D}} = \left(\frac{Fm^2}{Fm^2+\text{non-negative terms}}\right)^{\frac{3}{2}}$, as the term $mz\alpha\sin(\varphi) - \cos(\varphi)$ is non-negative. Hence one can immediately see that $\frac{\mathcal{A}}{\mathcal{D}}$ can not be larger than 1, however if $z = 0$ and $\varphi = \pi/2$, then it *is equal to* 1. Therefore we can not apply our former reasoning in a direct way and we consider T^2.

The only bad point on $\overline{\mathcal{M}_1}$ where the estimate for T fails is $u = (h_u, z_u) = (m/2, 0)$; which is *not a fixed point* of T as $T|_{R_0}(m/2, 0) = (m/2, -2) \ne (m/2, 0)$.

To prove the statement of the Lemma for T^2, choose a small, open neighbourhood N of $u = (m/2, 0)$ such that $u \in N \subseteq \overline{R_0} \subset \overline{\mathcal{M}_1}$ and $T(N) \cap N = \emptyset$. Let

$$
a_1 := \max\{G_{1,T}(h,z,\varphi) | (h,z) \in \mathcal{M}_1 \setminus N, (\cos(\varphi), \sin(\varphi)) \in C^u_{(h,z)}\} < 1
$$

Hence for any point $(h, z) \in N$:

$$
|G_{T^2}(h,z,\varphi,\kappa)| \le a_1 \cdot \underbrace{(|\kappa| + K)}_{|G_T| \le} + K
$$

After the first action of T we can not guarantee any decrease in the curvature, but since $T(h, z)$ is not in N, we can guarantee a_1 as a linear factor. On the other hand, for any $(h, z) \in \mathcal{M}_1 \setminus N$ the first action of T has a factor a_1 but we can not guarantee a small linear factor for the image of such a point. Summarizing

$$
|G_{T^2}(h,z,\varphi,\kappa)| \le (a_1|\kappa| + K) + K.
$$

□

Definition 3.15. *Given a positive number $C < \infty$, an unstable curve γ is* <u>*C-regular*</u> *if its curvature κ is less than C in all its points. Let, furthermore, $K' := \frac{2K}{1-a_1}$, where K and a_1 are the constants of Lemma 3.14. K'-regular unstable curves will be simply called regular unstable curves.*

Lemma 3.14 has the following corollaries.

Corollary 3.16. *(a) If γ is a regular unstable curve, then every component of $T^2(\gamma)$ is a regular unstable curve.*
(b) For almost every $x \in \mathcal{M}_1$, the unstable manifold $W^u(x)$ of x is a regular unstable curve.

Proof. Let $\gamma^{(i)}$ denote the components of $T^2\gamma$, where γ is a regular unstable curve. For any $y \in \gamma^{(i)}$ there exist $x \in \gamma$ with $T^2x = y$. Let κ and κ' denote the curvatures of γ at x and of $\gamma^{(i)}$ at y, respectively. Then we have

$$\kappa' \le a_1\kappa + 2K \le a_1\left(\frac{2K}{1-a_1}\right) + 2K = \frac{1}{1-a_1}2K = K'$$

which completes the proof of statement (a). See [5], Lemma 4 for a proof of statement (b). □

3.5 Distortion bounds

In this section we prove Assumption A5. In view of Corollary 3.16, it is enough to prove the following Lemma.

Lemma 3.17. *There exist some $C > 0$, such that, for any $l \ge 1$ we have the following property. Let γ be a regular unstable curve on which T^l is smooth. Recall Notation 3.3, and our convention form the Appendix: for $x, y \in \gamma$, $\operatorname{dist}(x, y)$ denotes the distance of the two points within the curve γ. We have*

$$\sum_{i=0}^{l-1}\left|\frac{\mathcal{J}_\gamma T^i x}{\mathcal{J}_\gamma T^i y} - 1\right| \le C \operatorname{dist}(T^l x, T^l y)^{\frac{1}{3}}.$$

First let us prove the Lemma for $l = 1$; extension to higher iterates is essentially automatic by uniform hyperbolicity. Throughout, γ denotes a regular unstable curve. T is smooth on γ if and only if $\gamma \subset R_n$ for some $n \ge 0$. We handle the easier case of $\gamma \subset R_0$ separately; $\gamma \subset R_n, n \ge 1$ ($n \to \infty$ in particular) requires more care and a different reasoning as the derivative of T is unbounded. Throughout, $C > 0$ denotes some uniform constant; the value of C is irrelevant and may change form line to line.

Lemma 3.18. *If γ is a regular unstable curve in R_0 with tangent vector $\underline{t}(x)$ at the point $x \in R_0$ then the following distortion bound holds: $\left|\frac{\mathcal{J}_{\underline{t}}Tx}{\mathcal{J}_{\underline{t}}Ty} - 1\right| \le C \operatorname{dist}(x, y)$.*

Proof. Introduce some notations. Let $a(h, z) = 2m\alpha z$ and $b(h, z) = \frac{\sqrt{2}}{m\sqrt{1-\frac{h}{m}+\alpha z^2}}$ so that $D_x(T) = \begin{pmatrix} -1 & a(h,z) \\ b(h,z) & -1-a(h,z)b(h,z) \end{pmatrix}$ for $x = (h, z) \in R_0$. It is easy to check that both $a(h, z)$ and $b(h, z)$ are Lipschitz on R_0. Since $\underline{t}(x)$ is a unit unstable vector we can write it as $\begin{pmatrix} \cos\xi(x) \\ \sin\xi(x) \end{pmatrix}$ for some function $\frac{\pi}{2} \le \xi(x) \le \pi$. Regular unstable curves have bounded curvature, therefore $\xi(x)$ depends on x Lipschitz continuously. Note also that $|\mathcal{J}_{\underline{t}}Tx| \ge 1$, so that we have

$$\left|\frac{\mathcal{J}_{\underline{t}}Tx}{\mathcal{J}_{\underline{t}}Ty} - 1\right| \le |\mathcal{J}_{\underline{t}}Tx - \mathcal{J}_{\underline{t}}Ty|$$

with

$$\mathcal{J}_{\underline{t}}Tx = \sqrt{1 + b^2\cos^2(\xi) - (2a + 2b + 2ab^2)\sin(\xi)\cos(\xi) + (a^2 + 2ab + a^2b^2)\sin^2(\xi)}$$

where, according to the above observations, the quantities a, b and ξ depend on x Lipschitz continuously, which completes the proof of the Lemma. □

To proceed note that the derivative $\mathrm{D}_x T$, when considered as a matrix acting on $\mathbb{R}^2$, has two real eigenvalues out of which one (the expanding eigenvalue) has absolute value greater than 1. Throughout the section this expanding eigenvalue will be denoted by $\lambda(x)$. To prove Lemma 3.17 with $l = 1$ for $\gamma \subset R_n$ with $n \to \infty$, our strategy is to first estimate regularity of $\lambda(x)$, and then find the relation between $\lambda(x)$ and the expansion factor along γ.

First we derive an explicit formula for $\lambda(h, z)$.

$$F_2^n(F_1(h,z)) = \left(-h + m\left(1+\alpha z^2\right), -z - 2(n+1)\sqrt{2F}\right);$$

$$Tr = Trace\left(\mathrm{D}_{h,z}\, F_2^n\left(F_1(h,z)\right)\right) = -2 - \frac{\sqrt{2}(n+1)z\alpha}{\sqrt{F}};$$

$$\begin{aligned}\lambda(h,z) &= \left|\frac{1}{2}\left(Tr - \sqrt{Tr^2-4}\right)\right| = \\ &= 1 + \frac{\sqrt{2}\left(\sqrt{(n+1)z\alpha\left(\sqrt{2F}+(n+1)z\alpha\right)} + (n+1)z\alpha\right)}{\sqrt{F}} = \\ &= (n+1)^2\left(\frac{1}{(n+1)^2} + \frac{\sqrt{2}\left(\sqrt{z\alpha\left(\frac{\sqrt{2F}}{n+1}+z\alpha\right)}+z\alpha\right)}{\sqrt{F}(n+1)}\right)\end{aligned} \tag{3.9}$$

We took the negative sign in the formula (3.9) because $\lambda < 0$ and the expanding eigenvalue has the larger magnitude.

Lemma 3.19. *For any fixed n let $\gamma \subset R_n$ be a regular unstable curve, and $x, y \in \gamma$. Then there exists a constant $C > 0$ (independent of n) such that the following distortion bound holds:* $\left|\frac{\lambda(x)}{\lambda(y)} - 1\right| \le C \cdot \mathrm{dist}(x,y)^{\frac{1}{3}}$.

Proof. We have to concentrate on the limit $n \to \infty$, since the derivative blows up only in P. The Lemma will follow form the following three estimates (note that $x, y \to P$ as $n \to \infty$). There exist $c_1, c_2, c_3, c_4 > 0$ such that, for any regular unstable curve γ:

$$\|\mathrm{Grad}\,\lambda(x)\| \le c_3 \cdot n^4 \qquad \forall\, x \in R_n, \tag{3.10}$$

$$c_1 \cdot n^2 \le \lambda(x) \le c_2 \cdot n^2 \qquad \forall\, x \in R_n, \tag{3.11}$$

$$\mathrm{dist}(x,y) \le \frac{c_4}{n^3} \qquad \forall\, x, y \in \gamma \subset R_n. \tag{3.12}$$

Let us check (3.11) for $(h, z) \in R_n$.

$$\frac{\lambda(h,z)}{(n+1)^2} = \frac{1}{(n+1)^2} + \frac{\sqrt{2}\left(\sqrt{z\alpha\left(\frac{\sqrt{2F}}{n+1}+z\alpha\right)}+z\alpha\right)}{\sqrt{F}(n+1)}$$

Using Lemma 3.6. we can estimate this expression from both above and below.

$$\frac{1}{(n+1)^2} + \frac{\sqrt{2}\left(\sqrt{z\alpha\left(\frac{c_1\cdot 1/n}{n+1}+z\alpha\right)}+z\alpha\right)}{c_2\frac{1}{n}(n+1)} \le \frac{\lambda(h,z)}{(n+1)^2} \le$$

$$\le \frac{1}{(n+1)^2} + \frac{\sqrt{2}\left(\sqrt{z\alpha\left(\frac{c_2\cdot 1/n}{n+1}+z\alpha\right)}+z\alpha\right)}{c_1\frac{1}{n}(n+1)}$$

Now take $n \to \infty \Leftrightarrow (h,z) \to P \Rightarrow z \to \frac{-1}{\sqrt{1-m}}$.

$$C_1 \le \liminf_{n\to\infty} \frac{\lambda(h,z)}{(n+1)^2} \le \limsup_{n\to\infty} \frac{\lambda(h,z)}{(n+1)^2} \le C_2 \qquad \text{for some } 0 < C_1 < C_2 < \infty$$
$$\Updownarrow$$
$$c_1 \cdot n^2 \le \lambda(h,z) \le c_2 \cdot n^2 \qquad \text{for some } 0 < c_1 < c_2 < \infty,$$

which is exactly (3.11).

Now let us consider (3.10). Let $x = (h,z) \in R_n$ again.

$$\frac{\operatorname{Grad}\lambda(x)}{\lambda(x)} \overset{\text{by direct differentiation}}{=}$$
$$\frac{(n+1)\alpha}{Fm\sqrt{(n+1)z\alpha\left(\sqrt{2F}+(n+1)z\alpha\right)}} \begin{pmatrix} \frac{z}{2} \\ m-h \end{pmatrix},$$

hence the magnitude of $\frac{\operatorname{Grad}\lambda}{\lambda}$ is determined by the following term.

$$\frac{(n+1)\alpha}{Fm\sqrt{(n+1)z\alpha\left(\sqrt{2F}+(n+1)z\alpha\right)}} \overset{\text{Lemma 3.6.}}{\le}$$
$$\frac{(n+1)\alpha}{c_1 \cdot 1/n^2 \cdot m\sqrt{(n+1)z\alpha\left(\sqrt{2c_1}\cdot 1/n+(n+1)z\alpha\right)}} =$$
$$\frac{n^2\alpha}{c_1 \cdot m\sqrt{z\alpha\left(\frac{\sqrt{2c_1}}{n(n+1)}+z\alpha\right)}}.$$

As $n \to \infty$ the leading term is n^2.

$$\frac{\|\operatorname{Grad}\lambda(x)\|}{\lambda(x)} \le c \cdot n^2$$
$$\Updownarrow \ (3.11)$$
$$\|\operatorname{Grad}\lambda(x)\| \le c \cdot n^4$$

Which is exactly (3.10).

