SINGULARITIES AND NONHYPERBOLIC MANIFOLDS DO NOT COINCIDE

NÁNDOR SIMÁNYI

Abstract. We consider the billiard flow of elastically colliding hard disks on the flat 2-torus and prove that no singularity manifold can locally coincide with a manifold describing future non-hyperbolicity of the trajectories. As a corollary, we obtain the ergodicity (actually the Bernoulli mixing property) of all such systems.

1. Introduction

In this paper we prove the Boltzmann–Sinai Ergodic Hypothesis for hard disk systems on the 2-torus $\mathbb{R}^2/\mathbb{Z}^2$ without any assumed hypothesis or exceptional model.

This introduction is, to a large extent, an edited version of some paragraphs of the introductory sections §1 and §2 of my paper [Sim(2009)]. For a more detailed introduction into the topic of hard ball systems, please see these two introductory sections of [Sim(2009)].

In a loose form, as attributed to L. Boltzmann back in the 1880's, the Boltzmann hypothesis asserts that gases of hard balls are ergodic. In a precise form, which is due to Ya. G. Sinai in [Sin(1963)], it states that the gas of $N \geq 2$ identical hard balls (of "not too big" radius) on a torus $\mathbb{T}^\nu = \mathbb{R}^\nu/\mathbb{Z}^\nu$, $\nu \geq 2$, (a ν-dimensional box with periodic boundary conditions) is ergodic, provided that certain necessary reductions have been made. The latter means that one fixes the total energy, sets the total momentum to zero, and restricts the center of mass to a certain discrete lattice within the torus. The assumption of a not too big radius is necessary to have the interior of the arising configuration space connected.

Sinai himself pioneered rigorous mathematical studies of hard ball gases by proving the hyperbolicity and ergodicity for the case $N = 2$ and $\nu = 2$ in his seminal paper [Sin(1970)], where he laid down the foundations of the modern theory of chaotic billiards. The proofs there were further polished and clarified in [B-S(1973)]. Then

Date: May 2, 2012.
1991 *Mathematics Subject Classification.* 37D50, 34D05.
Key words and phrases. Semi-dispersing billiards, hyperbolicity, ergodicity, local ergodicity, invariant manifolds, Chernov–Sinai Ansatz.

Chernov and Sinai extended these results to ($N = 2$, $\nu \geq 2$), as well as proved a general theorem on "local" ergodicity applicable to systems of $N > 2$ balls [S-Ch(1987)]; the latter became instrumental in the subsequent studies. The case $N > 2$ is substantially more difficult than that of $N = 2$ because, while the system of two balls reduces to a billiard with strictly convex (spherical) boundary, which guarantees strong hyperbolicity, the gases of $N > 2$ balls reduce to billiards with convex, but not strictly convex, boundary (the latter is a finite union of cylinders) – and those are characterized by a very weak hyperbolicity.

Further development has been due mostly to A. Krámli, D. Szász, and the present author. We proved hyperbolicity and ergodicity for $N = 3$ balls in any dimension [K-S-Sz(1991)] by exploiting the "local" ergodic theorem of Chernov and Sinai [S-Ch(1987)], and carefully analyzing all possible degeneracies in the dynamics to obtain "global" ergodicity. We extended our results to $N = 4$ balls in dimension $\nu \geq 3$ next year [K-S-Sz(1992)], and then I proved the ergodicity whenever $N \leq \nu$ in [Sim(1992)-I] and [Sim(1992)-II]. At that point the existing methods could no longer handle any new cases, because the analysis of the degeneracies became overly complicated. It was clear that further progress should involve novel ideas.

A big step ahead was made by Szász and myself, when we used the methods of algebraic geometry in [S-Sz(1999)]. We assumed that the balls had arbitrary masses $m_1, \dots, m_N$ (but the same radius r). By taking the limit $m_N \to 0$, we were able to reduce the dynamics of N balls to the motion of $N - 1$ balls, thus utilizing a natural induction on N. Then algebro-geometric methods allowed us to effectively analyze all possible degeneracies, but only for typical (generic) $(N + 1)$-tuples of "external" parameters $(m_1, \dots, m_N, r)$; the latter needed to avoid some exceptional submanifolds of codimension one, which remained unknown. This approach led to a proof of full hyperbolicity (but not yet ergodicity) for all $N \geq 2$ and $\nu \geq 2$, and for generic $(m_1, \dots, m_N, r)$, see [S-Sz(1999)]. Later I simplified the arguments and made them more "dynamical", which allowed me to obtain full hyperbolicity for hard balls with any set of external geometric parameters $(m_1, \dots, m_N, r)$ [Sim(2002)]. The reason why the masses m_i are considered *geometric parameters* is that they determine the relevant Riemannian metric

$$||dq||^2 = \sum_{i=1}^{N} m_i ||dq_i||^2$$

of the system. Thus, the almost sure hyperbolicity has been fully established for all systems of hard balls on tori.

