**STUDIA MATHEMATICA**

Online First version

# Certain systems of three falling balls satisfy the Chernov–Sinai ansatz

by

Michael Hofbauer-Tsiflakos (Wien)

**Abstract.** The system of falling balls is an autonomous Hamiltonian system with a smooth invariant measure and non-zero Lyapunov exponents almost everywhere. For almost three decades now, the question of its ergodicity remains open. We contribute to the solution of the ergodicity conjecture for three falling balls with a specific mass ratio in the following two points: First, we prove the Chernov–Sinai ansatz. Second, we prove that there is an abundance of sufficiently expanding points. It is of special interest that for the aforementioned specific mass ratio, the configuration space can be unfolded to a billiard table, where the proper alignment condition holds.

**1. Introduction.** The system of falling balls was introduced by Wojtkowski [W90a, W90b]. It describes the motion of $N \geq 2$ point masses, with positions $q_1, \dots, q_N$, momenta $p_1, \dots, p_N$ and masses $m_1, \dots, m_N$, moving up and down a vertical line and colliding elastically with each other. The bottom particle collides elastically with a rigid floor placed at $q_1 = 0$. The standing assumptions are $0 \leq q_1 \leq \dots \leq q_N$ and $m_1 \geq \dots \geq m_N$, $m_1 \neq m_N$. For convenience, we will refer to the point particles as *balls*. The system is an autonomous Hamiltonian system, with Hamiltonian given by the sum of the kinetic and linear potential energy of each ball. It possesses a smooth invariant measure with respect to the Hamiltonian flow and with respect to a suitable Poincaré map $T$, describing the movement of the balls from one collision to the next. We denote the underlying Poincaré section for this map by $\mathcal{M}^+$ and its invariant measure by $\mu$.

One aspect that makes the investigation of the dynamics cumbersome is the presence of singularities. These are codimension one manifolds in phase space on which the dynamics are not well-defined, in particular it has two

2010 *Mathematics Subject Classification*: Primary 37D50; Secondary 37J10.
*Key words and phrases*: ergodic theory, hyperbolic dynamical systems with singularities.
Received 19 May 2018; revised 1 September 2018 and 5 November 2018.
Published online 15 July 2019.

DOI: 10.4064/sm180519-2-10  

different images. A point belongs to the singularity manifold if its next collision is either between three balls or two balls with the floor.

Dynamicists first tried to answer the question whether the system of $N \geq 2$ falling balls has $2N-2$ non-zero Lyapunov exponents on a positive measure set of the phase space. The exceptional two directions with a zero exponent are the direction of the flow and the directions transversal to the energy surface. Wojtkowski was able to prove that two and three falling balls have non-zero Lyapunov exponents almost everywhere [W90a]. He supplemented this result by proving that an arbitrary number of balls subjected to a certain family of non-linear potential fields have non-zero Lyapunov exponents almost everywhere [W90b]. The most general result regarding the linear potential field is due to Simányi: For $N \geq 2$ falling balls, $\mu$-a.e. point $x \in \mathcal{M}^+$ has non-zero Lyapunov exponents [S96]. In [W98] Wojtkowski found an elegant way of proving the existence of non-zero Lyapunov exponents for a large class of falling balls systems. He first considers balls falling next to each other on a moving floor. By applying concrete stacking rules it is possible to obtain a variety of falling ball systems, including the original one introduced in [W90a]. The study of hyperbolicity is carried out by equivalently looking at the system of a particle falling in a wedge.

The motivation for this work is to contribute to the solution of the long-standing open problem of ergodicity for three or more balls. For two balls, the system is known to be ergodic [LW92, pp. 70–72] provided $m_1 > m_2$. Since the system of three falling balls has non-zero Lyapunov exponents everywhere, the theory of Katok–Strelcyn [KS86] shows that the phase space partitions into at most countably many components on which the conditional smooth measure is ergodic. A reliable method to check the ergodicity of such systems is the local ergodic theorem [ChS87, KSSz90, LW92]. In the present work we will follow the local ergodic theorem version of Liverani and Wojtkowski [LW92]. For its application, the local ergodic theorem needs the following five conditions to hold:

(1) Chernov–Sinai ansatz,
(2) non-contraction property,
(3) continuity of Lagrangian subspaces,
(4) regularity of singularity sets,
(5) proper alignment.

The validity of these conditions guarantees the existence of an open neighbourhood, around a point with non-vanishing Lyapunov exponents, that lies (mod 0) in one ergodic component. To prove that there is only one ergodic component requires a transitivity argument. Namely, the set of points with a sufficient amount of expansion must have full measure and be arcwise connected. We will refer to this property as the abundance of sufficiently expanding points. If the latter is true, one can build a chain of aforementioned open neighbourhoods from any point with sufficient expansion to another. These neighbourhoods have pairwise positive measure intersections, and hence there can only be one ergodic component. For three or more balls only condition (3) is known to be true [LW92].

In their approach to ergodicity, Liverani and Wojtkowski [LW92] introduced the property of (strict) unboundedness for a sequence of derivatives $(d_{T^n x}T)_{n\in\mathbb{N}}$. It roughly says that the expansion (measured with respect to an indefinite quadratic form $Q$) of any vector from the contracting cone field goes to infinity. In their terminology, it follows immediately that if $(d_{T^n x}T)_{n\in\mathbb{N}}$ is strictly unbounded everywhere then the Chernov–Sinai ansatz holds. Additionally, the abundance of sufficiently expanding points follows as a simple corollary.

The proof of the strict unboundedness property for every phase point is the main task of this work (see Section 2 for more details). For this, we will partially use the techniques introduced in [W98], which allow us to study the system of falling balls as a particle falling in a wedge. The results obtained from that analysis will be used to slightly modify the approach to strict unboundedness in [LW92].

Another important issue is the proper alignment condition (see Subsection 5.1.1). By some experts it has been wrongly assumed not to hold. We will thoroughly explain that it may well hold and is thus an open problem. Further, we will use the strict unboundedness property to analyze in Subsection 5.1.2 how the set of points that are not properly aligned behaves under sufficiently large iterates. We point out that for a specific mass ratio the configuration space of the falling balls systems can be unfolded to a billiard table (see Subsection 7.3) where the proper alignment condition holds as shown by Wojtkowski [W16].

On the same subject, in the realm of semi-dispersing billiards, Chernov [Ch93] formulated a transversality condition, which can serve as a substitute for proper alignment. We will show that in the framework of symplectic maps, Chernov's transversality follows from proper alignment (see Lemma 5.2).

The paper is organized in the following way:

In Section 2 we briefly summarize the main results of this paper, which are the strict unboundedness for every orbit, the Chernov–Sinai ansatz and the abundance of sufficiently expanding points. It will also be shown that the last two results follow at once from the strict unboundedness of every orbit.

In Section 3 we introduce the system of three falling balls.

In Section 4 we recall the standard method for studying Lyapunov exponents in Hamiltonian systems [W91] and recall what has been done for the system of falling balls so far.

Section 5, dealing with ergodicity, contains a detailed discussion of the local ergodic theorem, the proper alignment condition, Chernov's transversality condition and the abundance of sufficiently expanding points.

In Section 6 we give the first part of the proof of the strict unboundedness property. This section is written in the language of Liverani and Wojtkowski [LW92] and explains the use of our new results to modify their proof of unboundedness.

In Section 7 we introduce the system of a particle falling in a three-dimensional wedge from [W98]. For a special type of wedges this system is equivalent to the system of falling balls with particular masses. In the last subsection we show that the proper alignment condition is valid in those special wedges.

In Section 8 we utilize the results of Sections 6 and 7 to complete the proof of the strict unboundedness property.

**2. Main results.** Denote by $\mathcal{M}^+$ the phase space, which is partitioned $(\mathrm{mod}\,0)$ into subsets $\mathcal{M}_i^+$, $i = 1, 2, 3$, where each subset describes the moment right after collision of balls $i-1$ and $i$. For $i-1 = 0$, we have a collision with the floor, i.e. $q_1 = 0$. The masses satisfy $m_1 > m_2 > m_3$ and the special relation given in (7.9). Let $T : \mathcal{M}^+ \circlearrowleft$ be the Poincaré map, describing the movement from one collision to the next. After applying Wojtkowski's convenient coordinate transformation $(q, p) \mapsto (h, v) \mapsto (\xi, \eta)$ (see (4.1), (4.2)) we get a contracting cone field

$$\mathcal{C}(x) = \{(\delta\xi, \delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : Q(\delta\xi, \delta\eta) > 0,\ \delta\xi_1 = 0,\ \delta\eta_1 = 0\} \cup \{\vec{0}\},$$

where $(\delta\xi, \delta\eta)$ denote the coordinates in the tangent space. The cone field is defined by the quadratic form

$$Q(\delta\xi, \delta\eta) = \sum_{i=1}^{3} \delta\xi_i \delta\eta_i.$$

Denote by $\overline{\mathcal{C}(x)}$ the closure of $\mathcal{C}(x)$ and let $d_xT^n = d_{T^n x}T \dots d_{Tx}T d_xT$. The sequence of derivatives along the orbit

$$(d_{T^n x}T)_{n\in\mathbb{N}} = (d_xT, d_{Tx}T, d_{T^2x}T, \dots)$$

is called *unbounded* if

$$\lim_{n\to+\infty} Q(d_xT^n v) = +\infty, \qquad \forall v \in \mathcal{C}(x) \setminus \{\vec{0}\},$$

and *strictly unbounded* if

$$\lim_{n\to+\infty} Q(d_xT^n v) = +\infty, \qquad \forall v \in \overline{\mathcal{C}(x)} \setminus \{\vec{0}\}.$$

Main Theorem. *For every $x \in \mathcal{M}^+$, we have*

$$\lim_{n\to+\infty} Q(d_xT^n(\delta\xi,\delta\eta)) = +\infty \qquad \text{for all } (\delta\xi,\delta\eta) \in \overline{\mathcal{C}(x)} \setminus \{\vec{0}\}.$$

Due to [LW92, Proposition 6.2 and Theorem 6.8], the Main Theorem also implies the strict unboundedness for the orbit in negative time, $(d_{T^n x}T)_{n\in\mathbb{Z}^-}$ (with respect to the complementary cone field of $\mathcal{C}(x)$).

The singularity manifold on which $T$ resp. $T^{-1}$ is not well-defined is denoted by $\mathcal{S}^+$ resp. $\mathcal{S}^-$. Let $\mu_{\mathcal{S}^+}$ resp. $\mu_{\mathcal{S}^-}$ be the measures induced on the codimension one hypersurfaces $\mathcal{S}^+$ resp. $\mathcal{S}^-$ by the smooth $T$-invariant measure $\mu$.

The Main Theorem immediately yields the Chernov–Sinai ansatz, which is one of the conditions of the local ergodic theorem.

Chernov–Sinai ansatz. *For $\mu_{\mathcal{S}^-}$-a.e. $x \in \mathcal{S}^-$, we have*

$$\lim_{n\to+\infty} Q(d_xT^n(\delta\xi,\delta\eta)) = +\infty \qquad \text{for all } (\delta\xi,\delta\eta) \in \overline{\mathcal{C}(x)} \setminus \{\vec{0}\}.$$

The *least expansion coefficient* $\sigma$, for $n \geq 1$ and $x \in \mathcal{M}^+$, is defined as

$$\sigma(d_xT^n) = \inf_{v\in\mathcal{C}(x)} \sqrt{\frac{Q(d_xT^n v)}{Q(v)}}.$$

A point $x \in \mathcal{M}^+$ is called *sufficiently expanding* if there exists $n = n(x) \geq 1$ such that $\sigma(d_xT^n) > 3$.

The last result is the abundance of sufficiently expanding points. It can be described as a transitivity result, which specifies the size of the ergodicity domain in phase space by connecting open neighbourhoods which lie (mod 0) in one ergodic component.

Abundance of sufficiently expanding points. *The set of sufficiently expanding points has full measure and is arcwise connected.*

The abundance of sufficiently expanding points follows at once from the Main Theorem (see Theorem 4.4) and the proper alignment property (see Subsection 7.3). The validity of the latter guarantees that the set of double singular collisions is of codimension two.