Now we have to prove (3.12), which is $\operatorname{dist}(x,y) \le \frac{c}{n^3}$ for $x, y \in \gamma$, where γ is a regular unstable curve in R_n. Along the lines of section 2.4 we calculate the following limit directly:

$$\lim_{n\to\infty} n^3\left(r_{n+1}(z) - r_n(z)\right) = \frac{m\left(1-\sqrt{1-mz}\right)^2}{4(1-m)},$$

which means that, for any fixed z, the length of the horizontal slices of the sets R_n decay as $1/n^3$. However one can construct a sequence of points $x_n, y_n \in R_n$, such that $\operatorname{dist}(x_n,y_n) \ge \frac{c}{n}$, because $\frac{1}{n}$ is the asymptotic width of R_n along the stable direction. Now we use the condition that γ – on which x and y are located – is an unstable curve. Since the graphs r_n are uniformly transversal to the unstable cone, the estimation $\frac{c}{n^3}$ is only distorted by a constant.

Finally we can combine (3.10)-(3.12).

$$\begin{array}{cr} \overset{(3.10)}{\|\text{Grad}\,\lambda(x)\| \le c\cdot n^4} & \forall\, x\in R_n \\ \Downarrow & \\ |\lambda_n(x)-\lambda_n(y)| \le c\cdot n^4\,\text{dist}(x,y) & \forall\, x,y\in R_n \end{array}$$

Hence for $x, y \in \gamma$, where $\gamma \subset R_n$ is a regular unstable curve:

$$\begin{aligned} |\lambda(x)-\lambda(y)| &\le c\cdot n^4\,\text{dist}(x,y) & \\ &\le c\cdot n^4\cdot \frac{c_1}{n^3} = & \text{using (3.11) and (3.12)} \\ &= c_2\cdot n & \\ &\le C\cdot \lambda(x)^{\frac{1}{2}} & \text{using (3.11)} \end{aligned}$$

Hence for any $0<\alpha<1$:

$$\begin{aligned} |\lambda(x)-\lambda(y)| &= |\lambda(x)-\lambda(y)|^{\alpha}\,|\lambda(x)-\lambda(y)|^{1-\alpha} \\ &\le \Bigg(c_1\cdot \underbrace{n^4}_{c\cdot\lambda(y)^2}\,\text{dist}(x,y)\Bigg)^{\alpha}\left(c_2\cdot\lambda(y)^{\frac{1}{2}}\right)^{1-\alpha} \\ &= c\cdot \lambda(y)^{2\alpha+(1-\alpha)\frac{1}{2}}\,\text{dist}(x,y)^{\alpha} \end{aligned}$$

In particular for $\alpha=\frac{1}{3}$, $2\alpha+(1-\alpha)\frac{1}{2}=1$. Then dividing by $\lambda(y)$ gives

$$\left|\frac{\lambda(x)}{\lambda(y)}-1\right| \le c\cdot\text{dist}(x,y)^{\frac{1}{3}}.$$

□

Corollary 3.20. *Similar distortion bound holds for higher iterates of the map acting on unstable curves, i.e. for every $l\in\mathbb{N}$ and every x, y lying on the same regular unstable curve γ on which T^l is nonsingular:* $\left|\prod_{i=0}^{l-1}\frac{\lambda(T^ix)}{\lambda(T^iy)}-1\right| \le C\cdot\text{dist}(T^lx,T^ly)^{\frac{1}{3}}$.

Proof. Since T^l is nonsingular on γ for every $i=0\dots l$, $T^i\gamma$ is a regular unstable curve in some R_{n_i}. Because of the uniform hyperbolicity of T we know that $\text{dist}\,(T^ix,T^iy)\le\text{dist}\,\left(T^lx,T^ly\right)$ for any $i<l$. Moreover $\text{dist}\,(T^ix,T^iy)^{\frac{1}{3}} \le \frac{\text{dist}\left(T^lx,T^ly\right)^{\frac{1}{3}}}{\sqrt[3]{\Lambda}^{l-i}}$, as the minimal expansion of T along unstable curves is $\Lambda>1$. Here dist means the distance along γ, not the euclidean distance (nevertheless those two norms are comparable). We also use that

$$\left|\frac{\lambda(x)}{\lambda(y)}-1\right| \le c_1\,\text{dist}(x,y)^{\frac{1}{3}} \iff |\log\left(\lambda(x)\right)-\log\left(\lambda(y)\right)| \le c_2\,\text{dist}(x,y)^{\frac{1}{3}}$$

as by (3.11) we know that $C_1\le\frac{\lambda(x)}{\lambda(y)}\le C_2$, if $x,y\in R_{n_i}$, independently of n_i.

Using the above properties, we plant the result into higher iterates of T.

$$\left|\log \prod_{i=0}^{l-1} \frac{\lambda(T^i x)}{\lambda(T^i y)}\right| \leq \sum_{i=0}^{l-1} \left|\log\left(\lambda(T^i x)\right) - \log\left(\lambda(T^i y)\right)\right| \leq$$
$$\sum_{i=0}^{l-1} \operatorname{dist}\left(T^i x, T^i y\right)^{\frac{1}{3}} \leq \operatorname{dist}\left(T^l x, T^l y\right)^{\frac{1}{3}} \sum_{i=0}^{l-1} \sqrt[3]{\Lambda}^{-(l-i)} \leq c \cdot \operatorname{dist}\left(T^l x, T^l y\right)^{\frac{1}{3}}$$

Note that $\Lambda > 1$, therefore the geometrical series is sumable. □

Now we turn on to the verification of the distortion bound for regular unstable curves $\gamma \subset R_n$ with $n \geq 1$.

Lemma 3.21. *Denote by $\underline{e}_u(x)$ and $\underline{e}_s(x)$ the normalized unstable and stable eigenvectors of $D_x(T)$ respectively. Let γ be a regular unstable curve with unit tangent vector $\underline{t}(x)$ at the point $x \in \mathcal{M}_1 \setminus R_0$. Let the decomposition of $\underline{t}(x)$ in the eigendirections be $\underline{t}(x) = \alpha(x)\underline{e}_u(x) + \beta(x)\underline{e}_s(x)$ and denote the angle between $\underline{e}_u(x)$ and $\underline{e}_s(x)$ by $\varphi(x)$. Then there is a uniform constant $C > 0$ such that for every $x \in \mathcal{M}_1 \setminus R_0$ the following inequality holds: $\mathcal{J}_{\underline{t}} Tx \geq C|\alpha(x)\lambda(x)|$.*

Proof. The coefficients in the decomposition can be expressed using scalar products, for example $|\alpha(x)| = \left|\frac{\langle \underline{t}(x), \underline{e}_s^{\perp}(x)\rangle}{\langle \underline{e}_u(x), \underline{e}_s^{\perp}(x)\rangle}\right|$ and a similar formula gives $|\beta(x)|$. Away from the point $(\frac{m}{2}, 0)$ (which lies always in R_0) the slope of $\underline{e}_s(x)$ is positive and it is uniformly separated from 0 and from ∞ and the slope of $\underline{e}_u(x)$ is negative and it is uniformly separated from $-\infty$ and 0. These facts provide uniform upper bounds, say $C_{\alpha,1}$ and C_β on $|\alpha(x)|$ and $|\beta(x)|$ respectively and also a lower bound on $|\alpha(x)|$, say $C_{\alpha,2}$. Fixing any finite positive integer N_0 the statement of the lemma easily holds on the set $\cup_{n=1}^{N_0} R_n$ due to the fact that the dynamics expand vectors in unstable cones. Indeed

$$\mathcal{J}_{\underline{t}} Tx \geq 1 = \frac{1}{C_{\alpha,1} K_\lambda(N_0)} C_{\alpha,1} K_\lambda(N_0) \geq \frac{1}{C_{\alpha,1} K_\lambda(N_0)} |\alpha(x)\lambda(x)|$$

where $K_\lambda(N_0)$ is the maximum of $|\lambda(x)|$ on the set $\cup_{n=1}^{N_0} R_n$. It remains to examine the asymptotic behaviour as $n \to \infty$. From (3.11) we know that for $x \in R_n$, $|\lambda(x)| \to \infty$ as $n \to \infty$. Using this and the given bounds on $|\alpha(x)|$ and $|\beta(x)|$ it is clear that

$$\frac{\mathcal{J}_{\underline{t}} Tx}{|\alpha(x)\lambda(x)|} = \sqrt{1 + \frac{2\beta(x)\cos(\varphi(x))}{|\alpha(x)\lambda^2(x)|} + \frac{\beta^2(x)}{\alpha^2(x)\lambda^4(x)}}$$

converges to 1 for $x \in R_n$ as $n \to \infty$, which completes the proof. □

Lemma 3.22. *Within every stripe R_n, $n \geq 1$ the eigenvectors of $D(T)$ vary in a Hölder-continuous way, i.e. for every x and $y \in R_n$ we have $|\underline{e}_u(x) - \underline{e}_u(y)| \leq C \operatorname{dist}(x, y)^{\frac{1}{3}}$ and $|\underline{e}_s(x) - \underline{e}_s(y)| \leq C \operatorname{dist}(x, y)^{\frac{1}{3}}$.*

Proof. Here we prove only that the first coordinate of $\underline{e}_u(x)$ varies in a Hölder-continuous way, but similar calculations show that this is also the case for the second coordinate and for both coordinates of the stable eigenvectors. We introduce some notation. Let $a(x) = D_x(T)_{1,2}$, i.e. $a(x) = a(h, z) = 2m(1 - 3m + 2m^2)z$ for all $x = (h, z) \in \mathcal{M}_1$. The only point where $z = 0$ is $(\frac{m}{2}, 0)$, which lies always in R_0. Hence on the set $\cup_{n=1}^{\infty} R_n$ there is a uniform constant a_0 such that $a(x) \geq a_0 > 0$ and we know that here $D_x(T)$ is uniformly hyperbolic, i.e. there exists a uniform constant $\lambda_0 > 1$ such that for the unstable eigenvalue of the Jacobian $\lambda(x) \leq -\lambda_0 < -1$ holds. An elementary calculation shows that the first coordinate of the normalized unstable eigenvector of $D_x(T)$ is $e_1^u(x) = \frac{a(x)}{\sqrt{1+a^2(x)+2\lambda(x)+\lambda^2(x)}}$. Using the given bounds one can conclude that

$$|e_1^u(x) - e_1^u(y)| \leq C_1|a(x) - a(y)| + C_2\frac{|\lambda(x) - \lambda(y)|}{|1 + \lambda(x)|}$$

for some positive constants C_1, C_2. As a consequence of the upper bound on $\lambda(x)$ the inequality $\left|\frac{1-\lambda_0}{\lambda_0}\lambda(x)\right| \leq |1 + \lambda(x)|$ holds. Hence

$$|e_1^u(x) - e_1^u(y)| \leq C_1|a(x) - a(y)| + C_2\left|\frac{\lambda_0}{1 - \lambda_0}\right| \cdot \left|\frac{\lambda(y)}{\lambda(x)} - 1\right|$$

Applying Lemma 3.19 and the trivial fact that $a(x)$ is Hölder (even Lipschitz) we have $|e_1^u(x) - e_1^u(y)| \leq C \cdot \mathrm{dist}(x, y)^{\frac{1}{3}}$. ☐

Lemma 3.23. *On the stripes R_n, $n \geq 1$ the coefficients defined in Lemma 3.21 $\alpha(x)$, $\beta(x)$ and the cosine of the angle $\varphi(x)$ are all Hölder-continuous with exponent $\frac{1}{3}$, i.e. for x and $y \in R_n$ we have $|\alpha(x)-\alpha(y)| \leq C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$, $|\beta(x) - \beta(y)| \leq C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$ and $|\cos(\varphi(x)) - \cos(\varphi(y))| \leq C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$.*

Proof. Here we only discuss the case of $\alpha(x)$, for the other quantities similar computations should be done. As we have already mentioned in the proof of Lemma 3.21, for the unit eigenvectors the scalar product $\left\langle \underline{e}_u(x), \underline{e}_s^{\perp}(x)\right\rangle$ is uniformly separated from 0 on the set $\cup_{n=1}^{\infty} R_n$. So for a positive constant C_0 we have $\left\langle \underline{e}_u(x), \underline{e}_s^{\perp}(x)\right\rangle > C_0$. To prove the present statement one should write the difference $|\alpha(x) - \alpha(y)|$ as a difference of fraction of scalar products (recall the formula for $|\alpha(x)|$ from the proof of Lemma 3.21). Then by using the given uniform lower bound, the just proved Hölder-continuity of the eigenvectors of $D(T)$ and that tangent vectors of regular unstable curves vary also in a Hölder-continuous way (or even Lipschitz due to the proved bounded curvature property) after some triangular inequalities one can conclude that $|\alpha(x) - \alpha(y)| \leq C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$. ☐

