Singularities and non-hyperbolic manifolds do not coincide

2013 Nonlinearity 26 1703

(http://iopscience.iop.org/0951-7715/26/6/1703)

IOP Publishing — Nonlinearity
Nonlinearity **26** (2013) 1703–1717 doi:10.1088/0951-7715/26/6/1703

Singularities and non-hyperbolic manifolds do not coincide

Nándor Simányi

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

E-mail: simanyi@uab.edu

Received 1 May 2012, in final form 24 April 2013
Published 9 May 2013
Online at stacks.iop.org/Non/26/1703

Recommended by C Liverani

Abstract
We consider the billiard flow of elastically colliding hard balls on the flat ν-torus ($\nu \geqslant 2$), and prove that no singularity manifold can even 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, i.e. the verification of the Boltzmann–Sinai ergodic hypothesis.

Mathematics Subject Classification: 37D50, 34D05

1. Introduction

In this paper we prove the Boltzmann–Sinai ergodic hypothesis for hard ball systems on the ν-torus $\mathbb{R}^\nu/\mathbb{Z}^\nu$ ($\nu \geqslant 2$) 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 sections of [Sim(2009)].

In a loose form, as attributed to L Boltzmann back in the 1880s, the Boltzmann hypothesis asserts that gases of hard balls are ergodic. In a precise form, which is due to Sinai in [Sin(1963)], it states that the gas of $N \geqslant 2$ identical hard balls (of 'not too big' radius) on a torus $\mathbb{T}^\nu = \mathbb{R}^\nu/\mathbb{Z}^\nu$, $\nu \geqslant 2$, (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 Chernov and Sinai extended these results to ($N = 2$, $\nu \geqslant 2$), 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 weak hyperbolicity.

Further development has been due mostly to Krámli, Szász, and the present author. We proved the 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 analysing all possible degeneracies in the dynamics to obtain 'global' ergodicity. We extended our results to $N = 4$ balls in dimension $\nu \geqslant 3$ next year [K-S-Sz(1992)], and then I proved the ergodicity whenever $N \leqslant \nu$ in [Sim(1992)-I, 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 forward was made by D Szász and myself, when we used the methods of algebraic geometry in [S-Sz(1999)]. We assumed that the balls had arbitrary masses $m_1, \ldots, m_N$ (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 analyse all possible degeneracies, but only for typical (generic) $(N + 1)$-tuples of 'external' parameters $(m_1, \ldots, m_N, r)$; 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 $N \geqslant 2$ and $\nu \geqslant 2$, and for generic $(m_1, \ldots, m_N, r)$, 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 $(m_1, \ldots, m_N, r)$ [Sim(2002)]. The reason why the masses m_i are considered *geometric parameters* is that they determine the relevant Riemannian metric

$$||\mathrm{d}q||^2 = \sum_{i=1}^{N} m_i ||\mathrm{d}q_i||^2$$

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

To upgrade the complete 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 $\geqslant 1$ in the phase space. For ergodicity, one has to show that its codimension is $\geqslant 2$, or find some other ways to prove that the (possibly) arising codimension-one manifolds of non-sufficiency are not capable 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 the systems of $N \geqslant 2$ balls on a two-dimensional torus are ergodic for typical (generic) $(N + 1)$-tuples of external parameters $(m_1, \ldots, m_N, r)$. The proof again involves some algebro-geometric techniques, thus the result is restricted to generic parameters $(m_1, \ldots, m_N; r)$. But there was a good reason to believe that systems in $\nu \geqslant 3$ 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 $N \geqslant 2$, $\nu \geqslant 2$, and for almost every selection $(m_1, \ldots, m_N; r)$ of the external geometric parameters the corresponding system of N hard balls on $\mathbb{T}^\nu$ is (completely 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)].

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 describing 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.

2. Formulation and proof of the theorem

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

(a) $S^T(U_0) \cap \partial \boldsymbol{M} = \emptyset$,
(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 in the increasing time order, and let the corresponding advances be $\alpha_i = \alpha(e_i)$, $i = 1, 2, \dots, n$. Let $t_i = t(e_i)$ be the time of the ith 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 section 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 a proof by contradiction, so from now on we assume that $J = K$, and the proof 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 and homogeneous) behavior. We note that this can be done, thanks to the algebraic nature of the dynamics.

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 $(\Phi(w))_i = \alpha_i(w)$ ($i = 1, 2, \dots, n$) 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 image $\Phi(\mathcal{N}_0(S^{[0,T]}x))$ will be denoted by $\overline{\mathcal{N}}_0(S^{[0,T]}x)$. 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, let $1 = k(1) < k(2) < \ldots < k(N-1) < n$ be the uniquely defined indices with the property that for every l ($1 \leqslant l \leqslant N-1$) the collision graph $(\mathcal{V}, \{e_1, e_2, \ldots, e_{k(l)}\})$ has exactly $N-l$ connected components, whereas the number of components of $(\mathcal{V}, \{e_1, e_2, \ldots, e_{k(l)-1}\})$ is $N-l+1$.

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

For every *non-essential edge* $e_m = \{i(m), j(m)\}$ $(1 \leqslant i(m) < j(m) \leqslant N)$ 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 $e_1, e_2, \ldots, e_{m-1}$ (with coefficients made up from some masses and advances) precisely as described by the CPF, see proposition 2.19 in [S-Sz(1999)]:

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

($1 \leqslant m \leqslant n$, e_m is not essential), where each $\Gamma^{(m)}_k$ 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 $\Phi(\mathcal{N}_0(S^{[0,T]}x)) = \overline{\mathcal{N}}_0(S^{[0,T]}x)$, 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 non-hyperbolicity 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 $\Sigma = (e_1, e_2, \ldots, e_n)$) as an initial segment.)