To upgrade the full hyperbolicity to ergodicity, one needs to refine the analysis of the mentioned degeneracies. For hyperbolicity, it was enough that the degeneracies made a subset of codimension ≥ 1 in the phase space. For ergodicity, one has to show that its codimension is ≥ 2, or to find some other ways to prove that the (possibly) arising codimension-one manifolds of non-sufficiency are incapable of separating distinct ergodic components. In the paper [Sim(2003)] I took the first step in the direction of proving that the codimension of exceptional manifolds is at least two: I proved that all systems of $N \geq 2$ balls on a 2-dimensional torus are ergodic for typical (generic) $(N+1)$-tuples of external parameters $(m_1, \dots, m_N, r)$. The proof again involves some algebro-geometric techniques, thus the result is restricted to generic parameters $(m_1, \dots, m_N; r)$. But there was a good reason to believe that systems in $\nu \geq 3$ dimensions would be somewhat easier to handle, at least that was indeed the case in earlier studies.

As the next step, in the paper [Sim(2004)] I was able to further improve the algebro-geometric methods of [S-Sz(1999)], and proved that for any $N \geq 2$, $\nu \geq 2$, and for almost every selection $(m_1, \dots, m_N; r)$ of the external geometric parameters the corresponding system of N hard balls on $\mathbb{T}^\nu$ is (fully hyperbolic and) ergodic.

Finally, in the paper [Sim(2009)] I managed to prove the Boltzmann-Sinai Ergodic Hypothesis in full generality (i. e. without exceptional models), by assuming that the so called Chernov-Sinai Ansatz is true for these models.

Remark 1.1. The Chernov-Sinai Ansatz states that for almost every singular phase point $x \in \mathcal{SR}_0^+$ (with respect to the hypersurface measure of $\mathcal{SR}_0^+$) the forward orbit $S^{(0,\infty)}x$ is sufficient (geometrically hyperbolic). This is the utmost important global geometric hypothesis of the Theorem on Local Ergodicity of [S-Ch(1987)], see also Condition 3.1 in [K-S-Sz(1990)].

Thus the only missing piece of the whole puzzle is to prove that no open piece of a singularity manifold can precisely coincide with a codimension-one manifold desribing the trajectories with a non-sufficient forward orbit segment corresponding to a fixed symbolic collision sequence. This is exactly what we prove in our Theorem below for the case of elastically colliding disks, i. e. $\nu = 2$.

2. Formulation and Proof of the Theorem

Let $U_0 \subset \mathbf{M} \setminus \partial\mathbf{M}$ be an open ball, $T > 0$, and assume that

(a) $S^T(U_0) \cap \partial\mathbf{M} = \emptyset$,

(b) S^T is smooth on U_0.

Next we assume that there is a *codimension-one*, smooth submanifold $J \subset U_0$ with the property that for every $x \in U_0$ the trajectory segment $S^{[0,T]}x$ is geometrically hyperbolic (sufficient) if and only if $x \notin J$. (J is a so called non-hyperbolicity or degeneracy manifold.) Denote the common symbolic collision sequence of the orbits $S^{[0,T]}x$ ($x \in U_0$) by $\Sigma = (e_1, e_2, \dots, e_n)$, listed this time in the increasing time order, and let the corresponding advances be $\alpha_i = \alpha(e_i)$, $i = 1, 2, \dots, n$. Let $t_i = t(e_i)$ the time of the i-th collision, $0 < t_1 < t_2 < \dots < t_n < T$.

Finally we assume that for every phase point $x \in U_0$ the first reflection $S^{\tau(x)}x$ *in the past* on the orbit of x is a singular reflection (i. e. $S^{\tau(x)}x \in \mathcal{SR}_0^+$) if and only if x belongs to a codimension-one, smooth submanifold K of U_0. For the definition of the manifold of singular reflections $\mathcal{SR}_0^+$ see, for instance, the end of §1 in [Sim(2009)].

Theorem 2.1. *Using all the assumtions and notations above, the submanifolds J and K of U_0 do not coincide.*

The rest of this section is devoted to the proof of this theorem. It will be subdivided into several lemmas and propositions.

First of all, we assume that the center x_0 of the open ball U_0 belongs to the exceptional set J. During the indirect proof of the theorem smaller and smaller open balls U_0 will be selected to guarantee a regular (smooth) behavior.