**3. The system of three falling balls.** Let $q_i = q_i(t)$ be the position, $p_i = p_i(t)$ the momentum and $v_i = v_i(t)$ the velocity of the $i$th ball. The balls are aligned on top of each other and are therefore confined to

$$\mathcal{N}(q,p) = \{(q,p) \in \mathbb{R}^3 \times \mathbb{R}^3 : 0 \leq q_1 \leq q_2 \leq q_3\}.$$

The momenta and the velocities are related by $p_i = m_i v_i$. We assume that the masses $m_i$ satisfy $m_1 > m_2 > m_3$. The movements of the balls are a result of a linear potential field and their kinetic energies. The total energy of the system is given by the Hamiltonian function

$$H(q,p) = \sum_{i=1}^{3} \frac{p_i^2}{2m_i} + m_i q_i.$$

The Hamiltonian equations are

$$\dot{q}_i = p_i/m_i, \qquad \dot{p}_i = -m_i. \tag{3.1}$$

The dots indicate differentiation with respect to time $t$ and the Hamiltonian vector field on the right hand side will be denoted by $X_H(q,p)$. The solutions to these equations are

$$q_i(t) = -\frac{t^2}{2} + t\frac{p_i(0)}{m_i} + q_i(0), \qquad p_i(t) = -tm_i + p_i(0), \tag{3.2}$$

which form parabolas $(t, q_i(t)) \subset \mathbb{R} \times \mathbb{R}_+$. It is clear from the choice of the linear potential field that the acceleration of each ball points downwards, and thus these parabolas cannot escape to infinity. Hence, for every initial condition $(q,p)$ the balls go through every collision in finite time and thus every collision happens infinitely often. The energy manifold $E_c$ and its tangent space $\mathcal{T}E_c$ are given by

$$E_c = \left\{ (q,p) \in \mathbb{R}_+^3 \times \mathbb{R}^3 : H(q,p) = \sum_{i=1}^{3} \frac{p_i^2}{2m_i} + m_i q_i = c \right\},$$

$$\mathcal{T}_{(q,p)}E_c = \left\{ (\delta q, \delta p) \in \mathbb{R}^3 \times \mathbb{R}^3 : d_{(q,p)}H(\delta q, \delta p) = \sum_{i=1}^{3} \frac{p_i \delta p_i}{m_i} + m_i \delta q_i = 0 \right\}.$$

Including the restriction of the balls positions amounts to considering the set $E_c \cap \mathcal{N}(q,p)$.

The Hamiltonian vector field (3.1) gives rise to the Hamiltonian flow

$$\phi : \mathbb{R} \times E_c \cap \mathcal{N}(q,p) \to E_c \cap \mathcal{N}(q,p), \qquad (t,(q,p)) \mapsto \phi(t,(q,p)).$$

For convenience, the image $\phi(t,(q,p))$ will also be written as $\phi^t(q,p)$.

The standard symplectic form $\omega = \sum_{i=1}^{3} dq_i \wedge dp_i$ induces the symplectic volume element $\Omega = \bigwedge_{i=1}^{3} dq_i \wedge dp_i$. The volume element on the energy surface is obtained by contracting $\Omega$ with any vector $u$ satisfying $dH(u) = 1$. Denote the contraction operator by $\iota$; then the volume element on the energy surface is $\iota(u)\Omega$. Since the flow preserves the standard symplectic form, it preserves the volume element, and hence the Liouville measure $\nu$ on $E_c \cap \mathcal{N}(q,p)$ obtained from it. We define the Poincaré section, which describes the states right after a collision, as

$$\mathcal{M}^+ = \mathcal{M}_1^+ \cup \mathcal{M}_2^+ \cup \mathcal{M}_3^+$$

with

$$\begin{aligned}\mathcal{M}_1^+ &:= \{(q,p) \in E_c \cap \mathcal{N}(q,p) : q_1 = 0,\, p_1/m_1 \geq 0\},\\ \mathcal{M}_i^+ &:= \{(q,p) \in E_c \cap \mathcal{N}(q,p) : q_{i-1} = q_i,\, p_{i-1}/m_{i-1} \leq p_i/m_i\}, \qquad i = 2,3.\end{aligned}$$

In the same way we define the set of states right before collision,

$$\mathcal{M}^- = \mathcal{M}_1^- \cup \mathcal{M}_2^- \cup \mathcal{M}_3^-,$$

by

$$\begin{aligned}\mathcal{M}_1^- &:= \{(q,p) \in E_c \cap \mathcal{N}(q,p) : q_1 = 0,\, p_1/m_1 < 0\},\\ \mathcal{M}_i^- &:= \{(q,p) \in E_c \cap \mathcal{N}(q,p) : q_{i-1} = q_i,\, p_{i-1}/m_{i-1} > p_i/m_i\}, \qquad i = 2,3.\end{aligned}$$

The $+$ resp. $-$ superscripts refer to the states right after resp. before collision. The system of falling balls is considered as a hard ball system with fully elastic collisions. During a collision of balls $i$ and $i+1$ the momenta and velocities change according to

$$\begin{aligned}p_i^+ &= \gamma_i p_i^- + (1+\gamma_i) p_{i+1}^-, & v_i^+ &= \gamma_i v_i^- + (1-\gamma_i) v_{i+1}^-,\\ p_{i+1}^+ &= (1-\gamma_i) p_i^- - \gamma_i p_{i+1}^-, & v_{i+1}^+ &= (1+\gamma_i) v_i^- - \gamma_i v_{i+1}^-,\end{aligned} \tag{3.3}$$

where $\gamma_i = (m_i - m_{i+1})/(m_i + m_{i+1})$, $i = 1,2$, and when the bottom particle collides with the floor the sign of its momentum resp. velocity is simply reversed:

$$p_1^+ = -p_1^-, \qquad v_1^+ = -v_1^-. \tag{3.4}$$

These collision laws are described by the linear collision map

$$\Phi_{i-1,i} : \mathcal{M}^- \to \mathcal{M}^+, \qquad (q,p^-) \mapsto (q,p^+).$$

We will write $\Phi$ if we do not want to refer to any specific collision.

Let $\tau : M \to \mathbb{R}_+$ be the first return time to $\mathcal{M}^-$. We define the *Poincaré map* as

$$T : \mathcal{M}^+ \to \mathcal{M}^+, \qquad (q,p) \mapsto \Phi \circ \phi^{\tau(q,p)}(q,p).$$

Thus $T$ is the collision map, which maps from one collision to the next. On $\mathcal{M}^+$, we obtain the volume element $\iota(X_H)\iota(u)\Omega$, by contracting the volume element $\iota(u)\Omega$ on the energy surface with the direction of the flow $X_H$. This exterior form defines a smooth measure $\mu$ on $\mathcal{M}^+$, which is $T$-invariant. Our dynamical system can be described as the triple $(\mathcal{M}^+, T, \mu)$. Matching the present state with the next collision in the future ('+') resp. the past ('−'), we obtain two (mod 0) partitions of $\mathcal{M}^+$ with elements

$$\mathcal{M}^{\pm}_{1,1} = \{x \in \mathcal{M}^+_1 : T^{\pm 1}x \in \mathcal{M}^+_1\},$$
$$\mathcal{M}^{\pm}_{i,j} = \{x \in \mathcal{M}^+_i : T^{\pm 1}x \in \mathcal{M}^+_j\}, \quad i,j \in \{1,2,3\},\ j \neq i.$$

It can be calculated that $\mu(\mathcal{M}^{\pm}_{i,j}) > 0$. The system of falling balls has codimension one singularity manifolds

$$\mathcal{S}^+_{1,2} = \mathcal{M}^+_{1,2} \cap \mathcal{M}^+_{1,3}, \quad \mathcal{S}^-_{1,2} = \mathcal{M}^-_{2,1} \cap \mathcal{M}^-_{3,1},$$
$$\mathcal{S}^+_{1,1} = \mathcal{M}^+_{1,1} \cap \mathcal{M}^+_{1,2}, \quad \mathcal{S}^-_{1,1} = \mathcal{M}^-_{1,1} \cap \mathcal{M}^-_{2,1},$$
$$\mathcal{S}^+_{3,1} = \mathcal{M}^+_{3,1} \cap \mathcal{M}^+_{3,2}, \quad \mathcal{S}^-_{3,1} = \mathcal{M}^-_{1,3} \cap \mathcal{M}^-_{2,3}.$$

The states in $\mathcal{S}^{\pm}_{1,2}$ face a triple collision next, while the states in $\mathcal{S}^{\pm}_{1,1}, \mathcal{S}^{\pm}_{3,1}$ experience a collision of the lower two balls with the floor next. The maps $T$ resp. $T^{-1}$ are not well-defined on the sets $\mathcal{S}^+_{1,1}, \mathcal{S}^+_{1,2}, \mathcal{S}^+_{3,1}$ resp. $\mathcal{S}^-_{1,1}, \mathcal{S}^-_{1,2}, \mathcal{S}^-_{3,1}$, because they have two different images. This happens because the compositions $\Phi_{0,1} \circ \Phi_{1,2}$ and $\Phi_{1,2} \circ \Phi_{2,3}$ do not commute. When the trajectory hits a singularity, we will continue the system on both branches separately. In this way, the results obtained in this work hold for every point.

We abbreviate

$$\mathcal{S}^{\pm} = \mathcal{S}^{\pm}_{1,1} \cup \mathcal{S}^{\pm}_{1,2} \cup \mathcal{S}^{\pm}_{3,1}, \quad \mathcal{S}^{\pm}_n = \mathcal{S}^{\pm} \cup T^{\mp 1}\mathcal{S}^{\pm} \cup \cdots \cup T^{\mp(n-1)}\mathcal{S}^{\pm}.$$

**4. Lyapunov exponents.** We subject our system to two well-discussed coordinate transformations $(q,p) \mapsto (h,v) \mapsto (\xi,\eta)$ introduced in [W90a]. The first one is given by

$$h_i = \frac{p_i^2}{2m_i} + m_i q_i, \quad v_i = \frac{p_i}{m_i}, \tag{4.1}$$

while the second one is a linear coordinate transformation

$$\xi_i = A^{-1}h_i, \quad \eta_i = A^T v_i, \tag{4.2}$$

where $A$ is an invertible matrix depending only on the masses $m_i$ [W90a, p. 520]. The energy manifold and its tangent space take the form

$$E_c = \{(\xi,\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : H(\xi,\eta) = \xi_1 = c\},$$
$$\mathcal{T}E_c = \{(\delta\xi,\delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : dH(\delta\xi,\delta\eta) = \delta\xi_1 = 0\}.$$

The Hamiltonian vector field $X_H(\xi,\eta) = (0,0,0,-1,0,0)$ becomes constant. In these coordinates, the derivative $d\phi^t$ of the flow equals the identity map. Thus, only the derivatives $d\Phi_{i-1,i}$ of the collision maps are relevant to the dynamics in the tangent space. In these coordinates the derivatives of the collision maps are given by

$$d\Phi_{0,1} = \begin{pmatrix} \mathrm{id}_3 & 0 \\ B & \mathrm{id}_3 \end{pmatrix}, \quad d\Phi_{1,2} = \begin{pmatrix} M_1 & U_1 \\ 0 & M_1^T \end{pmatrix}, \quad d\Phi_{2,3} = \begin{pmatrix} M_2 & U_2 \\ 0 & M_2^T \end{pmatrix},$$

where

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad U_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\alpha_1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\alpha_2 \end{pmatrix},$$

$$\mathrm{id}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1+\gamma_1 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1-\gamma_2 & -1 \end{pmatrix}.$$

The entries in the matrices are given by

$$\beta = -\frac{2}{m_1 v_1^-}, \qquad \alpha_i = \frac{2 m_i m_{i+1}(m_i - m_{i+1})(v_i^- - v_{i+1}^-)}{(m_i + m_{i+1})^2}. \tag{4.3}$$