Lemma 3.24. *If γ is a regular unstable curve in some stripe R_n with tangent vector $\underline{t}(x)$ at the point $x \in \mathcal{M}_1$ then the following distortion bound holds: $\left|\frac{\mathcal{J}_{\underline{t}}Tx}{\mathcal{J}_{\underline{t}}Ty} - 1\right| \leq C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$.*

Proof. In Lemma 3.18 we have already proved the statement on R_0, hence now we only need to verify it on the set $\cup_{n=1}^{\infty} R_n$. To do this we will use the results of the previous four lemmas. For brevity we introduce the notation $f(x) := \mathcal{J}_{\underline{t}}Tx$. Note that in Lemma 3.21 we proved that $f(x) \geq C_0|\alpha(x)\lambda(x)|$ and from the proof of Lemma 3.21 we also have $C_{\alpha,1} \geq |\alpha(x)| \geq C_{\alpha,2}$ and $C_\beta > 0$ as an upper bound on $|\beta(x)|$. We have

$$\left|\frac{\mathcal{J}_{\underline{t}}Tx}{\mathcal{J}_{\underline{t}}Ty} - 1\right| = \left|\frac{f(x)}{f(y)} - 1\right| = \left|\frac{f^2(x) - f^2(y)}{f(x)f(y) + f^2(y)}\right| \leq$$
$$\leq \left|\frac{\alpha^2(x)\lambda^2(x) - \alpha^2(y)\lambda^2(y)}{f(x)f(y) + f^2(y)}\right| + \left|\frac{2\alpha(x)\beta(x)\cos(\varphi(x)) - 2\alpha(y)\beta(y)\cos(\varphi(y))}{f(x)f(y) + f^2(y)}\right| + \left|\frac{\beta^2(x)\frac{1}{\lambda^2(x)} - \beta^2(y)\frac{1}{\lambda^2(y)}}{f(x)f(y) + f^2(y)}\right|. \tag{3.13}$$

Now we perform the calculation for the three terms separately.

1. For the first term one should estimate the denominator by using Lemma 3.21. Then the triangular inequality and the bounds on $|\alpha(x)|$ will lead to the estimate

$$\left|\frac{\alpha^2(x)\lambda^2(x) - \alpha^2(y)\lambda^2(y)}{f(x)f(y) + f^2(y)}\right| \leq \frac{C_{\alpha,1}}{C_0^2 C_{\alpha,2}}\left|\frac{\lambda(x)}{\lambda(y)} - 1\right| + \frac{1}{C_0^2 C_{\alpha,2}}|\alpha(x) - \alpha(y)|$$

According to Lemma 3.21 and Lemma 3.23 this is less than $C\,\mathrm{dist}(x, y)^{\frac{1}{3}}$.

2. For the denominator of the second term use again Lemma 3.21 and the lower bound on $|\alpha(x)|$ to conclude that this term is less than $\frac{2}{2C_0^2 C_{\alpha,2}^2}|\alpha(x)\beta(x)\cos(\varphi(x)) - \alpha(y)\beta(y)\cos(\varphi(y))|$. After this Lemma 3.23 shows that the expression can be estimated by $C\,\mathrm{dist}(x,y)^{\frac{1}{3}}$.

3. For the third term apply the method of the previous two cases in the denominator. Then use a triangular inequality and the upper bound on $|\beta(x)|$ (and also the fact that $|\lambda(x)| > 1$) to get the estimate

$$\left|\frac{\beta^2(x)\frac{1}{\lambda^2(x)} - \beta^2(y)\frac{1}{\lambda^2(y)}}{f(x)f(y) + f^2(y)}\right| \leq \frac{1}{C_{\alpha,2}}\left(C_\beta\left|\frac{\lambda(x)}{\lambda(y)} - 1\right| + |\beta(x) - \beta(y)|\right)\frac{C_\beta}{C_{\alpha,2}}$$

   Now Lemma 3.19 and Lemma 3.23 shows that this is less than $C\,\mathrm{dist}(x,y)^{\frac{1}{3}}$.

This completes the proof. □

Proof of Lemma 3.17. Lemma 3.24 can be iterated exactly as the estimate for $\lambda(x)$ in Corollary 3.20. □

3.6 Absolute continuity

In the present section we prove Assumption A6.

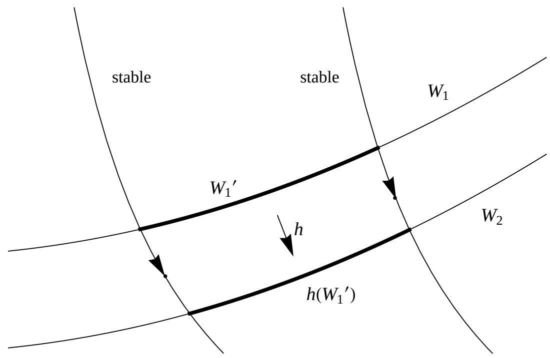


Figure 14: Sliding along stable manifolds

It is known (see [12], Theorem 5.39) that the Jacobian of the holonomy map h has the following form:

$$\mathcal{J}h(x) = \lim_{n\to\infty} \frac{\mathcal{J}_{W_1}T^n x}{\mathcal{J}_{W_2}T^n h(x)}.$$

Let γ denote the stable manifold which connects x and $h(x)$. As T^n is area preserving, and the angle between the stable and unstable directions is uniformly bounded away from 0, we have $c_1 \leq \mathcal{J}_{W_1}T^n x \cdot \mathcal{J}_\gamma T^n x \leq c_2$. Hence it is enough to prove $C_1 \leq \frac{\mathcal{J}_\gamma T^n h(x)}{\mathcal{J}_\gamma T^n x} \leq C_2$ for some constants and for all large n. At this point we can use the involution $(T = I \circ T^{-1} \circ I)$.

$$\frac{\mathcal{J}_\gamma(I \circ T^{-n} \circ I)(h(x))}{\mathcal{J}_\gamma(I \circ T^{-n} \circ I)(x)} = \frac{(\mathcal{J}_{T^{-n}I\gamma}I)\,(T^{-n}Ih(x))}{(\mathcal{J}_{T^{-n}I\gamma}I)\,(T^{-n}Ix)} \cdot \frac{(\mathcal{J}_{I\gamma}T^{-n})\,(Ih(x))}{(\mathcal{J}_{I\gamma}T^{-n})\,(Ix)} \cdot \frac{(\mathcal{J}_\gamma I)\,(h(x))}{(\mathcal{J}_\gamma I)\,(x)}.$$

Since we obtained uniform bounds on the expansion of I in Lemma 3.4, it is enough to consider only the middle factor. In other words the desired inequality reduces to

$$C_1 \leq \frac{(\mathcal{J}_{I\gamma}T^{-n})\,(Ih(x))}{(\mathcal{J}_{I\gamma}T^{-n})\,(Ix)} \leq C_2.$$

By Lemma 3.10 the involution maps γ into an unstable manifold $\gamma' = I\gamma$, where γ' connects the points Ix and $Ih(x)$. T^{-n} is smooth on γ', since unstable manifolds remain connected in backward time. Therefore we may use the rule for differentiating inverse functions.

$$\frac{(\mathcal{J}_{\gamma'}T^{-n})\,(Ih(x))}{(\mathcal{J}_{\gamma'}T^{-n})\,(Ix)} = \frac{(\mathcal{J}_{T^{-n}\gamma'}T^{n})\,(T^{-n}Ix)}{(\mathcal{J}_{T^{-n}\gamma'}T^{n})\,(T^{-n}Ih(x))}$$

$T^{-n}\gamma'$ is an unstable manifold, on which T^n is smooth, while $T^{-n}Ix$ and $T^{-n}Ih(x)$ are two points on $T^{-n}\gamma'$. Therefore the distortion bound (Lemma 3.24) implies:

$$\left|\log(\mathcal{J}_{T^{-n}\gamma'}T^{n})\left(T^{-n}Ix\right) - \log(\mathcal{J}_{T^{-n}\gamma'}T^{n})\left(T^{-n}Ih(x)\right)\right| \leq$$
$$c\cdot \mathrm{dist}(T^nT^{-n}Ix, T^nT^{-n}Ih(x))^{\frac{1}{3}} = c\cdot \mathrm{dist}(Ix, Ih(x))^{\frac{1}{3}}$$
$$\Downarrow$$
$$C_1 \leq \frac{(\mathcal{J}_{T^{-n}\gamma'}T^{n})\,(T^{-n}Ix)}{(\mathcal{J}_{T^{-n}\gamma'}T^{n})\,(T^{-n}Ih(x))} \leq C_2$$

which completes the verification of Assumption A6.

3.7 Regularity of the roof function

Proof of Proposition 1.10. First let us consider the roof function $\hat{\tau} : \mathcal{M} \to \mathbb{R}^+$. It is worth introducing $\tau_i = \hat{\tau}|_{\mathcal{M}_i}$, $i = 1, 2$, so that

$$\hat{\tau}(h,z) = \begin{cases} \tau_1(h,z) & \text{if } (h,z) \in \mathcal{M}_1, \\ \tau_2(h,z) & \text{if } (h,z) \in \mathcal{M}_2. \end{cases}$$

Let us calculate τ_2 first (when the balls do not collide). The lower ball is on the floor and has initial velocity $v_1 = \sqrt{\frac{2h}{m}}$. To fall back to the floor $\frac{2v_1}{g}$ time is required, so that

$$\tau_2(h,z) := \frac{\sqrt{\frac{8h}{m}}}{g}.$$

By the Gallilei law and the rules of elastic collision one can also calculate τ_1, that is, the flight time for a point in $\mathcal{M}_1$:

$$\tau_1(h,z) = \frac{\sqrt{2}\left(\sqrt{F} + \sqrt{\frac{h}{m}}\right) + 2(m-1)z}{g}.$$

Piecewise Hölder continuity of $\hat{\tau}$ is immediate.

Now let us consider the induced roof function $\tau : \mathcal{M}_1 \to \mathbb{R}^+$. For $(h,z) \in R_n$ we have

$$\tau(h,z) = \tau_1(h,z) + n\cdot\tau_2(F_1(h,z)) = \frac{\sqrt{2}\left(\sqrt{F}(2n+1) + \sqrt{\frac{h}{m}}\right) + 2(m-1)z}{g}.$$

Notice that for a fixed $n \in \mathbb{N}$ this function is differentiable on R_n, but the derivative has no finite bounds at two points: $h = 0$ and $(h,z) = P$. At the point $h = 0$ this function is Hölder-continuous with exponent $\frac{1}{2}$, because of the $\sqrt{h}$ term. The more problematic point is the accumulation point P. To obtain a uniform Hölder exponent we need to consider the asymptotic behaviour of $n \cdot \sqrt{F}$.

Lemma 3.25. *The roof function τ is uniformly piecewise Hölder continuous with the exponent $\frac{1}{3}$.*

Proof. For any $x, y \in R_n$, and any $\alpha > 0$ we have

$$\left|n\sqrt{F(x)} - n\sqrt{F(y)}\right| = n\underbrace{\frac{|F(x)-F(y)|}{\sqrt{F(x)}+\sqrt{F(y)}}}_{\text{Lemma 3.6.}} \le C_1 n^2 |F(x)-F(y)| =$$

$$C_1 n^2 \underbrace{|F(x)-F(y)|}_{\text{Lemma 3.7.}}{}^{1-\alpha} \cdot |F(x)-F(y)|^{\alpha} \le C_2 n^2 n^{-3(1-\alpha)} |F_1 - F_2|^{\alpha}.$$

To eliminate the dependence on n let $\alpha = \frac{1}{3}$. Also note that the quantity $F(x) = F(h, z)$ is a polynomial in (h, z), therefore $|F(x) - F(y)| \le C|x - y|$ for some $C > 0$. This completes the proof of the Lemma, and hence the proof of Proposition 1.10. □

□

4 Growth lemma

4.1 The first iterate of T

In this section we consider the growth lemma, that is, Assumption A7. It is to be checked that

$$\liminf_{\delta \to 0} \sup_{W:|W|<\delta} \sum_i \Lambda_i^{-1} < 1, \tag{4.1}$$

where W is an arbitrary unstable manifold with length at most δ, which is cut by singularities into pieces W_i indexed by i. Here Λ_i is the minimal expansion rate of T on the i-th piece, i.e. $\Lambda_i = \inf_{x \in W_i} \frac{\|DT(\underline{t}(x))\|}{\|\underline{t}(x)\|}$, where $\underline{t}(x)$ is a tangent vector of W_i at the point x.