Throughout the proof we shall assume that the masses of the elastically interacting balls are equal: $m_1 = m_2 = \ldots = m_N$. 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 $d = \nu(N-1)$ the dimension of the configuration space $\boldsymbol{Q}$.

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

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

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

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

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

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 $e_k = (i(k), j(k))$ $(1 \leqslant i(k) < j(k) \leqslant N)$, 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 neighbourhood $B_0 \subset M_n$ of the base point $(V_0(x_0);\ g_1(x_0), \dots, g_n(x_0))$. We begin with a useful definition.

Definition 2.6. The projections $R_k : \mathcal{Z} \to \mathbb{R}^\nu$ $(k = 1, 2, \dots, n)$ are defined by the equation

$$R_k(\delta q) = \delta q_{i(k)} - \delta q_{j(k)}$$

for $\delta q \in \mathcal{Z}$, where $\mathcal{Z}$ is the tangent space of $\boldsymbol{Q}$ in (2.3).

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

Proposition 2.7. *For any integer m, $2 \leqslant m \leqslant n$, the neutral space*

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

is determined by the directions *of all relative velocities $v^-_{i(l)} - v^-_{j(l)}$, $v^+_{i(l)} - v^+_{j(l)}$ $(2 \leqslant l \leqslant m-1)$, and by the directions of $v^+_{i(1)} - v^+_{j(1)}$ and $v^-_{i(m)} - v^-_{j(m)}$.* This property will be called the direction determination principle (DDP). *As a consequence, all the neutral spaces*

$$\mathcal{N}_k = \mathcal{N}_k(V_0;\ g_1, g_2, \dots, g_m) = \mathcal{N}_{t_k+0}(V_0;\ g_1, g_2, \dots, g_m)$$

$(0 \leqslant k \leqslant m)$ are determined by the relative velocities listed above and by $v^-_{i(1)} - v^-_{j(1)}$, $v^+_{i(m)} - v^+_{j(m)}$. We note that the neutral spaces $\mathcal{N}_k$ are connected to each other via the equations $\mathcal{N}_l = \mathcal{N}_k \cdot g_{k+1} \cdot \dots \cdot g_l$ for $k < l$, and the reflection g_s is (locally) determined by the directions of the relative velocities $v^-_{i(s)} - v^-_{j(s)}$ and $v^+_{i(s)} - v^+_{j(s)}$, $1 \leqslant s \leqslant m$.

Proof. Observe that for any tangent vector $\delta q = (\delta q_1, \dots, \delta q_N) \in \mathcal{Z}$ the relation $\delta q \in \mathcal{N}_1(V_0;\ g_1, \dots, g_m)$ holds true if and only if for every k, $2 \leqslant k \leqslant m$, 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 $v^+_{i(1)} - v^+_{j(1)}$. □

Proposition 2.8. *Use the notions and notations of the previous proposition, except that here we allow the values* 0 *and* 1 *for the number of collisions m. We claim that for fixed* directions *of all relative velocities $v^-_{i(k)} - v^-_{j(k)}$ and $v^+_{i(k)} - v^+_{j(k)}$ $(k = 1, 2, \dots, m)$ and for a given reference time $t_s + 0$ $(0 \leqslant s \leqslant m,\ t_0 = 0)$ all possible space and velocity variations δq and δv are precisely the elements of the neutral space $\mathcal{N}_s(V_0;\ g_1, g_2, \dots, g_m)$.*

Proof. Induction on m. The statement is obviously true for $m = 0$, since in this case $\mathcal{N}_0(V_0;\ g_1, g_2, \dots, g_m) = \mathcal{Z}$, the tangent space of the configuration space.

Assume now that $m \geqslant 1$ and the claim is true for all smaller numbers of collisions. Clearly it is enough to prove the proposition for the case $s = m-1$. The fixed directions of $v^-_{i(k)} - v^-_{j(k)}$ and $v^+_{i(k)} - v^+_{j(k)}$ for $k = 1, 2, \ldots, m-1$ mean that the possible values of either δq or δv are precisely the elements of the neutral space $\mathcal{N}_{m-1}(V_0;\ g_1, g_2, \ldots, g_{m-1})$. If, in addition, we also fix the direction of $v^-_{i(m)} - v^-_{j(m)}$, then this leaves for us the space $\mathcal{N}_{m-1}(V_0;\ g_1, g_2, \ldots, g_m)$ as the set of all available values for δv. Furthermore, by also fixing the direction of $v^+_{i(m)} - v^+_{j(m)}$ (i.e. also fixing the reflection g_m) restricts the space of available values for δq to the neutral space $\mathcal{N}_{m-1}(V_0;\ g_1, g_2, \ldots, g_m)$. □

Proposition 2.9. *For every m, $1 \leqslant m \leqslant n$, the generic ($\Longleftrightarrow$minimal) dimension (both in measure-theoretical and topological senses) of the neutral spaces*

$$\mathcal{N}_0(V_0;\ g_1, \ldots, g_m)$$

on the phase space M_m is equal to the generic ($\Longleftrightarrow$minimal) value of

$$\dim \mathcal{N}_0(V_0(x);\ g_1(x), g_2(x), \ldots, g_m(x))$$

for all $x \in U_0$. (key lemma 3.19 in [Sim(2002)].)

The value of this typical dimension will be denoted by $\Delta(e_1, e_2, \ldots, e_m)$. Plainly, it only depends on the symbolic sequence $(e_1, e_2, \ldots, e_m)$.