A *Lagrangian subspace* is a linear space of maximal dimension on which the symplectic form vanishes. In general, every vector $v \in \mathbb{R}^6$ can be uniquely decomposed according to any pair of transversal Lagrangrian subspaces $(V_1, V_2)$, i.e. $v = v_1 + v_2$, $v_i \in V_i$, $i = 1, 2$. For such a pair $(V_1, V_2)$ we can define a quadratic form $Q$ by

$$Q : \mathbb{R}^6 \to \mathbb{R}, \qquad v \mapsto Q(v) = \omega(v_1, v_2).$$

The canonical pair of transversal Lagrangian subspaces in $\mathbb{R}^6$ is given by

$$W_1 = \{(\delta\xi, \delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : \delta\eta_1 = \delta\eta_2 = \delta\eta_3 = 0\},$$
$$W_2 = \{(\delta\xi, \delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : \delta\xi_1 = \delta\xi_2 = \delta\xi_3 = 0\}.$$

Restricting both to $\mathcal{T}E$ and excluding the direction of the flow gives

$$\begin{aligned} L_1 &= \{(\delta\xi, \delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : \delta\xi_1 = 0,\ \delta\eta_i = 0,\ i = 1, 2, 3\}, \\ L_2 &= \{(\delta\xi, \delta\eta) \in \mathbb{R}^3 \times \mathbb{R}^3 : \delta\eta_1 = 0,\ \delta\xi_i = 0,\ i = 1, 2, 3\}. \end{aligned} \tag{4.4}$$

For the pair $(L_1, L_2)$, the quadratic form $Q$ becomes the Euclidean inner product

$$Q(\delta\xi, \delta\eta) = \langle \delta\xi, \delta\eta \rangle.$$

We see immediately that $Q(L_i) = 0$. Also, $Q$ is continuous and homogeneous of degree two. Using the quadratic form $Q$ we can define the open cones

$$\mathcal{C}(x) = \{(\delta\xi, \delta\eta) \in L_1 \oplus L_2 : Q(\delta\xi, \delta\eta) > 0\} \cup \{\vec{0}\},$$
$$\mathcal{C}'(x) = \{(\delta\xi, \delta\eta) \in L_1 \oplus L_2 : Q(\delta\xi, \delta\eta) < 0\} \cup \{\vec{0}\}.$$

Denote by $\overline{\mathcal{C}(x)}$ the closure of $\mathcal{C}(x)$.

Definition 4.1.

(1) The cone field $\{\mathcal{C}(x) : x \in \mathcal{M}^+\}$ is called *invariant* at $x \in \mathcal{M}^+$ if

$$d_x T \overline{\mathcal{C}(x)} \subseteq \overline{\mathcal{C}(Tx)}.$$

(2) The cone field $\{\mathcal{C}(x) : x \in \mathcal{M}^+\}$ is called *eventually strictly invariant* at $x \in \mathcal{M}^+$ if there exists a $k \geq 1$ such that

$$d_x T^k \overline{\mathcal{C}(x)} \subset \mathcal{C}(T^k x).$$

(3) Let $x \in \mathcal{M}^+$. The monodromy map $d_x T$ is called $Q$-monotone if

$$Q(d_x T(\delta\xi, \delta\eta)) \geq Q(\delta\xi, \delta\eta) \quad \text{for all } (\delta\xi, \delta\eta) \in L_1 \oplus L_2.$$

(4) Let $x \in \mathcal{M}^+$. The monodromy map $d_x T$ is called *eventually strictly $Q$-monotone* if there exists a $k \geq 1$ such that

$$Q(d_x T^k(\delta\xi, \delta\eta)) > Q(\delta\xi, \delta\eta) \quad \text{for all } (\delta\xi, \delta\eta) \in L_1 \oplus L_2 \setminus \{\vec{0}\}.$$

The condition in (1) resp. (2) is equivalent to that in (3) resp. (4) (see e.g. [LW92, Theorem 4.1]). The following lemma establishes eventual strict $Q$-monotonicity by using only the evolution of the Lagrangian subspaces $L_1$ and $L_2$ (see e.g. [W90a, Lemma 2]).

LEMMA 4.2. *Let $x \in \mathcal{M}^+$. The monodromy map $d_x T$ is eventually strictly $Q$-monotone if there exists $k \geq 1$ such that for all $(\delta\xi, 0) \in L_1$ and $(0, \delta\eta) \in L_2$,*

$$Q(d_x T^k(\delta\xi, 0)) > 0 \quad \text{and} \quad Q(d_x T^k(0, \delta\eta)) > 0.$$

In order to get non-zero Lyapunov exponents Wojtkowski [W90a, p. 516] introduced a criterion which links eventual strict $Q$-monotonicity to nonuniform hyperbolic behaviour:

$Q$-CRITERION. *If $d_x T$ is eventually strictly $Q$-monotone for $\mu$-a.e. $x \in \mathcal{M}^+$, then all Lyapunov exponents, except for two*(1)*, are non-zero.*

For $N \geq 2$ balls, Wojtkowski proved [W90a] that $d_x T$ is $Q$-monotone for every $x$ in $\mathcal{M}^+$. He strengthened this statement in the case of three balls with upward decreasing masses, by proving eventual strict $Q$-monotonicity for every (2) point in $\mathcal{M}^+$ [W90a, Proposition 3]. Afterwards Simányi [S96] proved that $d_x T$ is eventually strictly $Q$-monotone for $\mu$-a.e. $x \in \mathcal{M}^+$ and an arbitrary number of balls.

We close this subsection by defining the (strict) unboundedness property and the least expansion coefficient, which will be used to establish criteria for ergodicity.

The *least expansion coefficient* $\sigma$, for $n \geq 1$ and $x \in \mathcal{M}^+$, is defined as

$$\sigma(d_x T^n) = \inf_{v \in \mathcal{C}(x)} \sqrt{\frac{Q(d_x T^n v)}{Q(v)}}. \tag{4.5}$$

(1) The exceptional directions with zero Lyapunov exponents are the direction of the flow and the ones contained in the subset $\{v : dH(v) \neq 0\}$.

(2) Even though Proposition 3 in [W90a] is stated for almost every $x \in \mathcal{M}^+$, an inspection of the proof shows that it actually holds for every $x \in \mathcal{M}^+$.

Definition 4.3. The sequence $(d_{T^n x}T)_{n\in\mathbb{N}}$ is called *unbounded* if

$$\lim_{n\to+\infty} Q(d_xT^nv) = +\infty, \qquad \forall v \in \mathcal{C}(x)\setminus\{\vec{0}\},$$

and *strictly unbounded* if

$$\lim_{n\to+\infty} Q(d_xT^nv) = +\infty, \qquad \forall v \in \overline{\mathcal{C}(x)}\setminus\{\vec{0}\}.$$

The least expansion coefficient and the property of strict unboundedness relate to each other in the following way.

Theorem 4.4 ([LW92, Theorem 6.8]). *The sequence $(d_{T^n x}T)_{n\in\mathbb{N}}$ is strictly unbounded if and only if $\lim_{n\to+\infty}\sigma(d_xT^n) = +\infty$.*

**5. Ergodicity.** The theory of Katok–Strelcyn [KS86] implies, that since our system has non-zero Lyapunov exponents almost everywhere, we can partition the phase space $\mathcal{M}^+$ into countably many components on which the conditional smooth measure is ergodic. To prove that there is only one ergodic component, the following two points need to be verified:

(1) local ergodicity,
(2) abundance of sufficiently expanding points.

**5.1. Local ergodicity.** We start with

Definition 5.1. A compact subset $X \subset \mathcal{M}^+$ is called *regular* if

(1) $X = \bigcup_{i=1}^n I_i$, where $I_i$ are compact submanifolds with $I_i = \overline{\operatorname{int} I_i}$,
(2) $\dim I_i = 3$,
(3) $I_i \cap I_j \subset \partial I_i \cup \partial I_j$ for $i \neq j$,
(4) $\partial I_i = \bigcup_{j=1}^m H_{i,j}$, where $\dim H_{i,j} = 2$ and $H_{i,j}$ is compact.

Local ergodicity amounts to showing that around a point with least expansion coefficient larger than three, it is possible to find an open neighbourhood which lies (mod 0) in one ergodic component. To see this, one needs to check the following five conditions.

Condition 1 (Regularity of singularity sets). *The singularity sets $\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ are both regular sets for every $n \geq 1$.*

Condition 2 (Non-contraction property). *There exists $\zeta > 0$ such that for all $n \geq 1$, all $x \in \mathcal{M}^+ \setminus \mathcal{S}_n^+$, and all $(\delta\xi,\delta\eta) \in \overline{\mathcal{C}(x)}$, we have*

$$\|d_xT^n(\delta\xi,\delta\eta)\| \geq \zeta\|(\delta\xi,\delta\eta)\|.$$

Condition 3 (Chernov–Sinai ansatz). *For $\mu_{\mathcal{S}^-}$-a.e. $x \in \mathcal{S}^-$, we have*

$$\lim_{n\to+\infty} Q(d_xT^n(\delta\xi,\delta\eta)) = +\infty \quad \textit{for all } (\delta\xi,\delta\eta) \in \overline{\mathcal{C}(x)}.$$

Condition 4 (Continuity of Lagrangian subspaces). *The ordered pair of transversal Lagrangian subspaces $(L_1(x),\, L_2(x))$ varies continuously in $\operatorname{int}\mathcal{M}^+$.*

Condition 5 (Proper alignment). *There exists $N \geq 0$ such that for every $x$ in $\mathcal{S}^+$ resp. $\mathcal{S}^-$, $d_x T^{-N} v_x^+$ resp. $d_x T^N v_x^-$ belongs to $\overline{\mathcal{C}'(T^{-N}x)}$ resp. $\overline{\mathcal{C}(T^N x)}$, where $v_x^+$ resp. $v_x^-$ are on the characteristic lines* (3) *of $\mathcal{T}_x\mathcal{S}^+$ resp. $\mathcal{T}_x\mathcal{S}^-$.*

At the moment, for three or more falling balls, only Condition 4 has been verified. This is in fact easy to see, because the canonical pair of transversal Lagrangian subspaces (4.4) does not depend on the base point $x$ and is therefore constant in $\mathcal{M}^+$. Note that Conditions 2 and 3 also have to hold in negative time.

Local Ergodic Theorem. *If Conditions* 1–5 *are satisfied, then for any $x \in \mathcal{M}^+$ and $n \geq 1$ such that $\sigma(d_x T^n) > 3$ there exists an open ergodic neighbourhood $\mathcal{U}(x)$ that lies* (*mod* 0) *in one ergodic component.*

Chernov [Ch93] postulated a condition weaker than Condition 5. Denote by $W^u(x)$ resp. $W^s(x)$ the unstable resp. stable manifold at $x$.

Condition 6 (Transversality). *For $\mu_{\mathcal{S}^\pm}$-a.e. $x$, the subspace $W^s(x)$ resp. $W^u(x)$ is transversal to $\mathcal{S}^-$ resp. $\mathcal{S}^+$.*

Lemma 5.2. *The proper alignment condition implies the transversality condition.*

*Proof.* Assume that at $x \in \mathcal{S}^-$, the singularity manifold and $W^s(x)$ are not transversal but still properly aligned, i.e. $\mathcal{T}W^s(x) \subset \mathcal{T}\mathcal{S}^-$ and $\{v_x^-\} \cap \mathcal{T}W^s(x) = \emptyset$. Since transversality is not satisfied and $v_x^-$ is on the characteristic line, we have $\omega(v_x^-, v) = 0$ for all $v \in \mathcal{T}W^s(x)$. This means, that $v_x^- \in (\mathcal{T}W^s(x))^\perp_\omega$, where $(\mathcal{T}W^s(x))^\perp_\omega$ is the $\omega$-orthogonal complement of $\mathcal{T}W^s(x)$. But $\mathcal{T}W^s(x)$ is a Lagrangian subspace, and thus, $(\mathcal{T}W^s(x))^\perp_\omega = \mathcal{T}W^s(x)$. Hence, $v_x^- \in \mathcal{T}W^s(x)$, a contradiction. ∎

Even though proper alignment implies transversality, it is presently unclear whether for the local ergodic theorem (in the Liverani–Wojtkowski framework) to hold it is enough to have proper alignment only on a set of full measure with respect to the measure $\mu_{\mathcal{S}^\pm}$.