The first, immediate problem is that even though unstable manifolds are locally unique we do not know exactly their directions, hence we can not calculate Λ_i exactly. But we can use that tangent vectors of unstable manifolds always lie in unstable cones.

Lemma 4.1. *For any $x \in \mathcal{M}_1$, the minimum expansion factor on the cone C_x^u is attained at the vertical direction $(0, 1)^T$.*
Consequently, is W is an unstable manifold with unit tangent vector $\underline{t}(x)$ at the point x, then we have $\left\|DT \cdot (0, 1)^T\right\| \le \|DT(\underline{t}(x))\|$.

Proof. The statement of the Lemma follows from the following two claims.

1. For any $x \in \mathcal{M}_1$, the minimum expansion factor of $D_x T$ on C_x^u is always attained at one of the boundaries, that is, either on the vertical vector $(0, 1)^T$, or on the horizontal vector $(1, 0)^T$. To see this, let us parametrize unit vectors $\underline{v} \in \mathbb{R}^2$ with the angle $\varphi \in [0, \pi]$ they make with the horizontal direction ($\varphi \in [\frac{\pi}{2}, \pi]$ corresponds to $\underline{v} \in C_x^u$). It can be checked by direct differentiation that for any 2 by 2 real matrix A the expansion rate, as a function of φ, is either constant (when A is a constant multiple of an orthogonal matrix) or it has exactly one local maximum and one local minimum. Now in our case the matrix $D_x T$ is hyperbolic, and the global minimum of the expansion rate is attained in the complement of C_x^u, which implies the claim.

2. For any x, $D_x T$ expands the horizontal direction at least as much as the vertical direction, that is $d(x) = d(h, z) = \left\|DT \cdot (0, 1)^T\right\|^2 - \left\|DT \cdot (1, 0)^T\right\|^2 \le 0$. To see this, we calculate

$$d(h, z) = 4m^2\alpha^2 z^2 F + 4\sqrt{2}(n+1)\alpha z\sqrt{F} + \frac{8m^2\alpha^2 z^2 - 2}{m^2}(n+1)^2 \tag{4.2}$$

for any $(h,z) \in R_n$. We claim that (4.2), when viewed as a quadratic polynomial of the independent variable $(n+1)$, is decreasing for $n \geq 0$. Indeed, by $\alpha < 0$ and $\frac{-1}{\sqrt{m(1-m)}} \leq z \leq 0$:

$$\alpha z \leq \frac{-(1-m)(1-2m)}{\sqrt{m(1-m)}} = \frac{\sqrt{1-m}(2m-1)}{\sqrt{m}} \leq \sqrt{2}\sqrt{1-m}(2m-1) \leq \frac{2}{3\sqrt{3}} \tag{4.3}$$

so that the leading coefficient of the quadratic is negative. The location of its maximum is $\frac{4\sqrt{2}\alpha z\sqrt{F}m^2}{4-16m^2\alpha^2 z^2}$ which, by $F \leq \frac{1}{2m}$ and Formula (4.3), is less than 1, which completes the proof of the claim. Hence it is enough to show that for $n=0$ (4.2) is negative. Substituting $n=0$ and estimating F in (4.2) by its global maximum $\frac{1}{2m}$ the result is $\frac{-2}{m^2}+\frac{4\alpha z}{\sqrt{m}}+(8+2m)\alpha^2 z^2$, which can be treated as a quadratic in "αz". This quadratic is convex, hence it is negative if αz is in between its two roots $\rho_1 = -\frac{\sqrt{m}+\sqrt{2}\sqrt{2+m}}{4m+m^2}$ and $\rho_2 = \frac{-\sqrt{m}+\sqrt{2}\sqrt{2+m}}{4m+m^2}$. Since $\alpha z \geq 0$ it is always larger than $\rho_1 < 0$. On the other hand by Formula (4.3) $\alpha z \leq \frac{\sqrt{1-m}(2m-1)}{\sqrt{m}}$, which can be shown to be less than ρ_2 for $m \in [\frac{1}{2}, 1]$, which completes the proof.

□

The singularities of T are the curves $r_n(z)$ for $n = 0, 1, \ldots$, therefore we may distinguish three cases as far as the number of terms in (4.1) is concerned.

Classification 4.2.

1. *If the unstable manifold W does not cross a singularity curve then the sum consists of only one term which is the inverse of the minimal expansion of $D(T)$ along W. Since T is uniformly hyperbolic this is always less than* 1.

2. *If W runs into the accumulation point of the singularities then the sum consists of infinitely many terms. But we already know the asymptotic behaviour of the expansion rates from (3.11) and Lemma 3.21. For every $x \in R_n$ we have $\Lambda(x) \geq C \cdot n^2$, hence the sum can be estimated as $\sum_{n=N_0}^{\infty} \frac{1}{Cn^2}$ where $N_0 = \min\{n \in \mathbb{N} | W \cap R_n \neq \emptyset\}$. Clearly by decreasing the length of W (but keeping it attached to the accumulation point), N_0 can be made arbitrary large and since $\frac{1}{Cn^2}$ is summable, the tail of the sum is less than* 1 *for a sufficiently large N_0. So* (4.1) *is verified in this case, because one can take δ (the length of W) arbitrarily small.*

3. *In any other case a sufficiently short unstable manifold W can intersect only one singularity curve, say $r_n(z,m)$. Let us denote their intersection by Q. This point cuts W into two subcurves $W_1 \subset R_n$ and $W_2 \subset R_{n+1}$. Let the infimum of the expansion rate of $D(T)$ on W_i be Λ_i. Now to show that* (4.1) *holds qualitative information are not enough. One has to do exact calculations and estimates on Λ_i's. This will be done in the rest of this subsection.*

To give lower bounds on the Λ_i's, according to Lemma 4.1 it is enough to investigate the expansion rate of $D(T)$ on the vertical direction. As a preparation we describe the shape of the sets R_n. Recall the notations from section 2.1, in particular Convention 2.1.

For $n \geq 1$, R_n is topologically a rectangle with boundaries $r_{n-1}(z,m)$, $r_n(z,m)$ (we will call them horizontal sides) and either $l_1(z,m)$ and $l_2(z,m)$ from "right" and "left" or just $l_1(z,m)$ from both sides (let us call them vertical). Exceptionally, R_0 is topologically a triangle with bounding functions $l_1(z,m)$, $l_2(z,m)$ and $r_0(z,m)$. To get an impression about the shape of these sets, we recall the observations concerning the monotonicity of these bounding curves from section 3.2.

Now we calculate the exact minimum of the expansion along the vertical direction on each set R_n.

Lemma 4.3. *The minimum of $\left\|DT|_{R_n}\cdot(0,1)^T\right\|^2$ on the set $\overline{R_n}$ is attained at $X(m,n-1)$. The value of it is an increasing function of n.*

Proof. First we consider the gradient of the square of the expansion on the vertical direction.

- $\frac{\partial\left\|DT\cdot(0,1)^T\right\|^2}{\partial h}=2\left\langle DT\cdot(0,1)^T,DT'_h\cdot(0,1)^T\right\rangle$, where DT'_h is the 2 by 2 matrix given as the elementwise derivative of DT with respect to h, i.e. $DT'_h=\begin{pmatrix}0 & 0\\ \frac{\sqrt{2}(n+1)}{2m^2F^{3/2}} & -\frac{\sqrt{2}(n+1)\alpha z}{mF^{3/2}}\end{pmatrix}$. Since both α and z are negative $DT'_h\cdot(0,1)^T=(0,A)$ for some $A<0$. Using our knowledge on the invariant cone field and the fact that both eigenvalues of DT are negative it can be seen that $DT\cdot(0,1)^T$ lies in the lower right quarter of the plane and it is not horizontal. So the angle between the vectors in the scalar product is an acute angle and therefore $\frac{\partial\left\|DT\cdot(0,1)^T\right\|^2}{\partial h}>0$.

- $\frac{\partial\left\|DT\cdot(0,1)^T\right\|^2}{\partial z}=2\left\langle DT\cdot(0,1)^T,DT'_z\cdot(0,1)^T\right\rangle$, where DT'_z is the 2 by 2 matrix given as the elementwise derivative of DT with respect to z, i.e. $DT'_z=\begin{pmatrix}0 & 2m\alpha\\ -\frac{\sqrt{2}(n+1)\alpha z}{mF^{3/2}} & \frac{2\sqrt{2}(n+1)\alpha(h-m)}{mF^{3/2}}\end{pmatrix}$. We know that $\alpha<0$, $0\le h\le\frac{1}{2}$ and $\frac{1}{2}<m<1$ and from these it follows that $DT'_z\cdot(0,1)^T$ lies in the upper left quarter of the plane. For the same reason as before $DT\cdot(0,1)^T$ is a vector in the lower right quarter of the plane, so the angle of the vectors in the scalar product is either a right or an obtuse angle. This means that $\frac{\partial\left\|DT\cdot(0,1)^T\right\|^2}{\partial z}\le 0$.

According to these results and the already discussed shape of the set R_n one can see that the minimum must lie either on $l_2(z,m)$ ("left side" of R_n) or on $r_{n-1}(z,m)$ ("upper side" of R_n). We note that fortunately even the $n=0$ case behaves like this, because the formula $r_{-1}(z,m)$ coincides with $l_2(z,m)$. So now we investigate the derivative of the square of the expansion on the vertical direction along the curves $l_2(z,m)$ and $r_{n-1}(z,m)$. First let $h=f(z)$ be an arbitrary C^1 curve and let us calculate how $\left\|DT\cdot(0,1)^T\right\|$ behaves along this curve. Then we will apply the result for $l_2(z,m)$ and $r_{n-1}(z,m)$.

$$\frac{d\left\|DT_{(f(z),z)}\cdot(0,1)^T\right\|^2}{dz}=2\left\langle DT_{(f(z),z)}\cdot(0,1)^T,(DT'_{(f(z),z)})_z\cdot(0,1)^T\right\rangle$$

Again by the same arguments as above $DT_{(f(z),z)}\cdot(0,1)^T$ lies in the lower right quarter of the plane. The other vector has first coordinate $2m\alpha<0$ and second coordinate

$$-2\sqrt{2}\alpha(n+1)\Big(\frac{z}{\sqrt{1-\frac{f(z)}{m}+\alpha z^2}}\Big)'_z=-2\sqrt{2}\alpha(n+1)\Big(\frac{1-\frac{f(z)}{m}+\frac{zf'(z)}{2m}}{F^{3/2}}\Big)$$

where F is the usual quantity evaluated subject to $h=f(z)$. Since $\alpha<0$ the sign of the second coordinate is determined by the sign of $1-\frac{f(z)}{m}+\frac{zf'(z)}{2m}$. If this quantity is positive, then the angle of the vectors in the scalar product is an obtuse angle and therefore in this case $\frac{d\left\|DT_{(f(z),z)}\cdot(0,1)^T\right\|^2}{dz}\le 0$. Now we check positivity for $f(z)=l_2(z,m)$ and $f(z)=r_{n-1}(z,m)$.

- If $f(z)=l_2(z,m)$ then $1-\frac{f(z)}{m}+\frac{zf'(z)}{2m}=\frac{(m-1)z+\sqrt{1-mz^2+m^2z^2}}{2\sqrt{1-mz^2+m^2z^2}}>0$ because $z\le 0$ and $m-1<0$.

- If $f(z)=r_{n-1}(z,m)$ then $1-\frac{f(z)}{m}+\frac{zf'(z)}{2m}=\frac{-(m-1)(2n+1)z-\sqrt{1+(m-1)(mz^2-4n(n+1))}}{2(-1+4(m-1)n(n+1))\sqrt{1+(m-1)(mz^2-4n(n+1))}}>0$ because $z\le 0$, $m-1<0$ and $n\ge 0$.