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

$$\dim \mathcal{N}_0\left(V_0(y);\ g_1(y), \ldots, g_m(y)\right) = \Delta(e_1, e_2, \ldots, e_m) \quad \forall y \in B_0 \setminus \tilde{J}, \tag{2.10}$$

$$\dim \mathcal{N}_0\left(V_0(y);\ g_1(y), \ldots, g_m(y)\right) = \Delta_J(e_1, e_2, \ldots, e_m) \quad \forall y \in \tilde{J}, \tag{2.11}$$

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

Proposition 2.12. *(A corollary of the proof of key lemma 3.19 of [Sim(2002)].) Let $1 \leqslant m \leqslant n$, and $\mathcal{N}^* \subset \mathcal{Z}$ be a given subspace with $\mathcal{N}^* \cap \{V_m(x) \big| \ x \in U_0\} \neq \emptyset$. We claim that the typical (i.e. minimal) value of*

$$\dim \left[\mathcal{N}^* \cap \mathcal{N}_m(V_m;\ g_{m+1}, g_{m+2}, \ldots, g_n)\right]$$

for $V_m \in \mathcal{N}^*$ and $g_k \in \mathcal{P}_k$ ($m+1 \leqslant k \leqslant n$) is equal to the typical (i.e. minimal) value of

$$\dim \left[\mathcal{N}^* \cap \mathcal{N}_m(V_m(x);\ g_{m+1}(x), g_{m+2}(x), \ldots, g_n(x))\right]$$

for $x \in U_0$ with $V_m(x) \in \mathcal{N}^*$.

Proof. The proof of this statement can be obtained from the proof of Key lemma 3.19 of [Sim(2002)], hence it is omitted. □

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

Definition 2.13. The 'critical index' n_0 is the unique positive integer n_0, $1 \leqslant n_0 \leqslant n$, with the property that for any $x \in U_0$

(i) the directions of the relative velocities $v^-_{i(k)}(x) - v^-_{j(k)}(x)$, $v^+_{i(k)}(x) - v^+_{j(k)}(x)$, $k = 1, 2, \ldots, n_0$, determine in M_n if $\Pi(x) \in \tilde{J}$, whereas

(ii) the directions of the relative velocities $v^-_{i(k)}(x) - v^-_{j(k)}(x)$, $v^+_{i(k)}(x) - v^+_{j(k)}(x)$, $k = 1, 2, \ldots, n_0 - 1$, do not determine yet in M_n if $\Pi(x) \in \tilde{J}$.

The precise meaning of the notions above is the following: the manifolds $\mathcal{W}_{n_0} = \mathcal{W}_{n_0}(x) \subset U_0$ that are defined by fixing the directions of all the relative velocities listed in (i) (which form a smooth foliation of the local neighbourhood U_0 if U_0 is chosen small enough) are either subsets of J or they are disjoint from it, whereas the manifolds $\mathcal{W}_{n_0-1} = \mathcal{W}_{n_0-1}(x)$ that are defined by fixing the directions of all the relative velocities listed in (ii) (which also form a smooth foliation of the local neighbourhood U_0 for small enough U_0) are transversal to J.

Apply proposition 2.12 to $m = n_0$,

$$\mathcal{N}^* = \mathcal{N}_m\left(V_0(x);\ g_1(x), g_2(x), \ldots, g_{n_0}(x)\right)$$

($x \in U_0$) to realize that the directions of the relative velocities listed above in (i) also determine if the phase point $(V_0;\ g_1, g_2, \ldots, g_n) \in M_n$ belongs to $\tilde{J}$ or not. In the free velocity process $(V_0;\ g_1, g_2, \ldots, g_n) \in M_n$ there is absolutely no constraint on the velocities, other than that each g_k is an orthogonal reflection across a hyperplane determined by $e_k = (i(k),\ j(k))$. Because of this, the only way that the relative velocities listed above in (i) determine the status of $(V_0;\ g_1, \ldots, g_n) \in \tilde{J}$ is that a minor $\mathcal{M}$ (determinant of a square submatrix) of the system (2.2) with maximum column index n_0 vanishes. Observe that the n_0-th column of the system of CPFs (2.2), i.e. the coefficients of the unknown α_{n_0} in (2.2), depend on the pair of velocities

$$r(x) = \left(v^-_{i(n_0)}(x) - v^-_{j(n_0)}(x),\ v^+_{i(n_0)}(x) - v^+_{j(n_0)}(x)\right)$$

linearly (they are certain linear combinations of some coordinates of the two components of $r(x)$), hence the minor $\mathcal{M}$ also depends linearly on $r(x)$, and $(V_0;\ g_1, \ldots, g_n) \in \tilde{J}$ means that the solution set of (2.2) is atypically big. Using these two observations and the DDP of proposition 2.7 we obtain a useful description of the membership relation $x \in J$ as follows.

Proposition 2.14. *For any $x \in U_0$ the relation $x \in J$ holds true if and only if the pair of relative velocities*

$$r(x) := \left(v^-_{i(n_0)}(x) - v^-_{j(n_0)}(x),\ v^+_{i(n_0)}(x) - v^+_{j(n_0)}(x)\right) \in \mathbb{R}^\nu \times \mathbb{R}^\nu = \mathbb{R}^{2\nu} \tag{2.15}$$

belongs to a hyperplane $H(x) \subset \mathbb{R}^{2\nu}$ depending analytically on the directions

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

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

In order to make the mechanism discussed in proposition 2.14 more transparent, below we provide the reader with a brief analysis of the special example $\Sigma = (e_1,\ e_2,\ e_3)$ with $e_1 = (1, 2)$, $e_2 = (1, 3)$, and $e_3 = (2, 3)$. Since the relevant observation times for this sequence are t_1 and t_2 separating the first two and the second and third collisions, respectively, in this example we will consequently denote the velocities and space perturbations observed at time t_1 with a superscript $-$, whereas the velocities and space perturbations observed at time t_2 will be distinguished by a superscript $+$. (This is somewhat in contrast with the earlier notations, but here they come rather handy.)