**5.1.1.** *The current state of proper alignment.* There has been a substantial misconception about whether the system of falling balls is properly aligned or not. In brief, the correct answer is that on some part of the singularity manifold the system is properly aligned, while on the rest we simply do not know. The latter regards only the singularity manifolds $\mathcal{S}^\pm_{1,2}$, since every point on $\mathcal{S}^\pm_{1,1}$ and $\mathcal{S}^\pm_{3,1}$ is properly aligned. The original formulation of the proper alignment condition in [LW92] is more restrictive than the one

(3) The characteristic line is the set of vectors $\mathcal{T}_x\mathcal{S}^\pm$ such that $\omega(v_x^\pm, w) = 0$ for all $w \in \mathcal{T}_x\mathcal{S}^\pm$. In other words, it is the $\omega$-orthogonal complement of $\mathcal{T}_x\mathcal{S}^\pm$. Note that in symplectic geometry the $\omega$-orthogonal complement of a codimension one subspace is one-dimensional.

stated above. Namely, it demands the characteristic line $\{v_x^-\}$ resp. $\{v_x^+\}$ to lie in $\mathcal{C}(x)$ resp. $\mathcal{C}'(x)$ for every point of the singularity manifold. After the statement of the original proper alignment condition it says [LW92, p. 165]:

> It will be clear from the way in which the proper alignment of singularity sets is used in Section 12 that it is sufficient to assume that there is $N$ such that $T^N\mathcal{S}^-$ and $T^{-N}\mathcal{S}^+$ are properly aligned.

In [LW92, Section 12] the authors remind the reader that, in their constructive argument, the size of the neighbourhood $\mathcal{U}(x)$, appearing in the local ergodic theorem, was chosen so small that $\mathcal{U}(x)\cap\mathcal{S}_N^- = \emptyset$. Due to the regularity of singularity manifolds (see Condition 1), for every $M > N$ there exists a finite $p = p(M) > 0$ such that $\bigcup_{i=N}^M T^i\mathcal{S}^- = \bigcup_{k=1}^p I_k$, where the $I_k$ are compact submanifolds (see Definition 5.1). In the proof of their Proposition 12.2, Liverani and Wojtkowski make use of the fact that **every** point $x \in I_k$ is properly aligned (see [LW92, p. 185]). Hence, the relaxed version of the proper alignment condition (see Condition 5) is justified.

The authors continue (see [LW92, p. 165]) with the following assertion:

> We will show, in section 14, that for the system of falling balls even this weaker property [see Condition 5] fails.

This assertion is, however, erroneous. We will now illustrate what Liverani and Wojtkowski really did in Section 14: The argument is carried out for the singularity manifold $\mathcal{S}_{1,2}^-$. The characteristic line at $x \in \mathcal{S}_{1,2}^-$ is given by

$$\{v_x^-\} = \Bigg\{(\delta q,\delta p)\in\mathcal{T}_x\mathcal{S}_{1,2}^- : \delta q_1 = \delta q_2 = \delta q_3 = 0,$$
$$\sum_{i=1}^3 \delta p_i = 0, \sum_{i=1}^3 \frac{p_i\delta p_i}{m_i} = 0,\ \frac{p_1}{m_1} \le \frac{p_2}{m_2} \le \frac{p_3}{m_3}\Bigg\}.$$

The restrictions on the momenta follow from $\mathcal{S}_{1,2}^- \subset \mathcal{M}_2^+\cap\mathcal{M}_3^+$. We will look at the set of momenta in a little more detail: Without loss of generality we assume $t_0 < t_1$, $x = x(t_0) \in \mathcal{S}_{1,2}^-$ and $Tx = x(t_1) \in \mathcal{M}_1^+$. Since $p_1^+(t_0)/m_1 \le p_2^+(t_0)/m_2 \le p_3^+(t_0)/m_3$, applying the equations of motion (3.2) yields $p_1^-(t_1)/m_1 \le p_2^-(t_1)/m_2 \le p_3^-(t_1)/m_3$. As $x(t_1)\in\mathcal{M}_1^+$, we have $p_1^-(t_1)/m_1 < 0$. Consequently, we (mod 0) partition the set of eligible momenta at time $t_1$ into the subsets

$$\mathsf{Mom}_1(q(t_1),p^-(t_1)) = \left\{\frac{p_1^-(t_1)}{m_1} < 0 \le \frac{p_2^-(t_1)}{m_2} \le \frac{p_3^-(t_1)}{m_3}\right\},$$
$$\mathsf{Mom}_2(q(t_1),p^-(t_1)) = \left\{\frac{p_1^-(t_1)}{m_1} < \frac{p_2^-(t_1)}{m_2} \le 0 \le \frac{p_3^-(t_1)}{m_3}\right\},$$
$$\mathsf{Mom}_3(q(t_1),p^-(t_1)) = \left\{\frac{p_1^-(t_1)}{m_1} < \frac{p_2^-(t_1)}{m_2} \le \frac{p_3^-(t_1)}{m_3} \le 0\right\}.$$

Using again the equations of motion, we obtain, at time $t_0$,

$$\mathsf{Mom}_1(q(t_0), p^+(t_0)) = \left\{ \frac{p_1^+(t_0)}{m_1} < t_1 - t_0 \le \frac{p_2^+(t_0)}{m_2} \le \frac{p_3^+(t_0)}{m_3} \right\},$$
$$\mathsf{Mom}_2(q(t_0), p^+(t_0)) = \left\{ \frac{p_1^+(t_0)}{m_1} < \frac{p_2^+(t_0)}{m_2} \le t_1 - t_0 \le \frac{p_3^+(t_0)}{m_3} \right\},$$
$$\mathsf{Mom}_3(q(t_0), p^+(t_0)) = \left\{ \frac{p_1^+(t_0)}{m_1} < \frac{p_2^+(t_0)}{m_2} \le \frac{p_3^+(t_0)}{m_3} \le t_1 - t_0 \right\}.$$

Observe that the momenta can all be simultaneously negative only on the set $\mathsf{Mom}_3(q(t_0), p^+(t_0))$.

The quadratic form $Q$ of the contracting cone field in coordinates $(q, p)$ equals

$$Q(\delta q, \delta p) = \sum_{i=1}^{3} \delta q_i \delta p_i + \frac{p_i (\delta p_i)^2}{m_i^2}.$$

Inserting $v_x^-$ into $Q$ results in

$$Q(v_x^-) = \sum_{i=1}^{3} \frac{p_i (\delta p_i)^2}{m_i^2}. \tag{5.1}$$

The singularity manifold $\mathcal{S}_{1,2}^-$ at $x$ is properly aligned if and only if $Q(v_x^-) \ge 0$. It is easy to see that each of the sets $\mathsf{Mom}_i(q(t_0), p^+(t_0))$ contains a subset on which $\mathcal{S}_{1,2}^-$ is not properly aligned, i.e. $Q(v_x^-) < 0$. Hence, depending on $x \in \mathcal{S}_{1,2}^-$, (5.1) can take non-negative and negative values on every set $\mathsf{Mom}_i(q(t_0), p^+(t_0))$.

Additionally note that the image of the characteristic line is the characteristic line of the image:

$$d_x T^n v_x^- = v_{T^n x}^-. \tag{5.2}$$

Combining this with the fact that $d_x T$ is $Q$-monotone at every $x \in \mathcal{M}^+$ (see Definition 4.1(3)) we find that once a point is properly aligned, it remains properly aligned.

We summarize that on some parts of $\mathcal{S}_{1,2}^-$ the system of falling balls is properly aligned and on the complement we do not know, since an iterate of the characteristic line could very well be mapped into the contracting cone field. This is exactly what Liverani and Wojtkowski prove in Section 14. More importantly, they do **not** examine whether any iterate of $v_x^-$ gets mapped into the contracting cone field or not. This is currently unknown.

**5.1.2.** *Iterates of the characteristic line.* The Main Theorem allows us to compare the set of those iterated singular points which are not properly aligned to those points of the iterated singularity manifold which are not properly aligned. For this, an immediate consequence of the Main Theorem is that the monodromy matrix $d_xT$ is eventually strictly $Q$-monotone for every $x \in \mathcal{M}^+$ (see e.g. (6.1b) in Theorem 6.1), i.e. for every $x \in \mathcal{M}^+$, there exists $k = k(x) \geq 1$ such that $Q(d_xT^kv) > Q(v)$ for all $v \in L_1 \oplus L_2$. Define, for $n \geq 1$,

$$A(n, \mathcal{S}_{1,2}^-) = \{x \in \mathcal{S}_{1,2}^- : Q(v_x^-) < 0,\, Q(d_xT^nv) > Q(v),\, \forall v \in L_1 \oplus L_2\},$$

$$\bigcup_{n\geq 1} A(n, \mathcal{S}_{1,2}^-) = A(\mathcal{S}_{1,2}^-).$$

The set $A(n, \mathcal{S}_{1,2}^-)$ consists of all points in $\mathcal{S}_{1,2}^-$ which are not properly aligned and have an eventually strictly $Q$-monotone monodromy matrix after $n$ steps. We remark that $A(n, \mathcal{S}_{1,2}^-)$ is empty for small $n$. Once $A(n, \mathcal{S}_{1,2}^-) \neq \emptyset$, the $Q$-monotonicity of $d_xT$ for every $x \in \mathcal{M}^+$ implies that $A(n, \mathcal{S}_{1,2}^-) \subseteq A(n+1, \mathcal{S}_{1,2}^-)$. Due to the eventual strict $Q$-monotonicity of $d_xT$, we have

$$Q(d_{T^nx}T^{-n}v_{T^nx}^-) < Q(v_{T^nx}^-), \qquad \forall T^nx \in T^nA(n, \mathcal{S}_{1,2}^-).$$

Using this together with (5.2), we obtain

$$T^{-n}A(T^n\mathcal{S}_{1,2}^-) \subset A(n, \mathcal{S}_{1,2}^-) \subset A(\mathcal{S}_{1,2}^-).$$

However, the size of $T^{-n}A(T^n\mathcal{S}_{1,2}^-)$ and whether there exists a fixed $N \geq 1$, such that $A(T^N\mathcal{S}_{1,2}^-) = \emptyset$, remains unknown.

**5.2. Abundance of sufficiently expanding points.** Liverani and Wojtkowski require the point in the local ergodic theorem to have least expansion coefficient larger than three. However, after their formulation of the local ergodic theorem they point out (see [LW92, p. 167]) that there is no loss in generality in actually demanding that the least expansion coefficient be only larger than one. The reason is that the set of points with non-zero Lyapunov exponents has full measure (see [S96], [W98]). We quote:

> Let us note that the conditions of the last theorem are satisfied for almost all points $p \in \mathcal{M}$. Indeed, let
>
> $$\mathcal{M}_{n,\epsilon} = \{p \in \mathcal{M} \mid \sigma(D_pT^n) > 1 + \epsilon\}.$$
>
> Since almost all points are strictly monotone, then
>
> $$\bigcup_{n=1}^{+\infty} \bigcup_{\epsilon>0} \mathcal{M}_{n,\varepsilon}$$
>
> has full measure. By the Poincaré Recurrence Theorem and the supermultiplicativity of the coefficient $\sigma$, we conclude that
>
> $$\bigcup_{n=1}^{+\infty} \mathcal{M}_{n,3}$$
>
> has also full measure.

DEFINITION 5.3. Under the assumption $\mu\big(\{x \in \mathcal{M}^+ : \exists n = n(x) \geq 1, \sigma(d_xT^n) > 1\}\big) = 1$, a point $x \in \mathcal{M}^+$ is called *sufficiently expanding* if there exists an $n \geq 1$ such that $\sigma(d_xT^n) > 1$.

Once local ergodicity is established we know that every ergodic component is (mod 0) open. To obtain a single ergodic component one needs to verify

THEOREM 5.4 (Abundance of sufficiently expanding points). *The set of sufficiently expanding points has full measure and is arcwise connected.*

More precisely, this implies, that one can connect any two sufficiently expanding points by a curve which lies completely in the set of sufficiently expanding points. Consequently, the points on the curve can be chosen in such a way that the open neighbourhoods from the local ergodic theorem have pairwise intersections of positive measure. Hence, there can only be one ergodic component. For a more detailed proof see e.g. [ChM06, pp. 151–152].