Hence both along $l_2(z,m)$ and $r_{n-1}(z,m)$ the expansion on the vertical direction is a decreasing function of z. Combining all the results shows that the minimum is attained at $X(m,n-1)$. Its value is

$$\frac{4n^2\alpha^2m^2}{1+(1-m)n(n+2)}+(1-4n(n+1)\alpha)^2$$

by direct substitution. Since $\alpha<0$ the second term is increasing in n. Differentiating formally the first term with respect to n gives $\frac{8m^2\alpha^2n(1+n-mn)}{(1+(1-m)n(n+2))^2}$ which is positive because $m<1$ and $n\geq 0$, therefore the minimum is increasing as a function of n. □

Remark 4.4. *Later it will be useful to know the behaviour of the expansion on the vertical direction at the "lower" side of the set R_n, i.e. along the curve $r_n(z,m)$. Note that this has not been calculated yet, what we have just discussed was the behaviour of $\left\|DT|_{R_n}\cdot(0,1)^T\right\|^2$ along the curve $r_{n-1}(z,m)$ (the difference is in the index of the curve). However the same method works in this case also, hence to prove that $\left\|DT|_{R_n}\cdot(0,1)^T\right\|^2$ is decreasing as a function of z along the curve $r_n(z,m)$, it is enough to check that $1-\frac{r_n(z,m)}{m}+\frac{zr_n'(z,m)}{2m}>0$. Substituting the formula of the curve into this expression gives $\frac{-(m-1)(2n+3)z-\sqrt{(2n+3)^2+m^2z^2-m(8+12n+4n^2+z^2)}}{2(-(2n+3)^2+4m(2+3n+n^2))\sqrt{(2n+3)^2+m^2z^2-m(8+12n+4n^2+z^2)}}>0$ because $z\leq 0$, $m<1$ and $n\geq 0$. Therefore the minimum in this case is attained at $X(m,n)$. Similar calculation as in the previous lemma shows that the value of it is also an increasing function of n.*

Now we return to the verification of the growth lemma. There is only the final case left from Classification 4.2. In this case according to (4.1) it is enough to check that $\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}<1$ holds for any W that is sufficiently short. The dynamics acting on R_n has a continuous extension to the closure of this set, therefore in the limit as $|W|\to 0$ we have $\Lambda_1=\lim_{x\in R_n,x\to Q}\|DT|_{R_n}(\underline{t}(x))\|=\|DT|_{R_n}(\underline{t}(Q))\|$ where $\underline{t}(x)$ is a unit tangent vector of W at the point x, and a similar formula gives Λ_2 by replacing R_n with R_{n+1} everywhere. Hence $\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}$ depends only on $Q\in\{(r_n(z,m),z)|BX(m,n)_2\leq z\leq X(m,n)_2\}$ and n. We have to check whether the supremum of this along all curves $r_n(z,m)$ and over all natural numbers n is less than 1 or not. As the direction $\underline{t}(x)$ is unknown we can not calculate the Λ_i's explicitly. Instead we use the lower bound obtained in Lemma 4.1.

The first iterate does not expand enough. By Lemma 4.3 and Remark 4.4, the case when we have the worst estimate for $\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}$ is when an unstable manifold W (considered to be infinitesimally short) is cut by the singularity $r_0(z,m)$. In this case the estimate that we can give is $\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}\leq\frac{1}{\left\|D_{X(m,0)}T|_{R_0}\cdot(0,1)^T\right\|}+\frac{1}{\left\|D_{X(m,0)}T|_{R_1}\cdot(0,1)^T\right\|}$. The exact formula of this is quite complicated, but plotting it as a function of m between $\frac{1}{2}$ and 1 the result is quite convincing (see Figure 15). It is clear that it never goes below 1 and so we are not able to verify (4.1) in this way.

4.2 The second iterate

We proved in Subsection 3.3 that the dynamics is uniformly hyperbolic, therefore to get larger expansion rates it is a natural idea to take the second iterate of the map. In the rest of the paper we will show that (4.1) holds if we redefine Λ_i as $\inf_{x\in W_i}\frac{\left\|DT^2(\underline{t}(x))\right\|}{\|\underline{t}(x)\|}$, where $\underline{t}(x)$ is a tangent vector of W_i at the point x and W_i is the i-th piece of W when cut by the singularities of the second iterate. The price we have to pay for this is that the structure of the singularity set of T^2 is more complicated and also estimating the new Λ_i's is a much more difficult task than as it was for the first iterate. In Subsubsection 4.2.1 we chart the singularities of the second iterate and in Subsubsection 4.2.2 we give estimates on the expansion rates of DT^2. Then finally we show that for certain values of the parameter m the growth lemma holds for the second iterate.

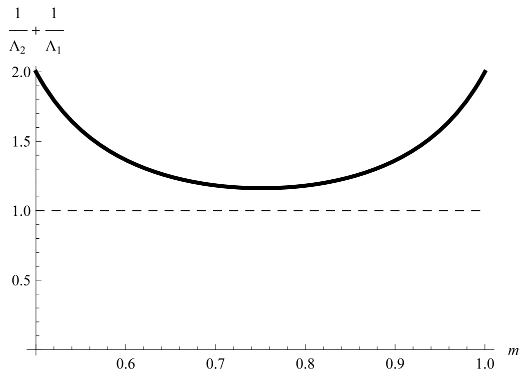


Figure 15: The first iterate does not expand enough

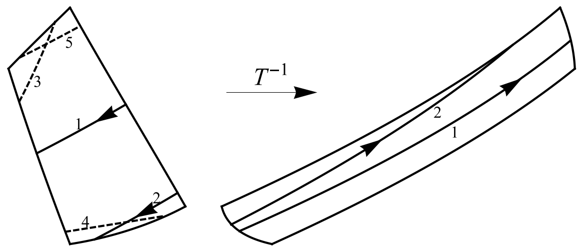


Figure 16: Preimage of singularities

4.2.1 Singularities of the second iterate

For the first iterate the singularity set is $S_1 = \partial\mathcal{M}_1 \cup (\cup_{n=0}^{\infty} r_n(z,m))$ which we have already understood well. For the second iterate the singularities are $S_2 = S_1 \cup T^{-1}(S_1)$. We have

$$T^{-1}(S_1) = T^{-1}(\partial\mathcal{M}_1) \cup T^{-1}\left(\cup_{n=0}^{\infty} r_n(z,m)\right) = \left(\cup_{n=0}^{\infty} r_n(z,m)\right) \cup \left(\cup_{k=0}^{\infty} T^{-1}(r_k(z,m))\right)$$

The first part consists of singularities inherited from the first iterate, so the new object to understand is $T^{-1}(r_k(z,m))$ for each k. To do so we first take the image $T(\mathcal{M}_1) = \cup_{n=0}^{\infty} T(R_n)$. At each point $x \in \mathcal{M}_1$ the Jacobian DT_x when viewed as a matrix acting on $\mathbb{R}^2$ has one stable and one unstable eigenvalue, both of which are negative. The corresponding unstable and stable eigendirections have negative and positive slopes, respectively. Hence under the action of T the stripe R_n gets contracted in the stable direction, expanded in the unstable direction and finally point reflected. So the stripe $T(R_n)$ will be rather vertical, its "horizontal" sides are sections of the boundary of $\mathcal{M}_1$ and its "vertical" sides are the curves $Ir_{n-1}(z,m)$ and $Ir_n(z,m)$ (see Figure 10(a)). For a fixed k the curve $r_k(z,m)$ crosses some of the "vertical" stripes (see Figure 17(a)). Except for a countable set of mass ratios (which can be exactly calculated) it can intersect only finitely many of them. As for the geometry of a nonempty intersection $r_k(z,m) \cap T(R_n)$ for a fixed pair of integers k and n, in principle five different regimes can be observed. The reader should take a look at the left side of Figure 16, where we numbered these cases explicitly. However some of the lines are dashed, because for a certain set of m's they can not exist as we will show next. Let us go through the five regimes.

Classification 4.5. *1. In a sense the most typical case is when $r_k(z,m)$ intersects both "vertical" sides of $T(R_n)$. The intersection $r_k(z,m) \cap T(R_n)$ is just a subcurve of $r_k(z,m)$ and will have different preimages for different n's. The reader might check (Figure 10(a) will help) that for a given n the preimage will be a curve in R_n running from one "vertical" side to the other, but with opposite orientation as $r_k(z,m)$ had (see Figure 16).*

2. *This kind of intersection appears only when the left end of $r_k(z,m)$ (which is the point $BX(m,k)$) lies on the lower boundary of $T(R_n)$. As this corresponds to the endpoint of $r_k(z,m)$, for every k there is a unique n for which this happens (on the other hand for a fixed n there may be $0, 1$ or several k's with this property). The preimage of the intersection will be a curve in R_n running from its "left side" to its "top" with opposite orientation as $r_k(z,m)$ had. For a better understanding take a look at Figure 16 again.*

3. *This case would happen if the right end of $r_k(z,m)$ (which is the point $X(m,k)$) was on the upper boundary of $T(R_n)$. But this is impossible because $X(m,k)$ always lies on the curve $l_1(z,m)$, while the upper boundary of $T(R_n)$ is always a subcurve of $l_2(z,m)$.*

4. *One can see such an intersection if the right end of $r_k(z,m)$, i.e. the point $X(m,k)$ lies on the lower boundary of $T(R_n)$. This is only possible if the accumulation point of the singularities P lies in $T(R_i)$ for some $i \geq n$. The lower right corner of $T(R_1)$ is $IBX(m,0) = (\frac{m}{2}, 2)$ and P has coordinates $(2(1-m)m, \frac{-1}{\sqrt{1-m}})$. Both points lie on the curve $l_1(z,m)$ and clearly for $m \in (\frac{1}{2}, \frac{3}{4})$ P has larger second coordinate than $IBX(m,0)$. This means that for $m \in (\frac{1}{2}, \frac{3}{4})$ the accumulation point is in $T(R_0)$ which exceptionally does not have a lower boundary, hence for this range of the mass ratio parameter an intersection of kind 4 can not exist.*

5. *Finally an intersection like this arises, when the left end of $r_k(z,m)$, which is the point $BX(m,k)$ lies on the upper boundary of $T(R_n)$. We now show that this can not happen either for m in $(\frac{1}{2}, \frac{3}{4})$. The upper left corner of $T(R_n)$ is the point $IX(m,n)$. The sequence of these points has a limit as n goes to infinity, which is $(0, \frac{-1}{\sqrt{1-m}})$, and the points have decreasing second coordinates. One can check by its formula, that $BX(m,k)$ has strictly smaller second coordinate than $\frac{-1}{\sqrt{1-m}}$ for all $k \geq 0$ if $m \in (\frac{1}{2}, \frac{3}{4})$ and so within this range of the parameter m such an intersection can not take place.*

From now on we restrict to the case when $\frac{1}{2} < m < \frac{3}{4}$. According to Classification 4.5 the singularity set of the second iterate has the structure displayed on Figure 17(b). There are two accumulation points of singularity curves. The accumulation point P, which is already present for the first iterate, lies on the left side of $T(R_0)$, which is part of the boundary of $\mathcal{M}_1$. Hence the other accumulation point, the preimage of P (to be denoted by P') is somewhere on the top of R_0, which is again a part of the boundary of $\mathcal{M}_1$. So there are infinitely many curves in R_0 running from its one side to the other. They intersect neither $r_0(z,m)$ nor each other and accumulate in P', on the top of R_0.
As for any other stripe R_k, $k \geq 1$, let us describe the singularity curves from top to bottom. The upper boundary of R_k is the curve $r_{k-1}(z,m)$ which is a singularity inherited from the first iterate. Then there are finitely many curves running from the the "left" side of R_k to its "top", terminating on $r_{k-1}(z,m)$; and finitely many curves running from one "vertical" side of R_k to the other. These curves are singularities of T^2 but not of T. Finally there is the lower boundary $r_k(z,m)$, which is singular already in the first step.
As a crucial consequence **at any point of the phase space at most one singularity curve can terminate on another singularity curve**. Consequently, **apart from** the case when it runs into one of **the accumulation points** of S_2, a **sufficiently short unstable manifold W can be cut by the singularities into at most three pieces**.

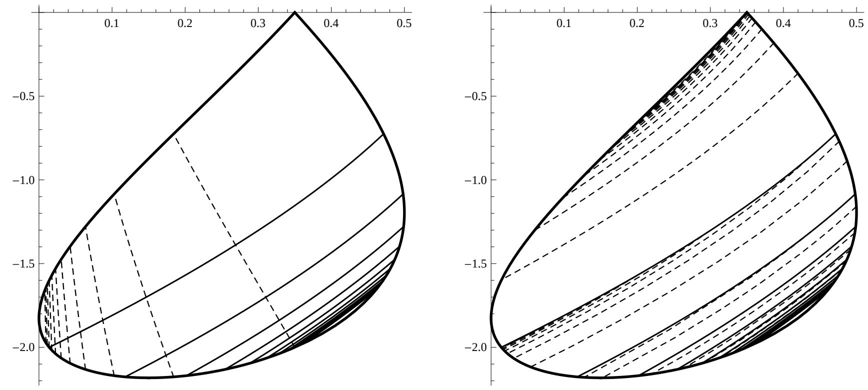
4.2.2 Growth lemma for the second iterate

We want to show that (4.1) holds for T^2. After two iterations of T, a sufficiently short unstable manifold will either remain connected, or will be cut by the singularities into two, three (which is a new phenomena) or infinitely many pieces. In case it remains connected or when it is cut into infinitely many pieces, the arguments of section 4.1 (see Classification 4.2) apply, because T^2 has even larger expansion rates than T.