Observe that the sufficiency of the orbit segments $S^{[0,T]}x$ ($x \in U_0 \setminus J$) immediately implies that the collision graph $\mathcal{G} = (\{1, 2, \dots, N\}, \{e_1, e_2, \dots, e_n\})$ is connected on the vertex set $\mathcal{V} = \{1, 2, \dots, N\}$. Therefore, according to Lemma 2.13 of [Sim(1992)-II], the linear map

$$\Phi : \ \mathcal{N}_0\left(S^{[0,T]}x\right) \to \mathbb{R}^n$$

defined by $\left(\Phi(w)\right)_i = \alpha_i(w)$ $(i = 1, 2, \dots, n)$ is a linear embedding for every $x \in U_0$. Here $\mathcal{N}_0(S^{[0,T]}x)$ denotes the neutral linear space of the trajectory segment $S^{[0,T]}x$, see Definition 2.5 of [Sim(2009)]. The sufficiency (geometric hyperbolicity) of a trajectory segment $S^{[0,T]}x$ means that the dimension of the neutral linear space $\mathcal{N}_0(S^{[0,T]}x)$ takes the minimum possible value 1, see Definition 2.7 in [Sim(2009)]. Moreover, just like in the previous section, let $1 = k(1) < k(2) < \dots < k(N-1) < n$ be the uniquely defined indices with the property that for every l $(1 \le l \le N-1)$ the collision graph $\left(\mathcal{V}, \, \{e_1, e_2, \dots, e_{k(l)}\}\right)$ has exactly $N - l$ connected components, whereas the number of components of $\left(\mathcal{V}, \, \{e_1, e_2, \dots, e_{k(l)-1}\}\right)$ is $N - l + 1$.

We shall call the edges (collisions) $e_{k(1)}, \dots, e_{k(N-1)}$ *essential.*

For every *non-essential edge* $e_m = \{i(m), j(m)\}$ $(1 \le i(m) < j(m) \le N)$ we express the relative displacement

$$\Delta q^-_{i(m)}(t_m) - \Delta q^-_{j(m)}(t_m) = \alpha_m \left[v^-_{i(m)}(t_m) - v^-_{j(m)}(t_m)\right]$$

as a linear combination of relative velocities of earlier collisions $e_1, e_2, \dots, e_{m-1}$ (with coefficients made up from some masses and advances) precisely as described by the CPF, see Proposition 2.19 in [S-Sz(1999)].

$$\alpha_m \left[v^-_{i(m)}(t_m) - v^-_{j(m)}(t_m)\right] = \sum_{k=1}^{m-1} \alpha_k \Gamma_k^{(m)} \tag{2.2}$$

($1 \le m \le n$, e_m is not essential), where each $\Gamma_k^{(m)}$ is a linear combination of the relative velocities $v_{i(k)}^- - v_{j(k)}^-$ and $v_{i(k)}^+ - v_{j(k)}^+$, and the coefficients in these linear combinations are fractional linear expressions of the masses $m_{i(k)}$ and $m_{j(k)}$, see the CPF as Proposition 2.19 in [S-Sz(1999)]. We observe that the solution set of the system of all equations (2.2) (taken for all m with a non-essential edge e_m) is precisely the linear space $\Phi\left(\mathcal{N}_0(S^{[0,T]}x)\right) = \overline{\mathcal{N}}_0(S^{[0,T]}x)$, having the same dimension as the neutral space $\mathcal{N}_0(S^{[0,T]}x)$, $x \in U_0$.

As follows, we are presenting an indirect proof (a proof by contradiction) by assuming that the nonhyperbolicity manifold J coincides with a past singularity so that no collision takes place between the mentioned singularity and J. (Otherwise those collisions between the singularity and J could be added to the symbolic sequence $\Sigma = (e_1, e_2, \dots, e_n))$ as an initial segment.)

Throughout the proof we shall assume not only that the dimension ν of the container torus is equal to 2, but also that the masses of the interactive disks are equal: $m_1 = m_2 = \cdots = m_N$. As a matter of fact, this assumption is not a serious restriction of generality: it is merely a technical-notational assumption, and the reader can easily re-write the present proof to cover the general case of arbitrary masses. We denote by $d = 2(N-1)$ the dimension of the configuration space $\mathbf{Q}$.

Following the ideas and notations of §3 of [S-Sz(2000)], we introduce the following concepts and notations.