**6. Strict unboundedness: part I.** In this section we begin the proof of the strict unboundedness of the sequence $(d_{T^nx}T)_{n\in\mathbb{N}}$ for every $x \in \mathcal{M}^+$. Due to [LW92, Theorem 6.8] we have the following equivalence:

THEOREM 6.1. *For every $x \in \mathcal{M}^+$, the sequence $(d_{T^nx}T)_{n\in\mathbb{N}}$ is strictly unbounded if and only if*

(6.1a) *For every $x \in \mathcal{M}^+$, the sequence $(d_{T^nx}T)_{n\in\mathbb{N}}$ is unbounded.*
(6.1b) *For every $x \in \mathcal{M}^+$, there exist $k_1, k_2 \in \mathbb{N}$ such that we have $Q(d_xT^{k_1}(\delta\xi, 0)) > 0$ and $Q(d_xT^{k_2}(0, \delta\eta)) > 0$ for all $(\delta\xi, 0) \in L_1$ and $(0, \delta\eta) \in L_2$.*

We will prove the strict unboundedness by equivalently proving properties (6.1a) and (6.1b). Let $\|\cdot\|$ denote the Euclidean norm.

The most important ingredient for (6.1a) is the following

THEOREM 6.2. *There exists a positive constant $\Lambda > 0$ such that for each $x \in \mathcal{M}^+$ there exists a sequence $(n_k)_{k\in\mathbb{N}} = (n_k(x))_{k\in\mathbb{N}}$ of strictly increasing positive integers such that for all $(0, \delta\eta) \in L_2$,*

$$Q(d_{T^{n_{2k-2}}x}T^{n_{2k-1}-n_{2k-2}}(0, \delta\eta)) > \Lambda\|(0, \delta\eta)\|^2. \tag{6.2}$$

In fact, we will prove that $dT^{n_{2k-1}-n_{2k-2}}$ equals either $d\Phi_{2,3}d\Phi_{1,2}$, $d\Phi_{1,2}d\Phi_{2,3}$, $d\Phi_{2,3}d\Phi_{0,1}d\Phi_{1,2}$ or $d\Phi_{1,2}d\Phi_{0,1}d\Phi_{2,3}$ [4]. Recursively define

$$(\delta\xi_n, \delta\eta_n) = \begin{cases} dT(\delta\xi_{n-1}, \delta\eta_{n-1}) & \text{for } n = 1, 2, \ldots, \\ (\delta\xi, \delta\eta) & \text{for } n = 0, \end{cases} \quad \text{and} \quad q_n = Q(\delta\xi_n, \delta\eta_n).$$

[4] The results remain valid if we allow multiple collisions with the floor, i.e. $d\Phi_{0,1}^{k_1}$ for every $k_1 \geq 1$.

From [W90a] we know that $d_xT$ is $Q$-monotone for every $x \in \mathcal{M}^+$, therefore $q_{n+1} \geq q_n$. Hence, in order to prove $\lim_{n\to+\infty} q_n = +\infty$, it is enough to prove this divergence along a subsequence $(q_{n_{2k-1}})_{k\in\mathbb{N}}$. We define

$$q_{n_{2k-1}} = Q(d_{T^{n_{2k-2}}x}T^{n_{2k-1}-n_{2k-2}}(\delta\xi_{n_{2k-2}}, \delta\eta_{n_{2k-2}})). \tag{6.3}$$

We will postpone the proof of Theorem 6.2 and property (6.1b) to Section 8, as they will both follow from our analysis of a particle moving inside a wedge (see Section 7). Here we will show how Theorem 6.2 is utilized to prove the unboundedness property (6.1a). In fact, (6.1a) will be obtained by using the estimate from Theorem 6.2 in a modified version of the unboundedness proof in [LW92, pp. 159–160]. First, we need to make some preparatory steps.

PROPOSITION 6.1. *For every $x \in \mathcal{M}^+$, we have*

$$q_{n_{2k+1}} > q_{n_{2k}} + \varLambda\|(0, \delta\eta_{n_{2k}})\|^2. \tag{6.4}$$

*Proof.* Without loss of generality let $d_{T^{n_{2k}}x}T^{n_{2k+1}-n_{2k}}$ be $d\varPhi_{1,2}d\varPhi_{2,3}$. Using (6.2), we estimate

$$\begin{aligned}
q_{n_{2k+1}} &= Q(d_{T^{n_{2k}}x}T^{n_{2k+1}-n_{2k}}(\delta\xi_{n_{2k}}, \delta\eta_{n_{2k}}))\\
&= Q\left(\begin{pmatrix} M_1M_2 & M_1U_2 + U_1M_2^T \\ 0 & M_1^TM_2^T \end{pmatrix}\begin{pmatrix}\delta\xi_{n_{2k}} \\ \delta\eta_{n_{2k}}\end{pmatrix}\right)\\
&= \langle M_1M_2\delta\xi_{n_{2k}} + (M_1U_2 + U_1M_2^T)\delta\eta_{n_{2k}}, M_1^TM_2^T\delta\eta_{n_{2k}}\rangle\\
&= \langle M_1M_2\delta\xi_{n_{2k}}, M_1^TM_2^T\delta\eta_{n_{2k}}\rangle\\
&\quad + Q\left(\begin{pmatrix} M_1M_2 & M_1U_2 + U_1M_2^T \\ 0 & M_1^TM_2^T \end{pmatrix}\begin{pmatrix} 0 \\ \delta\eta_{n_{2k}}\end{pmatrix}\right)\\
&> \langle\delta\xi_{n_{2k}}, \delta\eta_{n_{2k}}\rangle + \varLambda\|(0, \delta\eta_{n_{2k}})\|^2 = q_{n_{2k}} + \varLambda\|(0, \delta\eta_{n_{2k}})\|^2. \ \blacksquare
\end{aligned}$$

PROPOSITION 6.2. *Let $(a_{n_k})_{k\in\mathbb{N}}$ be a sequence of positive numbers and $C$ a positive constant. Then*

$$\sum_{i=0}^{+\infty} a_{n_{2i}} = +\infty \implies \sum_{k=0}^{+\infty} \frac{a_{n_{2k}}}{C + \sum_{i=0}^{k} a_{n_{2i}}} = +\infty.$$

*Proof.* For $1 \leq j \leq l$, we have

$$\sum_{k=j}^{l} \frac{a_{n_{2k}}}{C + \sum_{i=0}^{k} a_{n_{2i}}} > \frac{\sum_{k=j}^{l} a_{n_{2k}}}{C + \sum_{i=0}^{j-1} a_{n_{2i}} + \sum_{i=j}^{l} a_{n_{2i}}} \to 1 \qquad \text{as } l \to +\infty.$$

As the tail does not tend to zero, the series diverges. ■

Let $(q_{n_{2k-1}})_{k\in\mathbb{N}}$ be as in (6.3). Since $\prod_{k=1}^{+\infty} q_{n_{2k-1}}/q_{n_{2k-2}} = +\infty$ implies $\lim_{n\to+\infty} q_{n_{2k-1}} = +\infty$, we will estimate

$$\prod_{k=1}^{+\infty} \frac{q_{n_{2k-1}}}{q_{n_{2k-2}}} \geq \prod_{k=1}^{+\infty}(1+r_k),$$

and further prove that $\sum_{k=1}^{+\infty} r_k = +\infty$, which yields the unboundedness.

Before starting the proof of (6.1a) we need some preliminaries:

(1) From the definition of the monodromy maps, we immediately obtain

$$\begin{aligned} d\Phi_{0,1}(\delta\xi_{n-1},\delta\eta_{n-1}) &= \begin{pmatrix} \delta\xi_{n-1} \\ B\delta\xi_{n-1}+\delta\eta_{n-1}\end{pmatrix} = \begin{pmatrix}\delta\xi_n \\ \delta\eta_n\end{pmatrix}, \\ d\Phi_{i,i+1}(\delta\xi_{n-1},\delta\eta_{n-1}) &= \begin{pmatrix} M_i\delta\xi_{n-1}+U_i\delta\eta_{n-1} \\ M_i^T\delta\eta_{n-1}\end{pmatrix} = \begin{pmatrix}\delta\xi_n \\ \delta\eta_n\end{pmatrix}, \ i=1,2. \end{aligned} \tag{6.5}$$

(2) Cheng and Wojtkowski [ChW91] introduced the norm

$$\|\delta\xi\|_{CW}^2 = \sum_{i=1}^{2} \frac{(\delta\xi_{i+1}-\delta\xi_i)^2}{m_i}.$$

The maps $M_i$ are isometries with respect to this norm, i.e.

$$\|M_i\delta\xi\|_{CW} = \|\delta\xi\|_{CW}. \tag{6.6}$$

(3) The equivalence of norms gives us constants $D_1, D_2 > 0$ such that

$$D_1\|\delta\xi\|_{\max} \leq \|\delta\xi\|_{CW} \leq D_2\|\delta\xi\|_{\max}, \tag{6.7}$$

where $\|\cdot\|_{\max}$ denotes the maximum norm.

(4) Using the definitions of the Hamiltonian and the terms $\alpha_i$ of (4.3), we calculate

$$\max\{\alpha_1,\alpha_2\} \leq \frac{4\sqrt{2c}\,m_1^3}{m_3^2\sqrt{m_3}}, \tag{6.8}$$

where $c > 0$ is the energy of the system.

(5) Let $(i, i+1)$, $i = 0, 1, 2$, stand for a collision of ball $i$ with ball $i+1$, i.e. when $q_i = q_{i+1}$. When $i = 0$ the system experiences a collision with the floor.

*Proof of property (6.1a).* The proof is based on the scheme in [LW92, pp. 159–160].

We first estimate $\|\delta\xi_{n_{2k-1}}\|_{CW}$ between the points $T^{n_{2k-1}}x$ and $T^{n_{2k-2}}x$. Without loss of generality we assume $d_{T^{n_{2k-2}}x}T^{n_{2k-1}-n_{2k-2}}$ is $d\Phi_{1,2}d\Phi_{2,3}$ for every $k \in \mathbb{N}$. We estimate

$$\begin{aligned}
(6.9)\qquad \|\delta\xi_{n_{2k-1}}\|_{CW} &= \|M_1 M_2 \delta\xi_{n_{2k-2}} + (M_1 U_2 + U_1 M_2^T)\delta\eta_{n_{2k-2}}\|_{CW} \\
&\le \|\delta\xi_{n_{2k-2}}\|_{CW} + \|(M_1 U_2 + U_1 M_2^T)\delta\eta_{n_{2k-2}}\|_{CW} \\
&\le \|\delta\xi_{n_{2k-2}}\|_{CW} + D_2 \left\| \begin{pmatrix} \alpha_1 & (1+\gamma_1)\alpha_2 + (1-\gamma_2)\alpha_1 \\ 0 & \alpha_2 \end{pmatrix} \delta\eta_{n_{2k-2}} \right\|_{\max} \\
&\le \|\delta\xi_{n_{2k-2}}\|_{CW} + D_2 3 \max\{\alpha_1, \alpha_2\} \|\delta\eta_{n_{2k-2}}\|_{\max} \\
&\le \|\delta\xi_{n_{2k-2}}\|_{CW} + \frac{D_2 12\sqrt{2c}\, m_1^3}{m_3^2 \sqrt{m_3}} \|\delta\eta_{n_{2k-2}}\|_{\max}.
\end{aligned}$$

We abbreviate

$$K = \frac{D_2 12\sqrt{2c}\, m_1^3}{m_3^2 \sqrt{m_3}}.$$

Between the points $T^{n_{2k}}x$ and $T^{n_{2k-1}}x$ we have one of the following situations: Either a floor collision occurs, in which case $\|\delta\xi_{n_{2k}}\|_{CW} = \|\delta\xi_{n_{2k}-1}\|_{CW}$, or a ball-to-ball collision occurs, in which case $\|\delta\xi_{n_{2k}}\|_{CW} \le \|\delta\xi_{n_{2k}-1}\|_{CW} + \|U_{\kappa(n_{2k}-1)}\delta\eta_{n_{2k}-1}\|_{CW}$ (see (6.5)). Here $\kappa : \mathbb{N} \to \{1,2\}$ depends on the point and describes whether we have a $(1,2)$ or $(2,3)$ collision. Combining this with (6.9) we obtain

$$\begin{aligned}
(6.10)\qquad \|\delta\xi_{n_{2k}}\|_{CW} &\le \|\delta\xi_{n_0}\|_{CW} + \sum_{i=1}^{k} \sum_{j\in I_i} \|U_{\kappa(n_{2i}-j)}\delta\eta_{n_{2i}-j}\|_{CW} \\
&\quad + K \sum_{i=1}^{k} \|\delta\eta_{n_{2i-2}}\|_{CW},
\end{aligned}$$

where $|I_i|$ is the number of ball-to-ball collisions happening between $T^{n_{2i}}x$ and $T^{n_{2i-1}}x$. If $|I_i| = 0$, we set $\|U_{\kappa(n_{2i})}\delta\eta_{n_{2i}}\|_{CW} = 0$.