The other two situations require much more work. One has to estimate the expansion rates of DT^2 along the singularity curves $\{r_n(z,m)\}$ and $\{T^{-1}(r_n(z,m))\}$. The formulas for $\{T^{-1}(r_n(z,m))\}$, however, turn out to be much more complicated than the formulas for $\{Ir_n(z,m)\}$. Hence it is worth iterating one step forward to obtain the geometry of Figure 17(a): in general for $x \in R_n \cap T^{-1}(R_k) = T^{-1}(TR_n \cap R_k)$ (if this set is not empty) we will calculate the Jacobian D_xT^2 in terms of its image $y = Tx$, that is, as a function of $y \in TR_n \cap R_k$. From now on we use the notation $D_xT_n := D_xT|_{R_n}$, furthermore, for

$$D_x(T^2) = D_{Tx}T_k \cdot D_xT_n = D_{Tx}T_k \cdot (D_{Tx}T_n^{-1})^{-1} = D_yT_k \cdot (D_yT_n^{-1})^{-1},$$

which is to be investigated, we introduce the shorthand notation $D_yT^{k,n}$.



(a) The sets $TR_n \cap R_k$ (b) Singularities for T^2

Figure 17: T^{-1} maps figure (a) to figure (b)

The case when a short unstable manifold W is cut into two pieces:
We have seen that the growth lemma fails for the first iterate along the singularity $r_0(z,m)$. However for certain values of the parameter m this is the only location where it fails. As a consequence of Lemma 4.1 and Lemma 4.3 we can say that

$$\forall x \in \cup_{n=1}^{\infty} R_n = \mathcal{M}_1 \setminus R_0 \text{ and } \forall \underline{v} \in C_x^u, \quad \|D_xT_n \cdot \underline{v}\| > 2\,\|\underline{v}\|$$

if

$$\left\|D_{X(m,0)}T_1 \cdot (0,1)^T\right\| > 2. \tag{4.4}$$

By direct substitution this is equivalent to the inequality

$$\frac{196 - 1491m + 4212m^2 - 5496m^3 + 3380m^4 - 816m^5 + 16m^6}{4-3m} > 4$$

which holds for $m = 0.74$ for example. So if (4.4) is satisfied and either k or n is different from 0, then for a point $x \in R_n \cap T^{-1}(R_k)$ the expansion of the second iterate along the vertical direction is surely larger than 2 (we used Lemma 4.3 and the uniform hyperbolicity of the dynamics). Therefore if W is disjoint from the set $R_0 \cap T^{-1}(R_0)$ then the infimum of the expansion of DT^2 is larger than 2 on it and hence for curves that are cut into two pieces the growth lemma immediately holds, because then $\frac{1}{\Lambda_1} + \frac{1}{\Lambda_2} < \frac{1}{2} + \frac{1}{2} = 1$. If $W \cap (R_0 \cap T^{-1}(R_0)) \neq \emptyset$ and it will be cut within two iterations into two pieces, then there are two different scenarios:

- W intersects either $r_0(z,m)$ or,
- W intersects the segment of the preimage of $r_0(z,m)$, which lies in R_0.

First we consider the case when it intersects the above mentioned preimage of $r_0(z,m)$. Denote the intersection by Q. As we suggested in the beginning of this subsubsection we iterate the situation one step forward. The image of W remains connected and it will cross the curve $r_0(z,m) \cap T(R_0)$. To estimate the expansion rate of the second iterate at Q along W we will give lower bounds on $\left\|D_{T(Q)}T^{0,0} \cdot (0,1)^T\right\|$ and $\left\|D_{T(Q)}T^{1,0} \cdot (0,1)^T\right\|$ for $T(Q) \in r_0(z,m) \cap T(R_0)$. For this purpose we will use our knowledge on the expansion rate of the first iterate along the vertical direction and the invariance of unstable cones, but further preparations will be needed also. In general

$$\left\|D_xT^{k,n} \cdot (0,1)^T\right\|^2 = \left\langle D_xT_k(a(1,0)^T + b(0,-1)^T), D_xT_k(a(1,0)^T + b(0,-1)^T)\right\rangle =$$
$$a^2\left\|D_xT_k \cdot (1,0)^T\right\|^2 + b^2\left\|D_xT_k \cdot (0,-1)^T\right\|^2 + 2ab\left\langle D_xT_k \cdot (1,0)^T, D_xT_k \cdot (0,-1)^T\right\rangle \quad (4.5)$$

where $a > 0$ is the first, while $b > 0$ is the second coordinate of the vector $(D_xT_n^{-1})^{-1} \cdot (0,1)^T$. To obtain the lower bound, we will estimate below the three terms separately, for which purpose we will use the arguments in the proof of Lemma 4.3 and the forthcoming two lemmas repeatedly.

Lemma 4.6. *Both a and b have gradient vector lying in the upper right quarter of the plane.*

Proof. We perform the calculations for a and b separately.

- The exact formula for a is $a = (1,0) \cdot (D_xT_n^{-1})^{-1} \cdot (0,1)^T = -4\sqrt{2}\sqrt{hm}(n+1)\alpha - 2mz\alpha$. Therefore its gradient is the vector $(-\frac{2\sqrt{2}m(n+1)\alpha}{\sqrt{hm}}, -2m\alpha)$. Since $\alpha < 0$ this is indeed in the upper right quarter of the plane.

- The exact formula for b is $b = (0,1) \cdot (D_xT_n^{-1})^{-1} \cdot (0,1)^T = 1 - 8(n+1)^2\alpha - \frac{2\sqrt{2}m(n+1)\alpha z}{\sqrt{hm}}$. Its gradient is $(\frac{\sqrt{2}m^2(n+1)\alpha z}{(hm)^{3/2}}, -\frac{2\sqrt{2}m(n+1)\alpha}{\sqrt{hm}})$. Since both α and z are negative, this is again a vector in the upper right quarter of the plane.

□

Lemma 4.7. *The monotonicity of $\left\|D_{(g(z),z)}T_n \cdot (1,0)^T\right\|$ along the curve $h = g(z)$ is the opposite of the monotonicity of the quantity F along the same curve.*

Remark 4.8. *The monotonicity of F was discussed in Subsection 3.2.*

Proof. At the point $(h,z) \in \overline{R_n}$ we have $\left\|DT|_{R_n} \cdot (1,0)^T\right\|^2 = 1 + \frac{2(n+1)^2}{m^2F}$, from which the statement immediately follows. □

As a final step of the preparation before we derive the estimates, we give a lower bound on the cosine of the angle between the vectors $D_xT_k \cdot (1,0)^T$ and $D_xT_k \cdot (0,-1)^T$. As we showed in Corollary 3.12 both of these vectors lie in the unstable cone $C^u_{Tx,1}$. Hence $\cos(\angle(D_xT_k \cdot (1,0)^T, D_xT_k \cdot (0,-1)^T)) \geq \cos(\arctan(\frac{\sqrt{m}}{2})) = \frac{2}{\sqrt{4+m}}$.

Consider the curve $r_0(z,m) \cap T(R_0)$. We know that its right endpoint is $X(m,0)$ and its left end can be calculated by solving the equation $r_0(z,m) = Ir_0(z,m)$. This will lead to the point

$$J(m) = \left(\frac{m(24-24m+\sqrt{-558+54\sqrt{-7+16m}+16m(144-172m+64m^2-3\sqrt{-7+16m})})^2}{4(9-8m)^2(-31+8(13-8m)m+3\sqrt{-7+16m})}, \frac{-1}{\sqrt{2}}\sqrt{\frac{31-104m+64m^2+3\sqrt{-7+16m}}{-16+103m-231m^2+208m^3-64m^4}}\right).$$

We want to give a lower bound, using the representation (4.5), on $\Lambda_1 = \left\|D_xT^{0,0} \cdot (0,1)^T\right\|^2$ and on $\Lambda_2 = \left\|D_xT^{1,0} \cdot (0,1)^T\right\|^2$ for every x along the curve $r_0(z,m) \cap T(R_0)$. Then both in the estimate for Λ_1 and in the estimate for Λ_2 we make the following observations.

- Since this curve is *increasing*, the above introduced quantities a and b both have their minimum at $J(m)$ due to Lemma 4.6.
- According to Lemma 4.7 the minimum of the expansion on the horizontal direction is also attained in $J(m)$.
- From the argument in the proof of Lemma 4.3 it follows that the expansion along the vertical direction is minimal at $X(m,0)$.

Putting together these with the lower bound on the cosine leads to the following estimates:

- $\left\|D_xT_0\cdot(D_xT_0^{-1})^{-1}\cdot(0,1)^T\right\|^2 \geq a^2(J(m))\left\|D_{J(m)}T_0\cdot(1,0)^T\right\|^2 + b^2(J(m))\left\|D_{X(m,0)}T_0\cdot(0,-1)^T\right\|^2 + {} + 2a(J(m))b(J(m))\left\|D_{J(m)}T_0\cdot(1,0)^T\right\|\left\|D_{X(m,0)}T_0\cdot(0,-1)^T\right\|\frac{2}{\sqrt{4+m}}$ Let us denote this function of m as $ExpSquare00(m)$.
- $\left\|D_xT_1\cdot(D_xT_0^{-1})^{-1}\cdot(0,1)^T\right\|^2 \geq a^2(J(m))\left\|D_{J(m)}T_1\cdot(1,0)^T\right\|^2 + b^2(J(m))\left\|D_{X(m,0)}T_1\cdot(0,-1)^T\right\|^2 + {} + 2a(J(m))b(J(m))\left\|D_{J(m)}T_1\cdot(1,0)^T\right\|\left\|D_{X(m,0)}T_1\cdot(0,-1)^T\right\|\frac{2}{\sqrt{4+m}}$ Note that this slightly differs from the one above. The difference is that we have used the restriction of DT to the set R_1, while in the previous case it was restricted to R_0. Let us denote this estimate as $ExpSquare01(m)$.

Now the growth lemma is satisfied in this case if

$$\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}\leq\frac{1}{\sqrt{ExpSquare00(m)}}+\frac{1}{\sqrt{ExpSquare01(m)}}<1 \tag{4.6}$$

The reader can check that this holds for $m=0.74$ in particular and the plot of this function (Figure 18) shows that actually it is valid for a quite large set of m's.

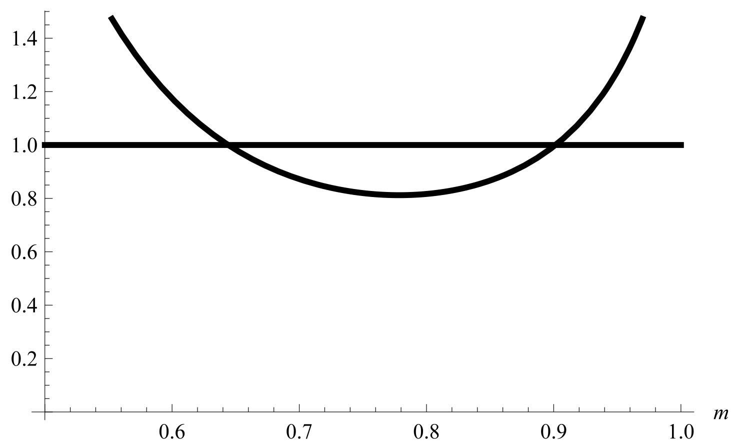


Figure 18: Plot of the estimate (4.6)

In the other case when W intersects the curve $r_0(z,m)$, its image under T is cut into two pieces immediately. The point $Q=W\cap r_0(z,m)$ is singular in this case, it has two images. One of them, say $T(Q)_1$ will be a point on the curve $l_2(z,m)$ between $IX(m,0)$ and $IX(m,-1)=(\frac{m}{2},0)$, while the other, say $T(Q)_2$ will be a point on the lower edge of the phase space between the points $IBX(m,1)$ and $IBX(m,0)$. The unstable manifold W can be arbitrarily short, so to verify the growth lemma it is enough to check that the sum of the inverse of $\left\|D_{T(Q)_1}T^{0,0}\cdot(0,1)^T\right\|$ and of $\left\|D_{T(Q)_2}T^{k,1}\cdot(0,1)^T\right\|$ is less than 1 for every Q lying on $r_0(z,m)$. In the second expression the index k is determined as the unique integer for which $T(Q)_2\in R_k$. Now we give a lower bound on this k for certain values of m.