The neutrality equations with respect to e_1 and e_2, along with the preservation of the center of mass are

$$\alpha_1(v_1^- - v_2^-) = \delta q_1^- - \delta q_2^-,$$

$$\alpha_2(v_1^- - v_3^-) = \delta q_1^- - \delta q_3^-,$$

$$\delta q_1^- + \delta q_2^- + \delta q_3^- = 0.$$

From these equations we immediately get

$$\delta q_1^- = \tfrac{1}{3}\alpha_1(v_1^- - v_2^-) + \tfrac{1}{3}\alpha_2(v_1^- - v_3^-),$$

$$\delta q_2^- = -\tfrac{2}{3}\alpha_1(v_1^- - v_2^-) + \tfrac{1}{3}\alpha_2(v_1^- - v_3^-),$$
$$\delta q_3^- = \tfrac{1}{3}\alpha_1(v_1^- - v_2^-) - \tfrac{2}{3}\alpha_2(v_1^- - v_3^-).$$

Using the transformation equations through e_2 and the neutrality with respect to this collision

$$\delta q_2^+ = \delta q_2^-,$$
$$\delta q_1^- + \delta q_3^- = \delta q_1^+ + \delta q_3^+,$$
$$(\delta q_1^+ - \delta q_3^+) - (\delta q_1^- - \delta q_3^-) = \alpha_2\left[(v_1^+ - v_3^+) - (v_1^- - v_3^-)\right]$$

one easily expresses the quantity $\delta q_2^+ - \delta q_3^+$ as follows:

$$\delta q_2^+ - \delta q_3^+ = -\alpha_1(v_1^- - v_2^-) + \tfrac{1}{2}\alpha_2\left[(v_1^- - v_3^-) + (v_1^+ - v_3^+)\right].$$

Note that the linear coordinates α_1 and α_2 independently parametrize the two-dimensional neutral space $\mathcal{N}_0(x;\ e_1, e_2)$. From the last equation we see that the non-hyperbolicity $x \in J$ holds true precisely when the vectors $v_1^- - v_2^-$ and $(v_1^- - v_3^-) + (v_1^+ - v_3^+)$ are parellel. This parallelity condition defines a subspace H for the vector $r(x) = (v_1^- - v_3^-,\ v_1^+ - v_3^+)$ with codimension $\nu - 1$, which codimension is 1 exactly when $\nu = 2$. (In the case $\nu \geqslant 3$ there is nothing to prove; the codimension is already big enough.)

The next result tells us that the collision e_{n_0} decreases the dimension of the neutral space.

Lemma 2.16.

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

Proof. Proof by contradiction: assume that $\Delta(e_1, \ldots, e_{n_0}) = \Delta(e_1, \ldots, e_{n_0-1})$. This assumption 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_0th column from the system of CPFs (2.2). This shows that the two relative velocity components of $r(x)$ in (2.15) have no effect on the solution set in question, and this contradicts to the properties (i)–(ii) of the critical index n_0 listed in definition 2.13. □

The upcoming lemma tells us that the critical collision e_{n_0} does not distinguish between the points of J and of $U_0 \setminus J$.

Lemma 2.17.

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

Proof. Again a proof by contradiction: assume that $\Delta(e_1, \ldots, e_{n_0}) < \Delta_J(e_1, \ldots, e_{n_0})$. According to proposition 2.7, the neutral space

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

is determined by the directions of the relative velocities $v_{i(l)}^-(x) - v_{j(l)}^-(x)$ and $v_{i(l)}^+(x) - v_{j(l)}^+(x)$ for $l = 1, 2, \ldots, n_0 - 1$, whereas, according to (ii) of definition 2.13, these relative velocities do not determine whether $x \in J$. On the other hand, the projection

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

of this neutral space onto $\delta q_{i(n_0)} - \delta q_{j(n_0)}$ determines if $x \in J$ is true or not. To see this we note that, due to the assumption $\Delta(e_1, \ldots, e_{n_0}) < \Delta_J(e_1, \ldots, e_{n_0})$, for the points $x \in U_0 \setminus J$ the dimension of

$$R_{n_0}\left[\mathcal{N}_{n_0-1}\left(V_0(x);\ g_1(x), g_2(x), \ldots, g_{n_0-1}(x)\right)\right]$$

(which is

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

is larger than the similar dimension for the points $x\in J$. This, in turn, means that the directions of the relative velocities $v^-_{i(l)}(x)-v^-_{j(l)}(x)$ and $v^+_{i(l)}(x)-v^+_{j(l)}(x)$ $(l=1,2,\ldots,n_0-1)$ determine if $x\in J$ is true or not, thus violating property (ii) of n_0 listed in definition 2.13. □

3. Finishing the proof of the theorem

First we present the closing part of the proof by assuming that $\nu=2$. We remind the reader that the entire proof of the theorem is a proof by contradiction, so the coincidence (in a neighbourhood U_0) of J and the past-singularity K is assumed all along. Right after that we present the proof for the case $\nu\geqslant 3$, which is just slightly more difficult technically than the case $\nu=2$. Thus, for now we assume that $\nu=2$.

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

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

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

This is possible, due to lemmas 2.16–2.17. 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) $v(\gamma_0(s))=\frac{v(y^*)+s\cdot w_0}{||v(y^*)+s\cdot w_0||}$,
(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) $v(\gamma_0(s))=\frac{v(y^*)+s\cdot w_0}{||v(y^*)+s\cdot w_0||}$,
(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

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

and on the lines

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

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.*

Remark 3.5. We should note here that there is an exceptional subcase of case B when the ellipse also degenerates to a point, just like in case A. This is the situation when the singularity at y^* is a tangency but the projection $R_0(w_0)$ is parallel to the outgoing relative velocity $v^+_{i(0)} - v^+_{j(0)}$ of the two particles $i(0)$ and $j(0)$ colliding tangentially at time zero. However, this degeneracy of the ellipse does not cause any problem in the proof, for it is treated as the degeneracy in case A.