With every collision $e_k = (i(k), j(k))$ ($1 \le k \le n$, $1 \le i(k) < j(k) \le N$) we associate the real projective line $\mathcal{P} \cong \mathbb{RP}(1)$ of all orthogonal reflections of the common tangent space

$$\mathcal{Z} = \mathcal{T}\mathbf{Q} = \mathcal{T}_q\mathbf{Q} = \left\{ (\delta q_1, \dots, \delta q_N) \in (\mathbb{R}^2)^N \Big| \sum_{i=1}^{N} \delta q_i = 0 \right\} \cong \mathbb{R}^d \tag{2.3}$$

across all possible tangent hyperlanes H of the cylinder C_{e_k} corresponding to the collision e_k. In this way we obtain a map

$$\Phi:\ S^{d-1} \times \prod_{k=1}^{n} \mathcal{P}_k \to S^{d-1} \tag{2.4}$$

which assignes to every $(n+1)$-tuple

$$(V_0;\, g_1, g_2, \dots, g_n) \in S^{d-1} \times \prod_{k=1}^{n} \mathcal{P}_k$$

the image velocity $V_n = V_0 g_1 g_2 \dots g_n$ of V_0 under the composite action $g_1 g_2 \dots g_n$. (Here, by convention, the composition is carried out from the left to the right, and S^{d-1} denotes the unit sphere of $\mathcal{Z}$ in 2.3.) The space $M_n = S^{d-1} \times \prod_{k=1}^{n} \mathcal{P}_k$ is called the phase space of the virtual velocity process $(V_0, V_1, \dots, V_n)$, where $V_k =$

$V_0 g_1 g_2 \dots g_k$. Clearly, the velocity process $(V_0, V_1, \dots, V_n)$ uniquely determines the sequence of reflections $g_1, g_2, \dots, g_n$. For any $x \in M_n$ or $x \in U_0$ we denote the velocity V_k by $V_k(x)$. Similarly, $v^+_{i(k)} - v^+_{j(k)}$ denotes the relative velocity of the colliding particles $i(k)$ and $j(k)$ right after the collision $e_k = (i(k),\, j(k))$ $(1 \le i(k) < j(k) \le N)$, and the definition of the pre-collision relative velocity $v^-_{i(k)} - v^-_{j(k)}$ is analogous, $k = 1, 2, \dots, n$. Thus we get a natural projection

$$\Pi :\ U_0 \to M_n \tag{2.5}$$

by taking $\Pi(x) = (V_0(x);\, g_1(x), \dots, g_n(x))$ for $x = (q(x),\, v(x)) \in U_0$, where $V_0(x) = v(x)$.

What is coming up is a local analysis in a small, open ball neighborhood $B_0 \subset M_n$ of the base point $(V_0(x_0);\, g_1(x_0), \dots, g_n(x_0))$.

The Connecting Path Formula (2.2) together with the results of [S-Sz(2000)] and [Sim(2002)] yield the following results.

Proposition 2.6. *For any integer n_0, $2 \le n_0 \le n$, the neutral space*

$$\mathcal{N}_{t_1+0}(V_0;\, g_1, g_2, \dots, g_{n_0}) = \mathcal{N}_1(V_0;\, g_1, g_2, \dots, g_{n_0})$$

is fully determined by the directions *of all relative velocities $v^-_{i(l)} - v^-_{j(l)}$, $v^+_{i(l)} - v^+_{j(l)}$ $(2 \le l \le n_0 - 1)$, and by the directions of $v^+_{i(1)} - v^+_{j(1)}$ and $v^-_{i(n_0)} - v^-_{j(n_0)}$.* This property will be called the Direction Determination Principle, or DDP. *As a consequence, the neutral space $\mathcal{N}_0(V_0;\, g_1, g_2, \dots, g_{n_0})$ is fully determined by the relative velocities listed above and by $v^-_{i(1)} - v^-_{j(1)}$.*

Proof. Observe that for any tangent vector $\delta q = (\delta q_1, \dots, \delta q_N) \in \mathcal{Z}$ the relation $\delta q \in \mathcal{N}_1(V_0;\, g_1, \dots, g_{n_0})$ holds true if and only if for every k, $2 \le k \le n_0$, the vector $R_k(\delta q \cdot g_2 \cdot g_3 \cdot \dots \cdot g_{k-1})$ is parallel to the relative velocity vector $v^-_{i(k)} - v^-_{j(k)}$, and $R_1(\delta q)$ is parallel to $v^+_{i(1)} - v^+_{j(1)}$. (Here $R_k(\delta y) = \delta y_{i(k)} - \delta y_{j(k)}$ for any $\delta y = (\delta y_1, \dots, \delta y_N) \in \mathcal{Z}$.) $\square$

Proposition 2.7. *For every n_0, $1 \le n_0 \le n$, the generic ($\Longleftrightarrow$minimal) dimension (both in measure-theoretical and topological senses) of the neutral spaces*

$$\mathcal{N}_0(V_0;\, g_1, \dots, g_{n_0})$$

on the phase space M_{n_0} is equal to the generic ($\Longleftrightarrow$minimal) value of

$$dim \mathcal{N}_0(V_0(x);\, g_1(x), g_2(x), \dots, g_{n_0}(x))$$

for all $x \in U_0$. (Key Lemma 3.19 in [Sim(2002)]*.)*

The value of this typical dimension will be denoted by $\Delta(e_1, e_2, \dots, e_{n_0})$. Plainly, it only depends on the symbolic sequence $(e_1, e_2, \dots, e_{n_0})$.