Lemma 4.9. *If $\frac{2}{3} < m < \frac{3}{4}$ then the unique integer k for which $T(Q)_2 \in R_k$ is at least 2.*

Proof. We already mentioned that for $m < \frac{3}{4}$ the points $BX(m, k-1)$ (which are the upper left corner points of R_k) lie on the lower edge of the phase space. Their first coordinates are increasing as a function of k. The first coordinate of $T(Q)_2$ is at least $IBX(m,1)_1$, therefore the unique k_0 that satisfies $BX(m, k_0 - 1)_1 < IBX(m,1)_1 \le BX(m, k_0)$ will be a lower bound on k. If $\frac{2}{3} < m < \frac{3}{4}$ then these inequalities reduce to $\frac{7-12m}{6m-5} < k_0 < \frac{2-6m}{6m-5}$. The lower bound is an increasing function of m and for $m = \frac{2}{3}$ it takes the value 1, hence if $\frac{2}{3} < m$ then $k_0 \ge 2$ (since it is an integer). □

By using the monotonicity statement from Lemma 4.3 and Lemma 4.9, $\left\| D_{T(Q)_2} T^{k,1} \cdot (0,1)^T \right\|$ is greater than the product of the minimal expansion rates along the vertical direction on the set R_1 and R_2. We know that if (4.4) holds then this is larger than 4.

So what left is to show that $\left\| D_{T(Q)_1} T^{0,0} \cdot (0,1)^T \right\|$ is larger than $\frac{4}{3}$. This will be verified again using the representation (4.5). Recall that $T(Q)_1$ is a point on the curve $l_2(z, m)$ between the points $IX(m, 0)$ and $IX(m, -1)$. Since this curve is *increasing* in z, we have the following:

- Lemma 4.6 tells us that the minimum of a and b is attained at the point $IX(m, 0)$.
- The same holds for the minimum of the one-step expansion rate along the horizontal direction due to Lemma 4.7.
- On the other hand the expansion along the vertical direction attains its minimum at $IX(m, -1) = (\frac{m}{2}, 0)$ and its value is 1.[2]

Putting all these results together leads to the estimate

$$\begin{aligned}
\left\| D_{T(Q)_1} T^{0,0} \cdot (0,1)^T \right\|^2 \ge a^2(IX(m,0)) \left\| D_{IX(m,0)} T_0 \cdot (1,0)^T \right\|^2 + \\
+ b^2(IX(m,0)) \left\| D_{IX(m,-1)} T_0 \cdot (0,-1)^T \right\|^2 + 2a(IX(m,0)) b(IX(m,0)) \cdot \\
\cdot \left\| D_{IX(m,0)} T_0 \cdot (1,0)^T \right\| \left\| D_{IX(m,-1)} T_0 \cdot (0,-1)^T \right\| \frac{2}{\sqrt{4+m}} \quad (4.7)
\end{aligned}$$

Let us denote this function of m as $ExpSquareAtTop(m)$. The reader can check that this is larger than $\left(\frac{4}{3}\right)^2$ in particular for $m = 0.74$ and the plot of it (Figure 19) shows that the inequality holds for a much larger set of m's.

To complete the proof the growth lemma we still need to investigate **the case when a short unstable manifold is cut into three by the singularities.**

When restricting to $\frac{1}{2} < m < \frac{3}{4}$ such a situation occurs if W is close to a point where a preimage of a singularity ends on another singularity. Let us denote the three pieces into which the unstable manifold W is cut by W_i, $i = 1, 2, 3$. Due to the description of the singularity structure of the second iterate (see case No.2) there is an integer k such that $W_1 \subset R_k$ and $W_2 \cup W_3 \subset R_{k+1}$. We will show that for an open set of m values $\Lambda_1 > 2$ while both $\Lambda_2 > 4$ and $\Lambda_3 > 4$, which is enough for (4.1).

[2] To avoid confusion let us recall from section 3.3 that the vertical direction is excluded from the unstable cone at the point $(\frac{m}{2}, 0)$, hence this does not contradict uniform expansion of the first iterate. However, the only way to obtain information on the minimum expansion rate is by considering expansion on the vertical direction.

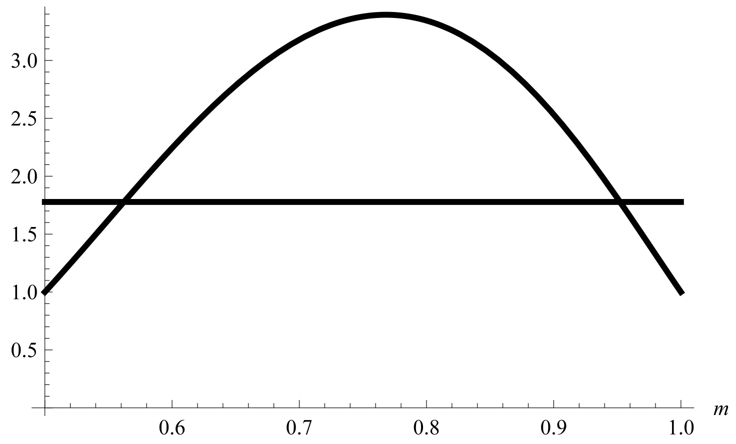


Figure 19: Plot of the estimate (4.7)

- First we consider the case when $k \geq 1$. As a consequence of Lemma 4.1 and Lemma 4.3 we can say that

$$\forall x \in \cup_{n=2}^{\infty} R_n = \mathcal{M}_1 \setminus (R_0 \cup R_1) \text{ and} \forall \underline{v} \in C_x^u, \quad \|D_x T_n \cdot \underline{v}\| > 4 \|\underline{v}\|$$

if

$$\left\| D_{X(m,1)} T_2 \cdot (0,1)^T \right\| > 4. \tag{4.8}$$

By direct substitution this is equivalent to the inequality

$$\frac{4761 - 34040m + 93040m^2 - 121440m^3 + 76240m^4 - 18624m^5 + 64m^6}{9 - 8m} > 16.$$

Now if $W \subset (R_k \cup R_{k+1})$ for $k \geq 1$, and the mass ratio m is such that both (4.4) and (4.8) holds (which happens when $m = 0.74$ for example), then (4.1) is satisfied (by $\Lambda_1 > 2$, $\Lambda_2 > 4$ and $\Lambda_3 > 4$).

- The absolutely last remaining case is when $k = 0$. Note that, by (4.7), for an open set of mass ratios we have already obtained the lower bound $\Lambda_1 > \frac{4}{3}$ in this situation (cf. also Figure 19). To estimate Λ_2 and Λ_3 we iterate W one step forward. Its subcurve $W_2 \cup W_3$ remains connected after one step and will be attached to the lower side of the phase space. It came from the set R_1 so assuming (4.4) we know that it got expanded by a factor larger than 2. Now as $W_2 \cup W_3$ is attached to the lower side of $\mathcal{M}_1$, it belongs to a set R_k for some $k \geq 2$ by Lemma 4.9. Hence if we assume (4.8), iterating it one more step forward it gets expanded by an additional factor larger than 4 . Summarizing, for an open set of m values, we have $\Lambda_2 > 8$, $\Lambda_3 > 8$ and $\Lambda_1 > \frac{4}{3}$, which is enough for (4.1).

In a summary, if all the inequalities (4.4), (4.6), (4.7) and (4.8) hold for a specific value of m, then the growth lemma is satisfied for this value of the parameter. Moreover since all these are open conditions if they hold for an m_0, then there exists an $\varepsilon > 0$ such that all of them remain valid on the interval $(m_0 - \varepsilon, m_0 + \varepsilon)$. We checked that these conditions hold for $m = 0.74$ so we can conclude that the growth lemma is satisfied on the interval $I = (0.74 - \varepsilon, 0.74 + \varepsilon)$ for an appropriate $\varepsilon > 0$.

5 Outlook

In this last section we mention two possible directions of future research closely related to our results.

Extension of the results to a larger set of mass ratios. As we have pointed out in section 1.2, the only point where we need to put restriction on the value of m is the verification of Assumption A7 (the growth lemma). The growth lemma is about controlling the competition between expansion of unstable curves and their partitioning by the singularities of the dynamics. As we have done it in section 4, we may increase the expansion of unstable curves by considering T^n, a fixed higher iterate of the map T. When doing so, we need to treat the more complicated singularity set S_n of T^n and it can happen that short unstable curves are partitioned into more pieces. For each n, the set S_n consists of countably many smooth curves (singularity curves). There are two effects by which S_n may partition a short unstable curve:

- *accumulation* of singularity curves. We have already discussed this phenomenon in Classification 4.2: for a short unstable manifold partitioned by such an accumulating structure, the expansion rates of the relevant pieces diverge as $\Lambda_k \sim k^2$, so that $\sum(\Lambda_k)^{-1}$ is summable, which is suitable for proving the growth lemma.
- *branching* of S_n, when certain singularity curves $\gamma' \subset S_n$ terminate on another singularity curve $\gamma \subset S_n$. In such a case, a short unstable curve can be partitioned into several pieces, on which it is hard to get explicit lower bounds on the expansion.

A useful notion to quantify this branching effect is as follows:

Definition 5.1. *For each $x \in \mathcal{M}_1$, let $k_n(x)$ denote the number of smooth curves $\gamma \subset S_n$ (singularity curves) such that $x \in \gamma$. The n-step complexity of S_n is defined as $K_n = \sup_{x\in\mathcal{M}_1} k_n(x)$. We refer to the growth rate of K_n as the growth of the complexity for the singularity set; in particular, by "subexponential complexity" we mean that $K_n = o(\alpha^n)$ for any $\alpha > 1$; and by "finite complexity" we mean that there exists some $K > 0$ such that $K_n \leq K$ for all n.*

For further accounts on complexity we refer to [7] and references therein. What is important for us is that the number of pieces into which a sufficiently short unstable manifold may be cut as a consequence of the branching of S_n is at most $K_n + 1$. On the other hand, the minimum expansion of the T^n on unstable curves is Λ^n, where $\Lambda > 1$ is the expansion constant from Lemma 3.8. As a consequence, if the complexity is subexponential, there exists n_0 such that $K_{n_0} + 1 < \Lambda^{n_0}$; then the growth lemma should follow for T^{n_0} as we have at most $K_{n_0} + 1$ terms, each of which is at most Λ^{-n_0}. Now we formulate two conjectures concerning the extension of our results to a larger class of mass ratios $m \in (1/2, 1)$.

Conjecture 5.2. *For all $m \in (1/2, 1)$ the singularity set has subexponential complexity. Consequently, the statements of Theorems 1.1, 1.2 and 1.3 hold for all $m \in (1/2, 1)$.*

Conjecture 5.3. *There exists a countable set $\Sigma \subset (1/2, 1)$ such that for $m \in (1/2, 1) \setminus \Sigma$ the complexity is finite. This way Theorems 1.1, 1.2 and 1.3 extend to all $m \in (1/2, 1) \setminus \Sigma$.*

Conjecture 5.2 would have stronger implications, but it seems more difficult to prove. In fact, subexponential complexity is often assumed in the literature (see eg. [6],[10],[1]). On the other hand, subexponential (actually, linear) bounds on the complexity were obtained by Bunimovich for two dimensional Sinai billiards without corner points ([8], [14]). Nonetheless, Bunimovich's argument does not seem to be generalizable to the system of falling balls as it relies, in particular, on the continuity of the flow which we do not have in our context. Proving Conjecture 5.3 seems to be a more realistic goal; such an argument could use the detailed geometric analysis of the present paper. We plan to address this question in a separate project.

Decay of correlations for the flow. An important open problem is to obtain bounds on the rate of mixing for the flow $(\tilde{\mathcal{M}}, S^t, \tilde{\mu})$. In fact, we expect the flow to mix faster than the map, as many consecutive applications of F_2 (long series of bounces of the lower ball on the floor before colliding with the upper ball) take place within uniformly bounded flow time. Similar phenomena has been observed in dispersing billiards with cusps ([4]). In particular, we expect that the following result from [26] applies to $(\tilde{\mathcal{M}}, S^t, \tilde{\mu})$.