The Cauchy–Schwarz inequality gives

$$q_{n_k} = \langle \delta\xi_{n_k}, \delta\eta_{n_k} \rangle \le \|\delta\xi_{n_k}\|\, \|\delta\eta_{n_k}\|,$$

which yields

$$(6.11)\qquad \|\delta\eta_{n_k}\| \ge \frac{q_{n_k}}{\|\delta\xi_{n_k}\|}.$$

From Proposition 6.1 and the Cauchy–Schwarz inequality, we get

$$\begin{aligned}
q_{n_{2k+1}} &> q_{n_{2k}} + \Lambda \|\delta\eta_{n_{2k}}\|_{\max}^2 \ge q_{n_{2k}} + \Lambda \|\delta\eta_{n_{2k}}\|_{\max} \frac{q_{n_{2k}}}{\|\delta\xi_{n_{2k}}\|_{\max}} \\
&\ge q_{n_{2k}} \left(1 + \Lambda \frac{D_1 \|\delta\eta_{n_{2k}}\|_{\max}}{\|\delta\xi_{n_{2k}}\|_{CW}}\right).
\end{aligned}$$

Utilizing the above, we estimate

$$\frac{q_{n_{2k+1}}}{q_{n_{2k}}} \geq 1 + \frac{\Lambda D_1 \|\delta\eta_{n_{2k}}\|_{\max}}{\|\delta\xi_{n_0}\|_{CW} + \sum_{i=1}^{k}\sum_{j\in I_i}\|U_{\kappa(n_{2i}-j)}\delta\eta_{n_{2i}-j}\|_{CW} + K\sum_{i=1}^{k}\|\delta\eta_{n_{2i-2}}\|_{CW}}.$$

Let

$$r_k = \frac{\Lambda D_1 \|\delta\eta_{n_{2k}}\|_{\max}}{\|\delta\xi_{n_0}\|_{CW} + \sum_{i=1}^{k}\sum_{j\in I_i}\|U_{\kappa(n_{2i}-j)}\delta\eta_{n_{2i}-j}\|_{CW} + K\sum_{i=1}^{k}\|\delta\eta_{n_{2i-2}}\|_{CW}}.$$

Without loss of generality assume ([5]) that $\sum_{i=1}^{+\infty}\sum_{j\in I_i}\|U_{\kappa(n_{2i}-j)}\delta\eta_{n_{2i}-j}\|_{CW}$ is finite. The only thing left to show is that $\sum_{k=1}^{+\infty} r_k = +\infty$. In view of Proposition 6.2, it will follow once we prove that $\sum_{i=0}^{+\infty}\|\delta\eta_{n_{2i}}\|_{\max} = +\infty$. Assume otherwise. Then, by (6.9), the sequence $(\|\delta\xi_{n_{2k-1}}\|_{CW})_{k\in\mathbb{N}}$ is bounded from above. This and (6.11) imply that $(\|\delta\eta_{n_{2k}}\|_{\max})_{k\in\mathbb{N}}$ is bounded away from zero, which contradicts our assumption. This yields the unboundedness. ∎

**7. Particle falling in a wedge.** Wojtkowski [W98] analyzed the hyperbolicity of a particle moving along parabolic trajectories in a variety of wedges. The particle is subject to constant acceleration and collides with the walls of the wedge. We adopt his notation and call such a system a particle falling in a wedge, or a PW system for short. Heuristically, for special wedges, namely simple ones, a PW system is equivalent to a falling balls system (or FB system) with particular masses. After introducing the basic setup in three dimensions we are going to recall and expand some of the results in [W98] in order to prove Theorem 6.2 and property (6.1b) in Section 8.

Let $E$ be the three-dimensional Euclidean space. For three linearly independent vectors $\{e_1, e_2, e_3\}$ we define the wedge $W(e_1, e_2, e_3) \subset E$ by

$$W(e_1, e_2, e_3) = \{e \in E : e = \lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3,\ \lambda_i \geq 0,\ i = 1, 2, 3\}.$$

The vectors $\{e_1, e_2, e_3\}$ are called the *generators* of the wedge. We denote by $S(e_1, \ldots, e_i)$, $1 \leq i \leq 3$, the linear subspace spanned by $\{e_1, \ldots, e_i\}$. A three-dimensional wedge is called *simple* if the generators can be ordered in such a way that the orthogonal projection of $e_1$ resp. $e_2$ onto $S(e_2, e_3)$ resp. $S(e_3)$ is a positive multiple of $e_2$ resp. $e_3$. The simplicity of a wedge can be verified by means of the following

([5]) If the sum is infinite, then we can apply the argument in [LW92, pp. 159–160] directly. The key point is that we do not control this sum, so we assume the worst case, namely, its finiteness.

Proposition 7.1 ([W98, Proposition 2.3]). *Let $\{e_1, e_2, e_3\}$ be a set of linearly independent unit vectors. The wedge $W(e_1, e_2, e_3)$ is simple if and only if*

$$\langle e_i, e_{i+1}\rangle > 0, \quad i = 1, 2, \tag{7.1a}$$

$$\langle e_1, e_3\rangle = \langle e_1, e_2\rangle\langle e_2, e_3\rangle. \tag{7.1b}$$

The angles $\alpha_i = \sphericalangle(e_i, e_{i+1})$, $i = 1, 2$, completely determine the geometry of the wedge. In a simple wedge the angles satisfy $0 < \alpha_i < \pi/2$, and if $\{e_1, e_2, e_3\}$ are unit vectors, we have

$$\cos\alpha_i = \langle e_i, e_{i+1}\rangle. \tag{7.2}$$

We also give another geometric characterization of the wedge by introducing another pair of angles $\beta_1$, $\beta_2$. Here $\beta_i$ is the angle between $S(e_i, e_{i+2})$ and $S(e_{i+1}, e_{i+2})$, where for $i = 2$, we set $\beta_2 = \alpha_2$. If the wedge is simple, then $0 < \beta_i < \pi/2$. The relation between $\beta_1$ and $\alpha_1, \alpha_2$ is given by

$$\tan\beta_1 = \frac{\tan\alpha_1}{\sin\alpha_2}. \tag{7.3}$$

Consider the FB system from Section 2. Its Hamiltonian is given by

$$\begin{aligned} H(q,p) &= \tfrac{1}{2}\langle Kp, p\rangle + \langle c_1, q\rangle, \\ K &= \operatorname{diag}(1/m_1, 1/m_2, 1/m_3), \qquad c_1 = (m_1, m_2, m_3). \end{aligned}$$

Here $K$ is the diagonal matrix with diagonal entries $1/m_1, 1/m_2, 1/m_3$. The unit vectors

$$e_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \quad e_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\1\end{pmatrix}, \quad e_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

span the configuration space

$$W_q(e_1, e_2, e_3) = \{(q_1, q_2, q_3) \in \mathbb{R}^3 : 0 \le q_1 \le q_2 \le q_3\}.$$

It carries the natural Riemannian metric given by the kinetic energy $\langle K\cdot, \cdot\rangle$. We subject the system to the coordinate transformation

$$x_i = \sqrt{m_i}\, q_i, \qquad w_i = \frac{p_i}{\sqrt{m_i}}, \tag{7.4}$$

and obtain the Hamiltonian

$$H(x, w) = \tfrac{1}{2}\langle w, w\rangle + \langle c_2, x\rangle, \qquad c_2 = (\sqrt{m_1}, \sqrt{m_2}, \sqrt{m_3}).$$

The natural Riemannian metric in these coordinates is the standard Euclidean inner product. The new generators of length one are

$$h_1 = \frac{1}{\sqrt{M_1}} \begin{pmatrix} \sqrt{m_1} \\ \sqrt{m_2} \\ \sqrt{m_3} \end{pmatrix}, \quad h_2 = \frac{1}{\sqrt{M_2}} \begin{pmatrix} 0 \\ \sqrt{m_2} \\ \sqrt{m_3} \end{pmatrix}, \quad h_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{7.5}$$

where $M_i = m_i + \cdots + m_3$, $i = 1, 2$. The configuration space changes to

$$W_x(h_1, h_2, h_3) = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \le \frac{x_1}{\sqrt{m_1}} \le \frac{x_2}{\sqrt{m_2}} \le \frac{x_3}{\sqrt{m_3}} \right\}. \tag{7.6}$$

With respect to the Euclidean inner product we have

$$\langle h_i, h_j \rangle = \frac{\sqrt{M_j}}{\sqrt{M_i}}, \qquad 1 \le i < j \le 3,$$

which immediately yields properties (7.1a), (7.1b) from Proposition 7.1, proving that $W_x(h_1, h_2, h_3)$ is a simple wedge. Further, using (7.2) and (7.3) we get a direct link between the angles characterizing the wedge and the masses of the FB system:

$$\cos^2 \alpha_i = \frac{M_{i+1}}{M_i}, \qquad \sin^2 \alpha_i = \frac{m_i}{M_i}, \qquad \tan^2 \beta_i = \frac{m_i}{m_{i+1}}. \tag{7.7}$$

Notice that the acceleration vector is directed along the first generator. We have arrived at the important conclusion that a PW system in a simple wedge with acceleration vector along the first generator is equivalent to a FB system with appropriate masses.

### 7.1. Wide wedges

DEFINITION 7.1. A three-dimensional wedge with generators $\{g_1, g_2, g_3\}$ is *wide* if the angles of the generators exceed $\pi/2$, i.e. $\langle g_i, g_j \rangle < 0$ for $1 \le i < j \le 3$.

Consider a PW system in a simple wedge $W_x(h_1, h_2, h_3)$ of (7.6). We will unfold $W_x(h_1, h_2, h_3)$ to a wide wedge by continuously reflecting it in the faces possessing the first generator, i.e. $W(h_1, h_2)$ and $W(h_1, h_3)$. It is not hard to see that this procedure creates a wide wedge if and only if the angle between $S(h_1, h_2)$ and $S(h_1, h_3)$ is exactly ([6]) $\pi/3$. This translates to the condition

$$1/2 = \cos(\pi/3) = \langle n_{S(h_1,h_2),0}, n_{S(h_1,h_3),0} \rangle, \tag{7.8}$$

where $n_{S(h_1,h_2),0}$ resp. $n_{S(h_1,h_3),0}$ are the unit normal vectors of the subspace in the subscript. Using (7.5) in (7.8) we deduce that the masses of the corresponding FB system satisfy

$$2\sqrt{m_1}\sqrt{m_3} = \sqrt{m_1 + m_2}\sqrt{m_2 + m_3}. \tag{7.9}$$

([6]) Otherwise the unfolded simple wedges would overlap.

In this way we obtain new generators $\{g_1, g_2, g_3\}$ and the wedge $W_x(g_1, g_2, g_3)$, which consists exactly of six simple wedges. With the help of (7.5) and elementary linear algebra it follows rather easily that the wedge $W_x(g_1, g_2, g_3)$ is wide.

The two-dimensional inner faces of the simple wedges possessing the first generator $h_1$ correspond to a collision of two balls in the associated FB system. When the particle hits one of the inner faces, we allow the particle to pass through the face to the adjacent wedge.

A collision of the particle with one of the faces of the wide wedge corresponds to a collision with the floor in the associated FB system. In this case, we do not allow the particle to pass through the face, but instead reflect the velocity vector across the face by using $w_1^+ = -w_1^-$.

Since the trajectory is parabolic, a natural question to ask is whether grazing collisions can occur. For our purposes we will confine ourselves to the simple wedge $W_x(h_1, h_2, h_3)$. The definition of a grazing collision is as follows:

Definition 7.2. A collision of the trajectory $x(t)$, at time $t_0$, with one of the faces of the simple wedge $W_x(h_1, h_2, h_3)$ is *grazing* if the velocity vector $\dot{x}(t_0)$ lies in the face of collision.