The value of $\dim \mathcal{N}_0(V_0(x);\, g_1(x), \dots, g_{n_0}(x))$ for typical $x \in J$ (either in measure-theoretical or in topological sense) will be denoted by $\Delta_J(e_1, e_2, \dots, e_{n_0})$. By selecting the open balls B_0 and U_0 ($B_0 \subset M_n$, $U_0 \subset \mathbf{M}$, $U_0 = \Pi^{-1}(B_0)$) small enough we may (and shall) assume that for every integer n_0, $1 \le n_0 \le n$,

$$\dim \mathcal{N}_0\left(V_0(y);\, g_1(y), \ldots, g_{n_0}(y)\right) = \Delta(e_1, e_2, \ldots, e_{n_0}) \quad \forall y \in B_0 \setminus \tilde{J}, \tag{2.8}$$

$$\dim \mathcal{N}_0\left(V_0(y);\, g_1(y), \ldots, g_{n_0}(y)\right) = \Delta_J(e_1, e_2, \ldots, e_{n_0}) \quad \forall y \in \tilde{J}, \tag{2.9}$$

where $\tilde{J} \subset B_0$ is an analytic submanifold of B_0 with $J = \Pi^{-1}(\tilde{J})$.

For variable points $x = (V_0;\, g_1, \ldots, g_n) \in M_n$ let us consider all minors (determinants of square submatrices) $\mathcal{M}_1(x), \mathcal{M}_2(x), \ldots, \mathcal{M}_s(x)$ of the system (2.2) of all CPFs that identically vanish on $\tilde{J}$ but are not identically zero on the phase space M_n. By again selecting the ball $B_0 \subset M_n$ small enough we may and shall assume that

$$\tilde{J} = \left\{ x \in B_0 \big|\, \mathcal{M}_i(x) = 0 \right\} \text{ for } i = 1, \ldots, s. \tag{2.10}$$

Now it is time to bring up the definition of the "critical index" n_0.

Definition 2.11. The critical index n_0 is the minimum value of all the maximum column indices of minors $\mathcal{M}_i(x)$ featuring (2.10). Let $\mathcal{M}_{i_0}(x)$ be a minor (satisfying 2.10) whose biggest column index is n_0.

Due to the facts that the columns of the minor $\mathcal{M}_{i_0}(x)$ depend on the coordinates of the relative velocities in a linear way and by the Direction Determination Principle (DDP) of Proposition 2.6 we obtain a useful description of the membership relation $x \in \tilde{J}$.

Proposition 2.12. *For any $x \in B_0$ the relation $x \in \tilde{J}$ holds true (i.e. $\mathcal{M}_{i_0}(x) = 0$) if and only if the pair of relative velocities*

$$r(x) := \left(v^-_{i(n_0)}(x) - v^-_{j(n_0)}(x),\, v^+_{i(n_0)}(x) - v^+_{j(n_0)}(x) \right) \in \mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4 \tag{2.13}$$

belongs to a hyperplane $H(x) \subset \mathbb{R}^4$ depending analytically on the directions

$$dir(v^-_{i(k)}(x) - v^-_{j(k)}(x)), \quad dir(v^+_{i(k)}(x) - v^+_{j(k)}(x))$$

of the indicated relative velocities for $k = 1, 2, \ldots, n_0 - 1$.

The next result tells that the collision e_{n_0} is "essential" in the sense that it decreases the dimension of the neutral space.

Lemma 2.14.

$$\Delta(e_1, e_2, \ldots, e_{n_0}) = \Delta(e_1, e_2, \ldots, e_{n_0-1}) - 1.$$

Proof. Proof by contradiction: assume that $\Delta(e_1, \ldots, e_{n_0}) = \Delta(e_1, \ldots, e_{n_0-1})$. This means that the actual CPF of (2.2) (in which $m = n_0$) can be dropped from the whole system without affecting the solution set. Furthermore, by making the standard reduction $\alpha_{n_0} = 0$ for the advance α_{n_0} (which can be done by modifying the solution by adding to it a solution with all advances equal, and this chops off the dimension of the solution set by 1) we can completely drop the n_0-th column from the system of CPFs (2.2). This shows that the two relative velocity components of $r(x)$ in (2.13) have no effect on the solution set in question, thus the essential minor $\mathcal{M}_{i_0}(x)$ from (2.10) cannot have the n_0-th column as one of its columns. $\square$

The upcoming lemma tells us that the critical collision e_{n_0} does not distinguish between the points of $\tilde{J}$ and of $B_0 \setminus \tilde{J}$.