We also note that in all of the cases above the directions of the lines $L(s)$ are properly changing at a non-zero rate, thanks to our choice of w_0 with

$$w_0 \notin \mathcal{N}_0\left(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0}(y^*)\right).$$

We remind the reader that, according to proposition 2.14, the vectors

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

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

$$\overline{v}^+(s) := v^+_{i(n_0)}(\gamma_0(s)) - v^+_{j(n_0)}(\gamma_0(s)) \tag{3.6}$$

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 reason why the hyperplanes $H(\gamma_0(s))$ are independent of s is the following: Both the space and velocity perturbations $q(\gamma_0(s)) - q(\gamma_0(0))$ and $v(\gamma_0(s)) - v(\gamma_0(0))$ belong to the neutral space $\mathcal{N}_0(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*))$ and, furthermore, they are proportional to each other. One proves by a standard 'continuous induction' that these properties remain true all the way until time t^*, thus the (incoming and outgoing) relative velocities of the collisions $g_1, g_2, \ldots, g_{n_0-1}$ are independent of the perturbation parameter s.

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

Proposition 3.7. *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 none of C_1 or C_2 is lying inside the other one, i.e. they have at least two common tangent lines. Suppose that $L(s)$, $|s| < \varepsilon_0$, is a smooth family of oriented lines in $\mathbb{R}^2$ with the direction vector $\overline{v}(s)$ satisfying the following conditions:*

- (i) *$L(s)$ is tangent to C_1 at the point of contact $A(s)$, and at $A(s)$ the direction vector $\overline{v}(s)$ agrees with a given orientation of C_1 (if C_1 is not a point),*
- (ii) *$L(s)$ intersects C_2 in two points, out of which the one whose position vector makes the smaller inner product with $\overline{v}(s)$ is denoted by $B(s)$,*
- (iii) *$\frac{\mathrm{d}}{\mathrm{d}s}\alpha(\overline{v}(s)) > 0$ for all s, $|s| < \varepsilon_0$.*

Here $\alpha(\overline{v}(s))$ denotes the direction angle of the vector $\overline{v}(s)$. 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. We can assume, without restricting generality, that $\|\overline{v}(s)\| = 1$. We prove the proposition in the case when C_1 and C_2 have at least two common, non-parallel tangent lines. The proof for the exceptional case, when this hypothesis is not satisfied, can be done with some modifications, which we will show below right after completing the proof by using the hypothesis.

First of all, we can assume that the lines $L(s)$ depend on the parameter s analytically. Then one can analytically extend the family of lines $L(s)$ to an 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 non-parallel and tangent 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, $a \leqslant s \leqslant b$. Now we have that

$$(\overline{v}(a),\ \overline{v}(a)) \in H,$$
$$(\overline{v}(b),\ \overline{v}(b)) \in H,$$

so H contains the diagonal $\{(x, x) \big|\ x \in \mathbb{R}^2\}$ and, consequently, the difference vectors $x - y$ for all $(x, y) \in H$ are parallel to each other. But this is impossible, for the difference vectors $\overline{v}^+(s) - \overline{v}(s)$ can obviously rotate as s varies in the parameter interval.

Finally, we show how to proceed in the case when $\overline{v}(a)$ and $\overline{v}(b)$ are parallel, i.e. $\overline{v}(b) = -\overline{v}(a)$. We assume, contrary to the claim of the proposition, that there exists a hyperplane $H \subset \mathbb{R}^2 \times \mathbb{R}^2$ containing all vectors $(\overline{v}(s),\ \overline{v}^+(s))$, $s \in I$. We take the limit

$$\lim_{s \to a^+} (s - a)^{-1/2} \left[(\overline{v}(s),\ \overline{v}^+(s)) - (\overline{v}(a),\ \overline{v}^+(a))\right] = (0,\ \xi) \in H,$$

where $\xi \in \mathbb{R}^2$, $\xi \neq 0$, $\xi \perp \overline{v}(a)$. This shows that for every $s \in I$ the vector

$$\eta(s) = \overline{v}(s) - \langle \overline{v}^+(s), \overline{v}(a)\rangle \cdot \overline{v}(a)$$

has the property that $(\eta(s), 0) \in H$. The vectors $\eta(s)$ $(s \in I)$ must be mutually parallel, otherwise the three-dimensional subspace H of $\mathbb{R}^2 \times \mathbb{R}^2$ would be equal to $\mathbb{R}^2 \times \langle \xi \rangle$, which would mean that all outgoing vectors $\overline{v}^+(s)$ are parallel to ξ, but this is clearly not the case.

Denote the common line containing all the vectors $\eta(s)$ by $\mathcal{L}$. Clearly $\mathcal{L}$ is not parallel to the vector $\overline{v}(a)$. We claim that $\mathcal{L} \perp \overline{v}(a)$. Indeed, the vectors $\overline{v}(s)$, $s \in I$, fill out one half of the unit circle, thus in the case $\mathcal{L} \not\perp \overline{v}(a)$ there would be a parameter value s, $a < s < b$, such that $\mathrm{dist}(\overline{v}(s),\ \eta(s)) > 1$, which is impossible, for

$$\mathrm{dist}(\overline{v}(s),\ \eta(s)) = \left|\langle \overline{v}^+(s),\ \overline{v}(a)\rangle\right| \leqslant 1.$$

The fact $\mathcal{L} \perp \overline{v}(a)$, however, implies that $\langle \overline{v}^+(s) - \overline{v}(s),\ \overline{v}(a)\rangle = 0$ for all $s \in I$, which is clearly a contradiction, since the non-zero difference vectors $\overline{v}^+(s) - \overline{v}(s)$ are parallel to the rotating collision normal. □

Finally, we complete the proof of the theorem in the (somewhat more difficult) case $\nu \geqslant 3$, as follows.