The next result gives equivalent conditions of a grazing collision with one of the faces possessing the first generator.

Proposition 7.2. *Let $t_0 < t_1$ be consecutive collision times of the trajectory in the simple wedge $W_x(h_1, h_2, h_3)$ and assume that $x(t_1) \in W_x(h_1, h_2)$ or $x(t_1) \in W_x(h_1, h_3)$. The following statements are equivalent:*

(1) *A collision with the face $W_x(h_1, h_3)$ resp. $W_x(h_1, h_2)$, at time $t_1$, is grazing.*
(2) *The difference $\frac{w_1^+(t_0)}{\sqrt{m_1}} - \frac{w_2^+(t_0)}{\sqrt{m_2}}$ resp. $\frac{w_2^+(t_0)}{\sqrt{m_2}} - \frac{w_3^+(t_0)}{\sqrt{m_3}}$ is zero.*
(3) *The trajectory segment $\{x(t) : t \in [t_0, t_1]\}$ is confined to $W_x(h_1, h_3)$ resp. $W_x(h_1, h_2)$.*

*Proof.* (1)⇒(2): Without loss of generality assume that $x(t_0)$ is in $W_x(h_1, h_2)$ or in $W_x(h_2, h_3)$. Further, let the particle experience a grazing collision with the face $W_x(h_1, h_3)$ at time $t_1$. In a grazing collision the velocity

$$w^-(t_1) = \begin{pmatrix} -\sqrt{m_1}(t_1 - t_0) + w_1^+(t_0) \\ -\sqrt{m_2}(t_1 - t_0) + w_2^+(t_0) \\ -\sqrt{m_3}(t_1 - t_0) + w_3^+(t_0) \end{pmatrix}$$

is parallel to the face

$$W_x(h_1,h_3) = \left\{(x_1,x_2,x_3) \in W_x(h_1,h_2,h_3) : \frac{x_1}{\sqrt{m_1}} = \frac{x_2}{\sqrt{m_2}}\right\}.$$

This is equivalent to

$$\frac{w_1^+(t_0)}{\sqrt{m_1}} = \frac{w_2^+(t_0)}{\sqrt{m_2}}.$$

The argument for a grazing collision with $W_x(h_1,h_2)$ is exactly the same.

(2)⇒(3): Without loss of generality assume again $x(t_0) \in W_x(h_1,h_2)$ or $x(t_0) \in W_x(h_2,h_3)$ and let the particle collide with $W_x(h_1,h_3)$ at time $t_1$. From the Hamiltonian equations, we calculate the first collision time

$$t_1 - t_0 = \frac{x_2(t_0)/\sqrt{m_2} - x_1(t_0)/\sqrt{m_1}}{w_1^+(t_0)/\sqrt{m_1} - w_2^+(t_0)/\sqrt{m_2}}.$$

Since the energy is fixed, $t_1 - t_0 < \infty$. It follows that if $w_1^+(t_0)/\sqrt{m_1} - w_2^+(t_0)/\sqrt{m_2} \to 0$, then $x_2(t_0)/\sqrt{m_2} - x_1(t_0)/\sqrt{m_1} \to 0$ with (at least) the same rate. Thus, in the case of equal velocities, we always have $x_1(t_0)/\sqrt{m_1} = x_2(t_0)/\sqrt{m_2}$, which implies that the trajectory moves inside the face $W_x(h_1,h_3)$. The argument for $w_2^+(t_0)/\sqrt{m_2} - w_3^+(t_0)/\sqrt{m_3} = 0$ is exactly the same.

(3)⇒(1): This is immediate. ∎

**7.2. Projection.** The Hamiltonian equations imply that the flow is an inverted parabola. Let $[t_0,t_c]$ be the time from one collision to the next. We define the planar subspace

$$\mathsf{P}_{x([t_0,t_c])} = S(\dot{x}(t_1),\dot{x}(t_2)), \quad \dot{x}(t_1) \neq \dot{x}(t_2),\ t_0 \le t_1 < t_2 \le t_c. \tag{7.10}$$

The movement of the parabolic trajectory is confined to that subspace, i.e.

$$\{x(t) : t \in [t_0,t_c]\} \subset \mathsf{P}_{x([t_0,t_c])}.$$

The acceleration vector $a = \ddot{x}(t)$ is always an element of $\mathsf{P}_{x([t_0,t_c])}$: Set

$$n_{\mathsf{P}}(t) = \dot{x}(t) \times \ddot{x}(t), \quad \|\dot{x}(t)\| = \|\ddot{x}(t)\| = 1, \quad \forall t \in [t_0,t_c].$$

The vector $n_{\mathsf{P}}(t)$ has unit length and since the trajectory moves inside a planar subspace, $n_{\mathsf{P}}(t)$ is constant, for all choices $t \in [t_0,t_c]$. Thus, $\dot{n}_{\mathsf{P}}(t) = 0$ Observe that

$$\langle n_{x(t)}, \dot{x}(t)\rangle = 0, \quad \forall t \in [t_0,t_c], \tag{7.11}$$

where $n_{x(t)}$ is a normal vector to $\dot{x}(t)$ at $x(t)$. Differentiating (7.11) gives

$$\langle n_{x(t)}, \ddot{x}(t)\rangle = -\langle \dot{n}_{x(t)}, \dot{x}(t)\rangle.$$

Replacing $\ddot{x}(t)$ with $a$ and $n_{x(t)}$ with $n_{\mathsf{P}}(t)$ yields

$$\langle a, n_{\mathsf{P}}(t)\rangle = -\langle \dot{n}_{\mathsf{P}}(t), \dot{x}(t)\rangle = 0.$$

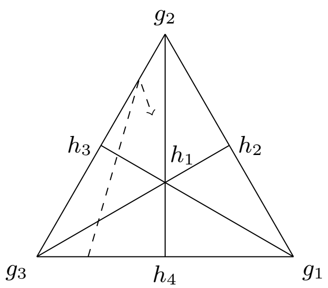


Fig. 1. The projected parabola moving inside the projected configuration space $\triangle$

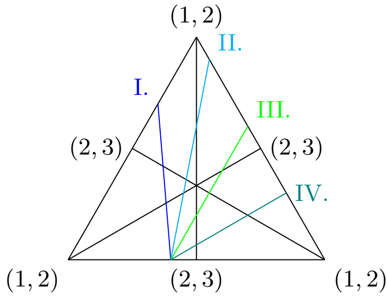


Fig. 2. An example of cases I–IV

We will use this fact to project the configuration space $W_x(g_1, g_2, g_3)$ along the first generator $h_1$ to the plane spanned by the normal vectors $n_{S(h_1,h_2)}$, $n_{S(h_1,h_3)}$ of the subspaces $S(h_1, h_2)$, $S(h_1, h_3)$. The projected configuration space becomes an equilateral triangle. Its algebraic form is

$$\triangle : \sqrt{m_1}\, x_1 + \sqrt{m_2}\, x_2 + \sqrt{m_3}\, x_3 = d, \qquad d > 0, \tag{7.12}$$

where $d$ determines its displacement from the origin. Since the acceleration vector lies in the plane spanned by two velocity vectors of the flow, the parabola projected to $\triangle$ becomes a straight line (see Figure 1).

**7.3. Proper alignment in wide wedges.** The idea to unfold the simple wedge $W_x$ of (7.6) into a wide wedge comes from [W16]. It is evident that the triple collision states in the configuration space, which are represented by the first generator $h_1$, disappear in the wide wedge. More precisely, each trajectory which passes through the spot where $h_1$ was has a smooth continuation. Since the triple collision singularity manifold is the only obstacle to proper alignment, the system of a particle falling in the wide wedge, obtained for the special mass configuration (7.9), satisfies the proper alignment condition. However, in the simple wedge $W_x$, once a trajectory hits the corner $h_1$ it is impossible to continue it uniquely, since it has two images after the singular collision. This holds for any possible mass configuration. Therefore, proper alignment cannot be immediately deduced from the dynamics of the wide wedge. It remains unknown so far (see Subsection 5.1.1 for more details).

**8. Strict unboundedness: part II.** Consider a PW system in the simple wedge $W_x(h_1, h_2, h_3)$ of (7.6) and with mass restrictions (7.9). Due to the results of the last section we reflect the simple wedge in its faces possessing the first generator to obtain a wide wedge $W_x(g_1, g_2, g_3)$.

To get strict unboundedness, it remains to prove Theorem 6.2 and property (6.1b) from Section 6. The latter was already proven as part of the Main

Theorem 6.6 in [W98, pp. 327–331]. In essence, Wojtkowski proved that every orbit will eventually hit every face of the wide wedge. Subsequently, this yields all collisions necessary to eventually map the Lagrangian subspaces $L_1$ and $L_2$ inside the interior of the contracting cone field ([7]).

To prove Theorem 6.2 we first establish how many different collisions, involving all the balls, are possible between two consecutive collisions of the lowest ball with the floor. Using the projection to $\triangle$ (see (7.12)), we encounter the following four different possibilities (see Figure 2):

$$\begin{aligned} &\text{I.} && (0,1)\to(1,2)\to(2,3)\to(0,1),\\ &\text{II.} && (0,1)\to(1,2)\to(2,3)\to(1,2)\to(0,1),\\ &\text{III.} && (0,1)\to(2,3)\to(1,2)\to(0,1),\\ &\text{IV.} && (0,1)\to(2,3)\to(1,2)\to(2,3)\to(0,1). \end{aligned} \tag{8.1}$$

*Proof of Theorem 6.2.* Since every collision in the FB system happens infinitely often, we distinguish between two sets of orbits, $O_1$ and $O_2$:

(1) $O_1$ consists of all the orbits where at least one of the cases I–IV above happens infinitely often.
(2) $O_2$ consists of all the orbits where each case I–IV occurs at most finitely many times.

$O_2$ can be considered as a special case, where between two consecutive collisions of the lowest ball with the floor, at most one ball-to-ball collision occurs. We begin with $O_1$. By symmetry it is enough to consider the first two cases of (8.1). Without loss of generality we start at time $t_0$ on the face $W_x(h_4, g_3)$. In case I, the order of the faces crossed by the trajectory is $W_x(h_1, g_3)$, $W_x(h_1, h_3)$ before the particle hits the last face $W_x(h_3, g_2)$. In case II, the trajectory crosses the faces $W_x(h_1, g_3)$, $W_x(h_1, h_3)$, $W_x(h_1, g_2)$ before it reaches $W_x(h_2, g_2)$. We compactly display this information as

Case I: $W_x(g_3, h_4) \to W_x(h_1, g_3) \to W_x(h_1, h_3) \to W_x(h_3, g_2)$.

Case II: $W_x(g_3, h_4) \to W_x(h_1, g_3) \to W_x(h_1, h_3) \to W_x(h_1, g_2) \to W_x(h_2, g_2)$.

Case I. Let $t_1 < t_2 < t_3$ be the collision times with the faces $W_x(h_1, g_3)$, $W_x(h_1, h_3)$ and $W_x(h_3, g_2)$. When the particle crosses $W_x(h_1, g_3)$ resp. $W_x(h_1, h_3)$, we have

$$\frac{w_1^-(t_1)}{\sqrt{m_1}} - \frac{w_2^-(t_1)}{\sqrt{m_2}} > 0 \quad \text{resp.} \quad \frac{w_2^-(t_2)}{\sqrt{m_2}} - \frac{w_3^-(t_2)}{\sqrt{m_3}} > 0. \tag{8.2}$$

([7]) One can directly calculate that all $(\delta\xi, 0) \in L_1$ get mapped into $\mathcal{C}(x)$ after at most three returns to the floor and all $(0, \delta\eta) \in L_2$ as soon as the trajectory experiences the first two ball-to-ball collisions.