Lemma 2.15.

$$\Delta(e_1, e_2, \dots, e_{n_0}) = \Delta_J(e_1, e_2, \dots, e_{n_0}).$$

Proof. Again a proof by contradiction: assume that $\Delta(e_1, \dots, e_{n_0}) < \Delta_J(e_1, \dots, e_{n_0})$. According to Proposition 2.6, the neutral space

$$\mathcal{N}_{n_0-1}\left(V_0(x);\, g_1(x), \dots, g_{n_0-1}(x)\right)$$

is determined by the relative velocities $v^-_{i(l)}(x) - v^-_{j(l)}(x)$ and $v^+_{i(l)}(x) - v^+_{j(l)}(x)$ for $l = 1, 2, \dots, n_0 - 1$. On the other hand, the projection

$$R_{n_0}\left[\mathcal{N}_{n_0-1}\left(V_0(x);\, g_1(x), \dots, g_{n_0-1}(x)\right)\right]$$

of this neutral space onto $\delta q_{i(n_0)} - \delta q_{j(n_0)}$ determines if $x \in J$ is true or not: $x \notin J$ if and only if the dimension of the above projection is 2, and not 1. This, in turn, means that the relative velocities $v^-_{i(l)}(x) - v^-_{j(l)}(x)$ and $v^+_{i(l)}(x) - v^+_{j(l)}(x)$ ($l = 1, 2, \dots, n_0 - 1$) determine if $x \in J$ is true or not, thus some essential minor $\mathcal{M}_i(x)$ of (2.10) has all its column indices $< n_0$. $\square$

3. Finishing the proof of the Theorem

Consider an arbitrary point $y_0 \in J$. Let $\tau < 0$ be the unique number such that

(1) $S^\tau y_0 = y^* \in \mathcal{SR}_0^+$,

(2) $S^{(\tau,0)} y_0 \cap \partial\mathbf{M} = \emptyset$.

Here $\mathcal{SR}_0^+$ denotes the set of all singular reflections given with their outgoing (post-singularity) velocity.

Select and fix a vector w_0, $w_0 \perp v(y^*)$, such that

$$w_0 \in \mathcal{N}_0\left(V_0(y^*);\, g_1(y^*), \dots, g_{n_0-1}(y^*)\right) \setminus \mathcal{N}_0\left(V_0(y^*);\, g_1(y^*), \dots, g_{n_0}(y^*)\right). \tag{3.1}$$

This is possible, due to lemmas 2.14–2.15. Next we consider a smooth curve $\gamma_0(s)$, $|s| < \varepsilon_0$, $\gamma_0(0) = y^*$, $\gamma_0(s) \in \mathcal{SR}_0^+$, as follows:

Case A. If the singularity at y^* is a double collision (a corner of the configuration space)

(1) $v(\gamma_0(s)) = \dfrac{v(y^*) + s \cdot w_0}{||v(y^*) + s \cdot w_0||}$,

(2) $q(\gamma_0(s)) = q(\gamma_0(0)) = q(y^*)$

for $|s| < \varepsilon_0$.

Case B. If the singularity at y^* is a tangency

(1) $v(\gamma_0(s)) = \dfrac{v(y^*) + s \cdot w_0}{||v(y^*) + s \cdot w_0||}$,

(2) $q(\gamma_0(s)) = q(y^*) + \alpha \cdot w_0 + \beta \cdot v(\gamma_0(s)))$

($|s| < \varepsilon_0$) so that the relation $\gamma_0(s) \in \mathcal{SR}_0^+$ still holds true. We note that the orders of magnitude of the correction parameters α and β are $\alpha = \mathrm{O}(s^2)$, $\beta = \mathrm{O}(s)$, as a simple geometric observation shows.