We consider an arbitrary phase point $y_0 \in J$, select the time $\tau < 0$ and, correspondingly, the phase point $y^* = S^\tau y_0$ just as before. Furthermore, the selection of a suitable tangent vector w_0 of (3.1), the construction of the smooth curve $\gamma_0(s) \in \mathcal{SR}_0^+$ $(|s| < \varepsilon_0)$, the selection of the separating time t^*, the construction of the vectors

$$(\overline{q}(s),\ \overline{v}(s)) \in \mathbb{R}^\nu \times \mathbb{R}^\nu$$

of (3.2) and the construction of the lines

$$L(s) := \left\{\overline{q}(s) + t \cdot \overline{v}(s) \big|\ t \in \mathbb{R}\right\} \subset \mathbb{R}^\nu$$

of (3.3) are similar to what we did above in the case $\nu = 2$, but now we have to exercise more care in the selection of the neutral tangent vector w_0 of (3.1), see below.

Suppose, for a moment, that we have already chosen a suitable tangent vector

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

of (3.1).

The counterpart of proposition 3.4 is

Proposition 3.8. *All the lines* $L(s)$ $(|s| < \varepsilon_0)$ *lie in the same two-dimensional affine subspace* $\mathcal{P} = \mathcal{P}(y^*, w_0)$ *of* $\mathbb{R}^\nu$. *These lines rotate about a point A of* $\mathcal{P}$ *in Case A, whereas they are tangential to a given ellipse* C_1 *of* $\mathcal{P}$ *in case B.*

Remark 3.9. We note that here remark 3.5 again applies.

Consider the smallest linear subspace $S = S(y^*, w_0) \subset \mathbb{R}^\nu$ of $\mathbb{R}^\nu$ containing the affine plane $\mathcal{P}$. Clearly the dimension of S is 3 or 2. By algebraic reasons there are two possibilities: Either the space $S = S(y^*, w_0)$ is three-dimensional for a typical pair (y^*, w_0) $(y^* = S^\tau y_0 \in \mathcal{SR}_0^+,\ y_0 \in U_0,\ w_0 \in \mathcal{N}_0(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*)),\ w_0 \perp v(y_0))$, and in this situation we can assume that $\dim S(y^*, w_0) = 3$ *always* in our local analysis by choosing a small enough open set U_0, or $\dim S(y^*, w_0) = 2$ for *all* such considered pairs. The next lemma shows that the latter case is actually impossible.

Lemma 3.10. *It is not possible that* $\dim S(y^*, w_0) = 2$ *for every* $y^* \in \mathcal{SR}_0^+$ $(y^* = S^\tau y_0$, $y_0 \in U_0)$ *and for every*

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

$$w_0 \perp v(y^*).$$

Proof. By way of contradiction, assume that $\dim S(y^*, w_0) = 2$ is always the case. This means that the velocities of the phase points y^* can be rotated along the curves $\gamma_0(s) \subset \mathcal{SR}_0^+$ in such a way that we obtain an n_0th collision with a collision normal vector parallel to the relative velocity of the colliding particles $i(n_0)$ and $j(n_0)$. (A so called 'head-on collision'.) It is clear that the foliation of the manifold $\mathcal{SR}_0^+$ into the curves $\gamma_0(s)$ can be chosen to be smooth. Furthermore, in order to reach a head-on collision from a given phase point $y^* \in \mathcal{SR}_0^+ \cap U_0$ via the curve $\gamma_0(s)$ (with $\gamma_0(0) = y^*$) it may be necessary to leave the small-sized local neighbourhood U_0 in which we are working. During the perturbation along the curve $\gamma_0(s)$ the times $t_k = t(e_k)$ of the collisions e_k $(k = 1, 2, \ldots, n_0 - 1)$ also change, and this could change the symbolic collision structure of the considered orbit segments. To avoid this problem, during the considered perturbations along the curves $\gamma_0(s)$ we delete all hard core potentials of unduly arising new collisions, i.e. we allow two particles to freely overlap each other if they would produce a collision not in the prescribed symbolic sequence $(e_1, e_2, \ldots, e_{n_0-1})$. (A so called phantom dynamics.)

The above mean that the phase points $y^* \in \mathcal{SR}_0^+$ with head-on collisions e_{n_0} form a codimension-one submanifold inside $\mathcal{SR}_0^+$. However, this is impossible, since the singularity manifold $\mathcal{SR}_0^+$ can be smoothly foliated by convex, local orthogonal manifolds, see section 4 in [K-S-Sz(1990)], and this shows that the codimension in $\mathcal{SR}_0^+$ of the set of phase points y^* with a head-on collision e_{n_0} is $\nu - 1$, which is now at least 2, a contradiction. □

Therefore, we may and we shall assume that the phase point $y^* = S^\tau y_0 \in \mathcal{SR}_0^+$ is chosen (and fixed) in such a way that for the typical selection of w_0 in (3.1) it is true that $\dim S(y^*, w_0) = 3$.

It is clear that the vector

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

varies in the five-dimensional linear subspace

$$\mathcal{P}' \times S \subset \mathbb{R}^\nu \times \mathbb{R}^\nu$$

of $\mathbb{R}^{2\nu}$, where $\mathcal{P}' = \mathcal{P}'(y^*, w_0)$ is the two-dimensional linear subspace of $\mathbb{R}^\nu$ parallel to $\mathcal{P}$.