The velocity differences are invariant between collisions, i.e.

$$\begin{aligned} \frac{w_1^-(t_1)}{\sqrt{m_1}} - \frac{w_2^-(t_1)}{\sqrt{m_2}} &= \frac{w_1^+(t_0)}{\sqrt{m_1}} - \frac{w_2^+(t_0)}{\sqrt{m_2}}, \\ \frac{w_2^-(t_2)}{\sqrt{m_2}} - \frac{w_3^-(t_2)}{\sqrt{m_3}} &= \frac{w_2^+(t_1)}{\sqrt{m_2}} - \frac{w_3^+(t_1)}{\sqrt{m_3}}. \end{aligned} \tag{8.3}$$

Due to Proposition 7.2, the quantities (8.2) are arbitrarily close to zero if and only if the collisions with the respective faces are arbitrarily close to grazing ones. The first collision with $W_x(h_1, g_3)$ is almost grazing if and only if the planar subspace $\mathsf{P}_{x([t_0,t_3])}$ (see (7.10)) is almost perpendicular to $W_x(g_1, g_2)$, i.e. $x(t_3) \in W_x(g_1, g_2)$. But this contradicts the trajectory reaching the last face $W_x(h_3, g_2)$. Therefore, there exists $\psi_1 > 0$ such that for all $x(t_0) \in W_x(h_4, g_3)$,

$$\sphericalangle(\mathsf{P}_{x([t_0,t_3])}, W_x(h_1, g_3)) > \psi_1. \tag{8.4}$$

The second collision with $W_x(h_1, h_3)$ is almost grazing if and only if $\mathsf{P}_{x([t_0,t_c])}$ is almost perpendicular to $W_x(g_2, g_3)$, i.e. $x(t_0) \in W_x(h_4, g_1)$. But this contradicts $x(t_0) \in W_x(h_4, g_3)$. Therefore, there exists $\psi_2 > 0$ such that for all $x(t_0) \in W_x(h_4, g_3)$,

$$\sphericalangle(\mathsf{P}_{x([t_0,t_3])}, W_x(h_1, h_3)) > \psi_2. \tag{8.5}$$

Using the projection along the first generator (see (7.12) and Figure 1) we conclude that $\psi_1 = \psi_2 = \pi/6$.

Case II. Let $t_1 < t_2 < t_3 < t_4$ be the collision times of the particle with the faces $W_x(h_1, g_3)$, $W_x(h_1, h_3)$, $W_x(h_1, g_2)$ and $W_x(h_2, g_2)$. It is sufficient to prove that either

$$\frac{w_1^-(t_1)}{\sqrt{m_1}} - \frac{w_2^-(t_1)}{\sqrt{m_2}} \quad \text{and} \quad \frac{w_2^-(t_2)}{\sqrt{m_2}} - \frac{w_3^-(t_2)}{\sqrt{m_3}}, \tag{8.6}$$

or

$$\frac{w_2^-(t_2)}{\sqrt{m_2}} - \frac{w_3^-(t_2)}{\sqrt{m_3}} \quad \text{and} \quad \frac{w_1^-(t_3)}{\sqrt{m_1}} - \frac{w_2^-(t_3)}{\sqrt{m_2}} \tag{8.7}$$

are uniformly bounded away from zero.

In order to reach the last face $W_x(h_2, g_2)$, the quantity $w_2^-(t_2)/\sqrt{m_2} - w_3^-(t_2)/\sqrt{m_3}$ is always uniformly bounded away from zero. Otherwise, due to Proposition 7.2, $\mathsf{P}_{x([t_0,t_4])}$ would be perpendicular to $W_x(g_2, g_3)$ and thus never reach $W_x(h_2, g_2)$.

Due to Proposition 7.2, $w_1^-(t_1)/\sqrt{m_1} - w_2^-(t_1)/\sqrt{m_2}$ is arbitrarily close to zero if and only if $\mathsf{P}_{x([t_0,t_4])}$ is almost perpendicular to $W_x(g_1, g_2)$. But this implies that $w_1^-(t_3)/\sqrt{m_1} - w_2^-(t_3)/\sqrt{m_2}$ is uniformly bounded away from zero.

If $w_1^-(t_3)/\sqrt{m_1} - w_2^-(t_3)/\sqrt{m_2}$ is arbitrarily close to zero, then by the same reasoning as above, $w_1^-(t_1)/\sqrt{m_1} - w_2^-(t_1)/\sqrt{m_2}$ is uniformly bounded away from zero. Thus, in case II, either (8.6) or (8.7) are always uniformly bounded away from zero.

It is clear, due to the coordinate transformation (7.4), that $w_i/\sqrt{m_i} - w_{i+1}/\sqrt{m_{i+1}}$ is uniformly bounded from below if and only if $v_i - v_{i+1}$ is uniformly bounded from below.

Consider the FB system in $x = (\xi, \eta)$ coordinates. Along every orbit $(T^n x)_{n\in\mathbb{N}}$ we have obtained two subsequences $(T^{n_{2k}}x)_{k\in\mathbb{N}}$ and $(T^{n_{2k+1}}x)_{k\in\mathbb{N}}$, where we set $(T^{n_{2k}}x)_{k\in\mathbb{N}}$ to be the phase points before, and $(T^{n_{2k+1}}x)_{k\in\mathbb{N}}$ right after, two consecutive collisions with velocity differences bounded away from zero. This means that the derivative map $dT^{n_{2k-1}-n_{2k-2}}$ equals either $d\Phi_{2,3}d\Phi_{1,2}$ or $d\Phi_{1,2}d\Phi_{2,3}$. The two maps are both upper triangular matrices of the form

$$\begin{pmatrix} X_1 & X_2 \\ 0 & X_1^T \end{pmatrix}.$$

$X_1$ depends only on the masses, while $X_2 = X_2(\alpha_1, \alpha_2)$ depends on the masses and the velocity differences $v_i - v_{i+1}$ in $\alpha_1, \alpha_2$ (see (4.3)). Each pair of consecutive collisions with velocity differences bounded away from zero belongs to one of the cases I–IV in (8.1). Each of these velocity differences has a uniform lower bound. Set the minimum of these lower bounds to be $\Theta > 0$. Observe that

$$\begin{aligned} &Q\big(d_{T^{n_{2k-2}}x}T^{n_{2k-1}-n_{2k-2}}(0,\delta\eta)\big) \\ &\qquad = \left\langle X_2 \frac{1}{\|(0,\delta\eta)\|}\delta\eta, X_1^T \frac{1}{\|(0,\delta\eta)\|}\delta\eta \right\rangle \|(0,\delta\eta)\|^2. \end{aligned}$$

Let $X_2(\Theta)$ be the matrix in which the velocity differences in $X_2(\alpha_1, \alpha_2)$ are replaced by $\Theta$. Since $X_1(X_2(\alpha_1,\alpha_2) - X_2(\Theta))$ is positive semi-definite, we have

$$\begin{aligned} &\left\langle X_2(\alpha_1,\alpha_2) \frac{1}{\|(0,\delta\eta)\|}\delta\eta, X_1^T \frac{1}{\|(0,\delta\eta)\|}\delta\eta \right\rangle \\ &\qquad \ge \left\langle X_2(\Theta) \frac{1}{\|(0,\delta\eta)\|}\delta\eta, X_1^T \frac{1}{\|(0,\delta\eta)\|}\delta\eta \right\rangle. \end{aligned}$$

Denote by $\partial B_{\|\cdot\|}(0,1)$ the boundary of the unit ball in the tangent space with respect to the indicated norm. The functional $f(u) = \langle X_2(\Theta)u, X_1^T u\rangle$ is positive, independent of $x$ and continuous on the compact space $\partial B_{\|\cdot\|}(0,1)$. Thus, there exists a constant $\Lambda_1 > 0$ such that

$$Q(d_{T^{n_{2k-2}}x}T^{n_{2k-1}-n_{2k-2}}(0,\delta\eta)) > \Lambda_1 \|(0,\delta\eta)\|^2. \tag{8.8}$$

The special case $O_2$ reduces to the analysis of the reappearing collision sequence

$$(1,2) \to (0,1) \to (2,3) \to (0,1) \to (1,2).$$

Using the Hamiltonian flow, the collision laws and the collision times, it can be quickly calculated that the velocity differences of the $(1,2)$ collisions cannot both become arbitrarily small, since otherwise this state would leave the constant energy surface. Hence, one of them has a uniform lower bound. The same idea can be applied to obtain a uniform lower velocity difference bound for the $(2,3)$ collision, i.e. we observe a loss of energy when sufficiently reducing the value of the velocity differences of $(2,3)$ above and its successive $(2,3)$ collision. For orbits in $O_2$, $dT^{n_{2k-1}-n_{2k-2}}$ takes the form either $d\Phi_{1,2}d\Phi_{0,1}d\Phi_{2,3}$ or $d\Phi_{2,3}d\Phi_{0,1}d\Phi_{1,2}$. Since $d\Phi_{0,1}$ and $d\Phi_{2,3}$ commute, on account of the invariance of $L_2$ under $d\Phi_{0,1}$ and the $Q$-monotonicity we obtain the same estimate (8.8) with a different $\Lambda_2 > 0$. Setting $\Lambda = \min\{\Lambda_1, \Lambda_2\}$ finishes the proof of Theorem 6.2 and therefore also Theorem 6.1. ∎

As outlined in Section 2, the strict unboundedness for every orbit implies the Chernov–Sinai ansatz and the abundance of sufficiently expanding points.

**Acknowledgements.** I wish to cordially thank Maciej P. Wojtkowski for his help in outlining the difficulties of the problem and his hospitality during my visit to Opole in October 2017. Further, my gratitude goes to my advisor Henk Bruin and Péter Bálint, Nandor Simányi, Domokos Szász and Imre Péter Tóth for many helpful discussions.

The author was supported by a Marietta-Blau and Marshall Plan Scholarship.

## References

[ChW91] J. Cheng and M. Wojtkowski, *Linear stability of a periodic orbit in the system of falling balls*, in: The Geometry of Hamiltonian Systems, Tu. Ratiu (ed.), MSRI Publ., Springer, 1991, 53–71.

[Ch93] N. Chernov, *Local ergodicity in hyperbolic systems with singularities*, Funct. Anal. Appl. 27 (1993), 51–54.

[ChM06] N. Chernov and Markarian, *Chaotic Billiards*, Math. Surveys Monogr. 127, Amer. Math. Soc., 2006.

[ChS87] N. Chernov and Y. Sinai, *Ergodic properties of some systems of two-dimensional discs and three-dimensional spheres*, Russian Math. Surveys 42 (1987), no. 3, 181–207.

[KS86] A. Katok and J.-M. Strelcyn, *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities*, Lecture Notes in Math. 1222, Springer, 1986.

[KSSz90] A. Krámli, N. Simányi and D. Szász, *A transversal fundamental theorem for semi-dispersing billiards*, Comm. Math. Phys. 129 (1990), 535–560.

[LW92] C. Liverani and M. Wojtkowski, *Ergodicity in Hamiltonian systems*, in: Dynamics Reported 4, Springer, 1995, 130–202.

[S96] N. Simányi, *The characteristic exponents of the system of falling balls*, Comm. Math. Phys. 182 (1996), 457–468.

[W88] M. Wojtkowski, *Measure theoretic entropy of the system of hard spheres*, Ergodic Theory Dynam. Systems 8 (1988), 133–153.

[W90a] M. Wojtkowski, *A system of one dimensional balls with gravity*, Comm. Math. Phys. 126 (1990), 507–533.

[W90b] M. Wojtkowski, *The system of one-dimensional balls in an external field. II*, Comm. Math. Phys. 127 (1990), 425–432.

[W91] M. Wojtkowski, *Systems of classical interacting particles with nonvanishing Lyapunov exponents*, in: Lyapunov Exponents (Oberwolfach 1990), L. Arnold et al. (eds.), Lecture Notes in Math. 1486, Springer, 1991, 243–262.

[W98] M. Wojtkowski, *Hamiltonian systems with linear potential and elastic constraints*, Fund. Math. 157 (1998), 305–341.

[W16] M. Wojtkowski, talk at the conference "Hyperbolic Dynamics and Statistical Physics", dedicated to the 75th birthday of Domokos Szász, 2016; https://math.bme.hu/∼szasz75/BLT_Szasz75.pdf (2016).

Michael Hofbauer-Tsiflakos
Faculty of Mathematics
University of Vienna
Oskar-Morgenstern-Platz 1
1090 Wien, Austria
E-mail: michael.tsiflakos@univie.ac.at