Theorem 5.4 ([26]). *Consider a suspension flow with (i) base transformation that can be modelled by a Young tower with exponential tails; and with a roof function that is (ii) uniformly piecewise Hölder continuous and (iii) non-integrable (in an appropriate sense). Then the flow is rapid mixing: in continuous time, correlations for sufficiently regular observables decay faster than any polynomial.*

We refer to the original reference [26] for what "non-integrable roof function" and "sufficiently regular observable" means in this context. What we would like to point out that conditions (i) and (ii) have been verified for $(\tilde{\mathcal{M}}, S^t, \tilde{\mu})$ in the present paper. In fact, Theorem 5.4 was the main motivation for studying the Hölder regularity of the induced roof function τ on $\mathcal{M}_1$ (cf. Remark 1.11). The missing ingredient is (iii), the non-integrability of the roof function. In the context of hyperbolic billiards non-integrability of the roof function is a consequence of the fact that the billiard flow preserves a contact structure (see [4] and references therein). In the system of falling balls this is not available because of the presence of the gravity field. Hence non-integrability of the roof function should be studied directly, for which the geometric analysis of the present paper may provide a good basis.

Acknowledgements

We are grateful to Imre Péter Tóth for his encouragement and enlightening discussions. The work reported in this paper has been developed in the framework of the project "Talent care and cultivation in the scientific workshops of BME" project. This project is supported by the grant TÁMOP - 4.2.2.B-10/1–2010-0009. Our research is also connected to the scientific program of the "Development of quality-oriented and harmonized R+D+I strategy and functional model at BME" project. This project is supported by the New Széchenyi Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002). Part of this work was done while P.B. and A.N.V visited the Fields Institute, Toronto; and while G.B. and A.N.V. visited the MPIPKS Dresden; we thank both institutions for their hospitality and stimulating research atmosphere. The financial support of the Bolyai Scholarship of the Hungarian Academy of Sciences (P.B.), and OTKA (Hungarian National Fund for Scientific Research), grant K71693 (all authors) is thankfully acknowledged.

Appendix

In the following section we collect the conditions of Chernov and Zhang from [15], which ensure that a map can be modeled by a Young tower with exponential tails. To prove Proposition 1.5, we need to verify that the first return map $(\mathcal{M}_1, T, \mu)$ satisfies these assumptions.

A0 Phase space

We consider a map $\mathcal{T} : M \mapsto M$ where the phase space M is an open domain in a two-dimensional smooth compact Riemannian manifold with or without boundary.

A1 Piecewise Smoothness

The map $\mathcal{T}$ is a C^2 diffeomorphism of $M \setminus \mathcal{S}$ onto $\mathcal{T}(M \setminus \mathcal{S})$, where $\mathcal{S}$ is a countable collection of smooth and compact curves. Usually, $\mathcal{S}$ is the set of points at which $\mathcal{T}$ either is not defined or is singular (discontinuous or not differentiable).

A2 Hyperbolicity

The map $\mathcal{T}$ is uniformly hyperbolic in the following sense.

- There exist two families of cones $\{\mathcal{C}^u_x\}_{x\in M}$ (unstable) and $\{\mathcal{C}^s_x\}_{x\in M}$ (stable) such that the unstable is *strictly* forward invariant and the stable is *strictly* backward invariant: $\mathrm{D}_x\,\mathcal{T}\,(\mathcal{C}^u_x) \subset (int\,\mathcal{C}^u_{\mathcal{T}x}) \cup \{0\}$ and $int(\mathrm{D}_x\,\mathcal{T}\,(\mathcal{C}^s_x)) \cup \{0\} \supset \mathcal{C}^s_{\mathcal{T}x}$ if the tangent map exists.

- The expansions have to be uniformly bounded away from 1 in the following way. There is a $\Lambda > 1$ such that $\| \mathrm{D}_x \mathcal{T}(v)\| \geq \Lambda \|v\|$ for every $v \in \mathcal{C}_x^u$ and $\| \mathrm{D}_x \mathcal{T}^{-1}(v)\| \geq \Lambda \|v\|$ for every $v \in \mathcal{C}_x^s$ (whenever the tangent map exists).

- The two cone fields are uniformly transversal, that is, the angle between C_x^u and C_x^s is uniformly bounded away from zero.

- For all $n > 0$, tangent vectors to the singularity curves of $\mathcal{T}^n$ must lie in stable cones, and tangent vectors to the singularity curves of $\mathcal{T}^{-n}$ must lie in unstable cones.

By unstable manifold we mean a curve $W^u \subset M$ such that for all $n \geq 1$, $\mathcal{T}^{-n}$ is well-defined and smooth on W^u and for all $x, y \in W^u$ $d(\mathcal{T}^{-n}x, \mathcal{T}^{-n}y) \to 0$ exponentially fast as $n \to \infty$. Stable manifolds can be defined analogously. In [19] the existence of stable and unstable manifolds is shown under mild technical conditions (which have been verified for the system of falling balls in [34]) for hyperbolic systems with singularities. That is, with respect to any $\mathcal{T}$-invariant probability measure ν_0', almost every point $x \in M$ has an unstable and a stable manifold, which we denote by $W^u(x)$ and $W^s(x)$, respectively.

A3 SRB-measure
The map $\mathcal{T}$ preserves an ergodic and mixing measure ν, whose conditional distributions on unstable manifolds are absolutely continuous.

A4 Bounded Curvature
The curvature of unstable manifolds is uniformly bounded by a constant $B \geq 0$.

A5 Distortion Bounds
Fix $n \in \mathbb{N}$ arbitrary, and consider an unstable manifold W^u so short that $\mathcal{T}^n$ is well-defined and smooth on W^u. For $x, y \in W^u$, let $\mathrm{dist}(x, y)$ denote the distance of the two points within the curve W^u, and let $\mathcal{J}\mathcal{T}(x)$ denote the expansion factor of $\mathcal{T}$ along W^u in the point x. Then we have:

$$\left| \log \prod_{i=0}^{n-1} \frac{\mathcal{J}\mathcal{T}(\mathcal{T}^i x)}{\mathcal{J}\mathcal{T}(\mathcal{T}^i y)} \right| \leq \psi\left(\mathrm{dist}\left(\mathcal{T}^n x, \mathcal{T}^n y\right)\right)$$

where ψ is a function independent of W^u and n, such that $\lim_{s \to 0} \psi(s) = 0$.

A6 Absolute continuity
If W_1, W_2 are two small unstable manifolds close to each other, then the holonomy map $h : W_1 \mapsto W_2$ (defined by sliding along stable manifolds) is absolutely continuous with respect to the Lebesgue measures ν_{W_1}, ν_{W_2} , and its Jacobian is bounded:

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

for some $C > 0$. Here $W_1' \subset W_1$ denotes the subset where h is defined.

A7 Growth Lemma
To formulate the last condition, we recall the notion of the local expansion factor of $\mathcal{T}$ for $x \in M$, see Notation 3.3: $\mathcal{J}\mathcal{T}x$ is the expansion factor of $\mathcal{T}$ at x on the vector tangent to the unstable manifold $W^u(x)$. The growth lemma, or one-step expansion estimate requires that

$$\liminf_{\delta_0 \to 0} \sup_{W : |W| < \delta_0} \sum_i \Lambda_i^{-1} < 1$$

where

- the supremum is taken over unstable manifolds in M,

- i is indexing the connected components W_i of $W \setminus \mathcal{S}$,

- Λ_i is the minimum of the local expansion factor on the connected component W_i.

References

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

[2] P. Bálint, N. Chernov and D. Dolgopyat, *Limit Theorems for Dispersing Billiards with Cusps* , Comm. Math. Phys. **308** (2011) 479–510.

[3] P. Bálint and S. Gouëzel, *Limit theorems in the stadium billiard*, Comm. Math. Phys. **263** (2006) 461–512.

[4] P. Bálint and I. Melbourne; *Decay of correlations and invariance principles for dispersing billiards with cusps, and related planar billiard flows*, J. Stat. Phys. **133** (2008), 435–447.

[5] P. Bálint and I.P. Tóth, *Correlation decay in certain soft billiards*, Comm. Math. Phys. **243** (2003) 55–91.

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

[7] P. Bálint and I.P. Tóth, *Example for Exponential Growth of Complexity in a Finite Horizon Multidimensional Dispersing Billiard*, preprint, available at `http://www.math.bme.hu/~pet/pub/ellenpelda_20110913.pdf`

[8] L. A. Bunimovich, Ya. G. Sinai and N. I. Chernov N. I. *Statistical properties of two-dimensional hyperbolic billiards*, Russ. Math. Surv. **46** (1991), 47–106.

[9] J.-R. Chazottes and S. Gouëzel; *Optimal concentration inequalities for dynamical systems*, preprint, arXiv:1111.0849

[10] N. Chernov, *Decay of correlations and dispersing billiards* J. Statist. Phys. **94** (1999), no. 3–4, 513–556.

[11] N. Chernov and D. Dolgopyat; *Anomalous current in periodic Lorentz gases with infinite horizon*, Russian Mathematical Surveys, **64** (2009), 651–699.

[12] N. Chernov and R. Markarian, *Chaotic billiards*, Mathematical Surveys and Monographs **127**, American Mathematical Society, Providence, RI, 2006.

[13] N. Chernov and R. Markarian; *Dispersing billiards with cusps: slow decay of correlations* Comm. Math. Phys. **270** (2007), 727–758.

[14] N. Chernov and L.-S. Young; *Decay of correlations for Lorentz gases and hard balls* in: Hard Ball Systems and the Lorentz Gas, Ed. by D. Szasz, Encyclopaedia of Mathematical Sciences 101, pp. 89-120, Springer, 2000.

[15] N. Chernov, H.-K. Zhang, *Billiards with polynomial mixing rates*, Nonlinearity 18 (2005), no. 4, 1527–1553.

[16] R. Dorfmann; *A intorduction to chaos in non-equilibrium statistical mechanics*; Cambridge University Press, 1999.

[17] S. Gouëzel; *Central limit theorem and stable laws for intermittent maps*; Probability Theory and Related Fields, **128** (2004), 82–122.

[18] S. Gouëzel; Private communication.

[19] A. Katok and J.-M.Strelcyn; *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities*, Lecture Notes in Mathematics **1222**, Springer (1986).

[20] R. Klages, G. Radons and I.M. Sokolov (Editors); *Anomalous Transport: Foundations and Applications*, Wiley, 2008.

[21] C. Liverani, B. Saussol and S. Vaienti; *A probabilistic approach to intermittency*, Ergodic Theory Dyn. Syst. **19**(1999) 671–685.

[22] C. Liverani, M. P. Wojtkowski, *Ergodicity in Hamiltonian systems*, Dynamics Reported (1995), 130–202.

[23] Y. Pomeau and P. Manneville, *Intermittent transition to turbulence in dissipative dynamical systems*, Comm. Math. Phys. **74** (1980), 189–197.

[24] R. Markarian; *Billiards with polynomial decay of correlations*, Ergodic Theory Dyn. Syst. **24** (2004), 177–197.

[25] I. Melbourne; *Large and moderate deviations for slowly mixing dynamical systems*. Proc. Amer. Math. Soc. **137** (2009), 1735–1741.

[26] I. Melbourne; *Rapid decay of correlations for nonuniformly hyperbolic flows* Trans. Amer. Math. Soc. **359** (2007), 2421–2441.

[27] I. Melbourne, M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Ann. Probab. 37 (2009), no. 2, 478–505.

[28] I. Melbourne and D. Tershiu; *Decay of correlations for nonuniformly expanding systems with general return times*, preprint, available at
`http://personal.maths.surrey.ac.uk/st/I.Melbourne/papers/upper.pdf`

[29] I. Melbourne and A. Török; *Statistical limit theorems for suspension flows.* Israel Journal of Math. **144** (2004), 191–209.

[30] I. Melbourne and A. Török. *Convergence of moments for Axiom A and nonuniformly hyperbolic flows*. Ergodic Theory Dyn. Syst., to appear.

[31] M. Pollicott and R. Sharp; *Large deviations for intermittent maps*, Nonlinearity **22** (2009), 2079–2092.

[32] N. Simányi, *The characteristic exponents of the falling ball model*, Comm. Math. Phys. **182** (1996), no. 2, 457–468.

[33] D. Szász and T. Varjú, *Limit Laws and Recurrence for the Planar Lorentz Process with Infinite Horizon*, J. Statist. Phys. **129** (2007), 59–80.

[34] M. P. Wojtkowski, *A system of one dimensional balls with gravity*, Comm. Math. Phys. **126** (1990), 507–533.

[35] L.-S. Young, *Statistical properties of systems with some hyperbolicity including certain billiards*, Ann. Math., **147** (1998), 585–650.

[36] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188.