Let us focus on the hyperplane $H(\gamma_0(s)) = H(y^*)$ of $\mathbb{R}^{2\nu}$, defined as before. For the proof of the fact $H(\gamma_0(s)) = H(\gamma_0(0))$ please see the paragraph containing (3.6). The fact that the velocities $V_0(x), V_1(x), \ldots, V_{n_0-1}(x)$ do not determine if the relation $x \in J$ is true or not, has the following consequence.

Proposition 3.11. *For every singular phase point y^* the neutral vector w_0 of 3.1 can be chosen in such a way that the hyperplane $H(\gamma_0(s)) = H(y^*)$ does not contain the subspace $\mathcal{P}' \times S$, i.e.* $\dim[(\mathcal{P}' \times S) \cap H(y^*)] = 4$.

Remark 3.12. The propery $\dim[(\mathcal{P}' \times S) \cap H(y^*)] = 4$ is an open property and the system in which it is defined is algebraic, so either this property holds on an open set with full measure inside the singularity manifold (and then we can assume that it holds for every singular phase point in the local neighbourhood U_0 that is chosen suitably small), or this property holds nowhere on the singularity manifold. In the indirect proof below we will assume the latter.

Proof. A proof by contradiction. Assume that for every singular phase point y^* (in U_0) and for every choice

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

the set containment

$$\mathcal{P}'(w_0) \times S(w_0) \subset H = H(y^*) \subset \mathbb{R}^\nu \times \mathbb{R}^\nu$$

is true. (The phase point $y^* = S^\tau y_0$ is now fixed.) This means that

$$\bigcup_{w_0 \in \mathcal{N}_0} \mathcal{P}'(w_0) \times \{0\} \subset H, \tag{3.13}$$

and

$$\bigcup_{w_0 \in \mathcal{N}_0} \{0\} \times S(w_0) \subset H, \tag{3.14}$$

where $\mathcal{N}_0 = \mathcal{N}_0(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*))$. Here is now the key observation: If we fix the manifold $\mathcal{W}_{n_0-1}(y^*)$, that is, all the *directions* of all the relative velocities $v^-_{i(k)} - v^-_{j(k)}$ and $v^+_{i(k)} - v^+_{j(k)}$ $(1 \leqslant k \leqslant n_0 - 1)$ for a phase point $y^* \in U_0$ and let all the other data vary then, according to propositions 2.7 and 2.8, we also fix the neutral space $\mathcal{N}_0 = \mathcal{N}_0(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*))$, and at any time t^* between t_{n_0-1} and t_{n_0} the data δq and δv vary in the neutral space $\mathcal{N}_{n_0-1}(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*))$, which is also determined by the manifold $\mathcal{W}_{n_0-1}(y^*)$, see again proposition 2.7. Therefore, the set containment relations (3.13)–(3.14) mean that

$$R_{n_0}\left[\mathcal{N}_{n_0-1}\left(V_0(x);\ g_1(x), \ldots, g_{n_0-1}(x)\right)\right] \\ \times \operatorname{span}\left\{q_{i(n_0)} - q_{j(n_0)},\ R_{n_0}\left[\mathcal{N}_{n_0-1}\left(V_0(x);\ g_1(x), \ldots, g_{n_0-1}(x)\right)\right]\right\} \subset H. \tag{3.15}$$

for any phase point $x \in U_0 \cap \mathcal{W}_{n_0-1}(y^*)$. We note here that not only the first factor of the Cartesian product of (3.15) is constant on $\mathcal{W}_{n_0-1}(y^*)$, but the second one, as well. The reason for this is that on $\mathcal{W}_{n_0-1}(y^*)$ the possible variations of the vector $q_{i(n_0)} - q_{j(n_0)}$ belong to the space $R_{n_0}[\mathcal{N}_{n_0-1}(V_0(y^*);\ g_1(y^*), \ldots, g_{n_0-1}(y^*))]$, see proposition 2.8.

The last set containment means that for any $x \in U_0 \cap \mathcal{W}_{n_0-1}(y^*)$ it is true that $r(x) \in H(x) = H(y^*)$, so $x \in J$ for all such x, according to proposition 2.14. However, this contradicts to the fact that for the phase points $y \in U_0$ the manifolds $\mathcal{W}_{n_0-1}(y)$ are transversal to J, see definition 2.13. □

Our proof of the theorem will be completed as soon as we prove the following counterpart of proposition 3.7.

Proposition 3.16. *Let $\mathcal{P}'$ be a two-dimensional linear subspace of the Euclidean space $\mathbb{R}^3$, $\mathcal{P} = \mathcal{P}' + x_0$ a coset of $\mathcal{P}'$ not containing 0, $C_2 \subset \mathbb{R}^3$ be the unit sphere of $\mathbb{R}^3$, $C_1 \subset \mathcal{P}$ be an ellipse in $\mathcal{P}$, possibly degenerated to a single point. Assume that the unit sphere C_2 intersects the affine plane $\mathcal{P}$ in a circle C and none of C_1 and C lies completely inside the other one. Suppose that $L(s)$, $|s| < \varepsilon_0$, is a smooth family of oriented lines in $\mathcal{P}$ with the direction vector $\overline{v}(s)$ satisfying the following conditions:*

- (i) *$L(s)$ is tangent to C_1 at the point of contact $A(s)$, and at $A(s)$ the direction vector $\overline{v}(s)$ agrees with a given orientation of C_1 (if C_1 is not a point),*
- (ii) *$L(s)$ intersects C in two points, out of which the one whose position vector makes the smaller inner product with $\overline{v}(s)$ is denoted by $B(s)$,*
- (iii) *$\frac{\mathrm{d}}{\mathrm{d}s}\alpha(\overline{v}(s)) > 0$ for all s, $|s| < \varepsilon_0$.*

Here $\alpha(\overline{v}(s))$ denotes the direction angle of the vector $\overline{v}(s)$. Finally, let $\overline{v}^+(s)$ be the mirror image of $\overline{v}(s)$ under the orthogonal reflection across the tangent plane of the unit sphere C_2 at the point $B(s)$.