Fix a time t^*, $t_{n_0-1}(y^*) < t^* < t_{n_0}(y^*)$, and investigate the image $S^{t^*}(\gamma_0(s)) = \gamma^*(s)$ of the curve γ_0 under the t^*-iterate of the billiard flow. More precisely, let us focus our attention on the projection

$$\begin{aligned} &\big(q_{i(n_0)}(\gamma^*(s)) - q_{j(n_0)}(\gamma^*(s)),\, v_{i(n_0)}(\gamma^*(s)) - v_{j(n_0)}(\gamma^*(s))\big) \\ &= (\overline{q}(s),\, \overline{v}(s)) \in \mathbb{R}^2 \times \mathbb{R}^2, \end{aligned} \tag{3.2}$$

and on the lines

$$L(s) := \big\{\overline{q}(s) + t \cdot \overline{v}(s) \big|\, t \in \mathbb{R}\big\} \subset \mathbb{R}^2. \tag{3.3}$$

The following proposition directly follows from the definition (3.1) of w_0 and from the definition of the curve $\gamma_0 \subset \mathcal{SR}_0^+$.

Proposition 3.4. *The lines $L(s)$ rotate about a point A of $\mathbb{R}^2$ in Case A, whereas they are tangential to a given ellipse of $\mathbb{R}^2$ in Case B.*

We remind the reader that, according to Proposition 2.12, the vectors

$$r(\gamma_0(s)) = \big(\overline{v}(s),\, \overline{v}^+(s)\big)$$

belong to a given hyperplane $H(\gamma_0(0)) = H(y^*)$ of $\mathbb{R}^4$ not depending on the parameter s. Here

$$\overline{v}^+(s) := v^+_{i(n_0)}(\gamma_0(s)) - v^+_{j(n_0)}(\gamma_0(s)) \tag{3.5}$$

denotes the outgoing $(i(n_0), j(n_0))$ relative velocity right after the collision $e_{n_0} = (i(n_0), j(n_0))$.

The proof of the Theorem will be complete as soon as we prove our

Proposition 3.6. *Let $C_1 \subset \mathbb{R}^2$ be an ellipse, possibly degenerated to a single point, $C_2 \subset \mathbb{R}^2$ be a circle, so that C_1 and C_2 are lying outside of each other. Suppose that $L(s)$, $|s| < \varepsilon_0$, is a smooth family of lines in $\mathbb{R}^2$ with the following properties:*

(*i*) $L(s)$ *is tangent to* C_1 *at the point of contact* $A(s)$,

(*ii*) $L(s)$ *intersects* C_2 *in two points, out of which the one lying closer to* $A(s)$ *is denoted by* $B(s)$,

(*iii*) $\frac{d}{ds}\alpha\left(B(s)-A(s)\right)>0$ *for all* s, $|s|<\varepsilon_0$.

Here $\alpha\left(B(s)-A(s)\right)=\alpha(\overline{v}(s))$ *denotes the direction angle of the vector* $B(s)-A(s)=\overline{v}(s)$. *Finally, let* $\overline{v}^+(s)$ *be the mirror image of* $\overline{v}(s)$ *under the orthogonal reflection across the tangent line of the circle* C_2 *at the point* $B(s)$.

We claim that there is no hyperplane $H\subset\mathbb{R}^2\times\mathbb{R}^2$ *containing all the points* $(\overline{v}(s),\,\overline{v}^+(s))$ *for* $|s|<\varepsilon_0$.

Proof. A simple geometric inspection. Indeed, first one can assume that lines $L(s)$ depend on the parameter s analytically. Then one can analytically extend the family of lines $L(s)$ to a maximal interval of parameters $I=[a,b]\supset(-\varepsilon_0,\varepsilon_0)$ by preserving all properties (i)–(iii) above so that $L(a)$ and $L(b)$ are tangential to the circle C_2. If there was a hyperplane $H\subset\mathbb{R}^2\times\mathbb{R}^2$ containing all points $(\overline{v}(s),\,\overline{v}^+(s))$ for $|s|<\varepsilon_0$ then, by the reason of analyticity, the same containment $(\overline{v}(s),\,\overline{v}^+(s))\in H$ would be true for all s, $a\le s\le b$. Let $s_0\in(a,b)$ be such that $\overline{v}^+(s_0)=-\overline{v}(s_0)$. Now we have that

$$\begin{aligned}&(\overline{v}(a),\,\overline{v}(a))\in H,\\&(\overline{v}(b),\,\overline{v}(b))\in H,\\&(\overline{v}(s_0),\,-\overline{v}(s_0))\in H,\end{aligned}$$

so the hyperplane $H\subset\mathbb{R}^2\times\mathbb{R}^2$ must coincide with the hyperplane

$$\left\{(x,y)\in\mathbb{R}^2\times\mathbb{R}^2\big|\quad (y-x)\parallel\overline{v}(s_0)\right\}.$$

But this is impossible, for the difference vectors $\overline{v}^+(s)-\overline{v}(s)$ are obviously not parallel to $\overline{v}(s_0)$ for any s, unless $s=s_0$, or $s=a$, or $s=b$. □