We claim that there is no (four-dimensional) hyperplane $H \subset \mathcal{P}' \times \mathbb{R}^3$ containing all the points $(\overline{v}(s), \overline{v}^+(s))$ for $|s| < \varepsilon_0$.

Remark 3.17. In the proposition above the space $\mathbb{R}^3$ plays the role of the space S of proposition 3.11.

Proof. Very similar to the proof of proposition 3.7. We assume again that C_1 and C_2 possess at least two non-parallel tangent lines. (Otherwise the argument discussing the parallelity case $\overline{v}(b) = -\overline{v}(a)$ in the proof of proposition 3.7 applies here with obvious modifications, which are left to the reader.) We can also assume that the lines $L(s)$ depend on the parameter s analytically, and this analytic family of lines $L(s) \subset \mathcal{P}$ is already extended to a parameter interval $I = [a, b] \supset (-\varepsilon_0, \varepsilon_0)$ by keeping the properties (i)–(iii) above, so that $L(a)$ and $L(b)$ are non-parallel and tangent to the circle $C = \mathcal{P} \cap C_2$. Suppose there is a hyperplane H in the five-dimensional space $\mathcal{P}' \times \mathbb{R}^3$ containing all the points $(\overline{v}(s), \overline{v}^+(s))$ for $|s| < \varepsilon_0$. By reasons of analyticity, the same membership relation $(\overline{v}(s), \overline{v}^+(s)) \in H$ is true for all $s \in I$. The relations $(\overline{v}(a), \overline{v}(a)) \in H$, $(\overline{v}(b), \overline{v}(b)) \in H$ imply that H contains the diagonal $\{(v, v) \big| \ v \in \mathcal{P}'\}$, which diagonal is the kernel of the linear map $\Psi : H \to \mathbb{R}^3$, $\Psi(v_1, v_2) = v_1 - v_2$. Therefore, since $\dim H = 4$ by our assumption, we get that $\dim \Psi(H) \leqslant 2$. However, for the points $(\overline{v}(s), \overline{v}^+(s))$, $a < s < b$, the lines spanned by the vectors $\overline{v}(s) - \overline{v}^+(s)$ fill out a (nonempty) open part of a circular cone of $\mathbb{R}^3$, which cannot be the part of any subspace with dimension $\leqslant 2$, so the proposition and our non-coincidence theorem are now proved. □

4. Proof of the Boltzmann–Sinai Ergodic hypothesis for all hard ball systems

Proof. We carry out an induction on the number N of elastically interacting balls. For $N = 2$ this is the classic result of Sinai and Chernov [S-Ch(1987)]. 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 balls (of equal masses) on the flat ν-torus $\mathbb{T}^\nu$ with the number of balls 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 $J_1,\ J_2,\ \ldots$ of exceptional, codimension-one, smooth, non-hyperbolicity 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 balls. Finally, the Theorem of [Sim(2009)] gives us that the considered N-ball system is also ergodic, completely hyperbolic, hence Bernoulli mixing. □

Acknowledgments

Special thanks are due to N I Chernov and D Dolgopyat for their illuminating questions and remarks.

References

[B-S(1973)] Bunimovič L A and Sinaĭ Ja G 1973 The fundamental theorem of the theory of scattering billiards *Math. Sb. (N.S.)* **90** 415–31

[C-H(1996)] Chernov N I and Haskell C 1996 Nonuniformly hyperbolic K-systems are Bernoulli *Ergod. Theory Dyn. Syst.* **16** 19–44

[K-S-Sz(1990)] Krámli A, Simányi N and Szász D 1990 A 'transversal' fundamental theorem for semi-dispersing billiards *Commun. Math. Phys.* **129** 535–60

[K-S-Sz(1991)] Krámli A, Simányi N and Szász D 1991 The K-property of three billiard balls *Ann. Math.* **133** 37–72

[K-S-Sz(1992)] Krámli A, Simányi N and Szász D 1992 The K-property of four billiard balls *Commun. Math. Phys.* **144** 107–48

[O-W(1998)] Ornstein D and Weiss B 1998 On the Bernoulli nature of systems with some hyperbolic structure *Ergod. Theory Dyn. Syst.* **18** 441–56

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

[Sim(1992)-I] Simányi N 1992 The K-property of N billiard balls: I *Invent. Math.* **108** 521–48

[Sim(1992)-II] Simányi N 1992 The K-property of N billiard balls: II *Invent. Math.* **110** 151–72

[Sim(2002)] Simányi N 2002 The complete hyperbolicity of cylindric billiards *Ergod. Theory Dyn. Syst.* **22** 281–302

[Sim(2003)] Simányi N 2003 Proof of the Boltzmann–Sinai ergodic hypothesis for typical hard disk systems *Invent. Math.* **154** 123–78

[Sim(2004)] Simányi N 2004 Proof of the ergodic hypothesis for typical hard ball systems *Ann. Henri Poincaré* **5** 203–33

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

[Sin(1963)] Sinaĭ Ja G 1963 On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics *Sov. Math. Dokl.* **4** 1818–22

[Sin(1970)] Sinai Ya G 1970 Dynamical systems with elastic reflections *Russ. Math. Surv.* **25** 137–89

[S-Sz(1999)] Simányi N and Szász D 1999 Hard ball systems are completely hyperbolic *Ann. Math.* **149** 35–96

[S-Sz(2000)] Simányi N and Szász D 2000 Non-integrability of cylindric billiards and transitive Lie group actions *Ergod. Theory Dyn. Syst.* **20** 593–610