4. Proof of the Boltzmann-Sinai Ergodic Hypothesis for all hard disk systems

Proof. We carry out an induction on the number N of elastically interacting disks. For $N=2$ this is the classic result of Sinai, [Sin(1970)]. Suppose that $N>2$ and the result (ergodicity, the Chernov-Sinai Ansatz, and complete hyperbolicity, implying the Bernoulli mixing property, see [C-H(1996)] and [O-W(1998)]) has been proved for all systems of hard disks (of equal masses) on the flat 2-torus $\mathbb{T}^2$ with the number of disks less than N. According to Theorem 6.1 of [Sim(1992)-I], for almost every singular phase point $x\in\mathcal{SR}_0^+$ the forward orbit $S^{(0,\infty)}x$ of x

(1) contains no singularity, and
(2) contains infinitely many connected collision graphs following each other in time.

By Corollary 3.26 of [Sim(2002)] such forward orbits $S^{(0,\infty)}x$ are sufficient (geometrically hyperbolic), unless the phase point x belongs to a countable family $J_1, J_2, \dots$ of exceptional, codimension-one, nonhyperbolicity manifolds studied right here in this paper. By our Theorem all these exceptional manifolds J_k intersect $\mathcal{SR}_0^+$ in zero-measured subsets of $\mathcal{SR}_0^+$, and this proves the Chernov-Sinai Ansatz for our current system with N disks. Finally, the Theorem of [Sim(2009)] gives us that the considered N-disk system is also ergodic, completely hyperbolic, hence Bernoulli mixing. □

References

[B-S(1973)] Bunimovich, L. A., Sinai, Ya. G.: *On the fundamental theorem of dispersing billiards*, Math. Sb. **90** (1973), 415–431.

[C-H(1996)] Chernov, N. I., Haskell, C.: *Nonuniformly hyperbolic K-systems are Bernoulli*, Ergod. Th. & Dynam. Sys. **16** (1996), 19–44.

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

[K-S-Sz(1991)] Krámli, A., Simányi, N., Szász, D.: *The K–Property of Three Billiard Balls*, Annals of Mathematics **133** (1991), 37–72.

[K-S-Sz(1992)] Krámli, A., Simányi, N., Szász, D.: *The K–Property of Four Billiard Balls*, Commun. Math. Phys. **144** (1992), 107-148.

[O-W(1998)] Ornstein, D., Weiss, B.: *On the Bernoulli Nature of Systems with Some Hyperbolic Structure*, Ergod. Th. & Dynam. Sys. **18** (1998), 441–456.

[S-Ch(1987)] Sinai, Ya. G., Chernov, N. I.: *Ergodic properties of certain systems of 2–D discs and 3–D balls*, Russian Math. Surveys **42:3** (1987), 181-207.

[Sim(1992)-I] Simányi, N.: *The K-property of N billiard balls I*, Invent. Math. **108** (1992), 521–548.

[Sim(1992)-II] Simányi, N.: *The K-property of N billiard balls II*, Invent. Math. **110** (1992), 151–172.

[Sim(2002)] Simányi, N.: *The Complete Hyperbolicity of Cylindric Billiards*, Ergod. Th. & Dynam. Sys. **22** (2002), 281–302.

[Sim(2003)] Simányi, N.: *Proof of the Boltzmann-Sinai Ergodic Hypothesis for Typical Hard Disk Systems*, Invent. Math. **154:1** (2003), 123-178.

[Sim(2004)] Simányi, N.: *Proof of the Ergodic Hypothesis for Typical Hard Ball Systems*, Annales Henri Poincaré **5** (2004), 203-233.

[Sim(2009)] Simányi, N.: *Conditional proof of the Boltzmann-Sinai ergodic hypothesis*, Invent. Math. **177** (2009), 381–413.

[Sin(1963)] Sinai, Ya. G.: *On the Foundation of the Ergodic Hypothesis for a Dynamical System of Statistical Mechanics*, Soviet Math. Dokl. **4** (1963), 1818-1822.

[Sin(1970)] Sinai, Ya. G.: *Dynamical Systems with Elastic Reflections*, Russian Math. Surveys **25:2** (1970), 137–189.

[S-Sz(1999)] Simányi, N., Szász, D.: *Hard ball systems are completely hyperbolic*, Annals of Mathematics, **149** (1999), 35–96.

[S-Sz(2000)] Simányi, N., Szász, D.: *Non-integrability of Cylindric Billiards and Transitive Lie Group Actions*, Ergod. Th. & Dynam. Sys. **20** (2000), 593–610.

The University of Alabama at Birmingham, Department of Mathematics, 1300 University Blvd., Suite 452, Birmingham, AL 35294 U.S.A.

E-mail address: simanyi@uab.edu