October 17, 2002

# Proof of the Ergodic Hypothesis for Typical Hard Ball Systems

**Nándor Simányi**[1]

University of Alabama at Birmingham
Department of Mathematics
Campbell Hall, Birmingham, AL 35294 U.S.A.
E-mail: simanyi@math.uab.edu

**Abstract.** We consider the system of $N$ ($\geq 2$) hard balls with masses $m_1, \dots, m_N$ and radius $r$ in the flat torus $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$ of size $L$, $\nu \geq 3$. We prove the ergodicity (actually, the Bernoulli mixing property) of such systems for almost every selection $(m_1, \dots, m_N; L)$ of the outer geometric parameters. This theorem complements my earlier result that proved the same, almost sure ergodicity for the case $\nu = 2$. The method of that proof was primarily dynamical-geometric, whereas the present approach is inherently algebraic.

Primary subject classification: 37D50

Secondary subject classification: 34D05

## §1. Introduction

Hard ball systems or, a bit more generally, mathematical billiards constitute an important and quite interesting family of dynamical systems being intensively studied by dynamicists and researchers of mathematical physics, as well. These dynamical systems pose many challenging mathematical questions, most of them

[1]Research supported by the National Science Foundation, grant DMS-0098773.

concerning the ergodic (mixing) properties of such systems. The introduction of hard ball systems and the first major steps in their investigations date back to the 40's and 60's, see Krylov's paper [K(1942)] and Sinai's ground-breaking works [Sin(1963)] and [Sin(1970)], in which the author — among other things — formulated the modern version of Boltzmann's ergodic hypothesis (what we call today the Boltzmann–Sinai ergodic hypothesis) by claiming that every hard ball system in a flat torus is ergodic, of course after fixing the values of the trivial flow-invariant quantities. In the articles [Sin(1970)] and [B-S(1973)] Bunimovich and Sinai proved this hypothesis for two hard disks on the two-dimensional unit torus $\mathbb{T}^2$. The generalization of this result to higher dimensions $\nu > 2$ took fourteen years, and was done by Chernov and Sinai in [S-Ch(1987)]. Although the model of two hard balls in $\mathbb{T}^\nu$ is already rather involved technically, it is still a so called strictly dispersive billiard system, i. e. such that the smooth components of the boundary $\partial\mathbf{Q}$ of the configuration space are strictly concave from outside $\mathbf{Q}$. (They are bending away from $\mathbf{Q}$.) The billiard systems of more than two hard balls in $\mathbb{T}^\nu$ are no longer strictly dispersive, but just semi-dispersive (strict concavity of the smooth components of $\partial\mathbf{Q}$ is lost, merely concavity persists), and this circumstance causes a lot of additional technical troubles in their study. In the series of my joint papers with A. Krámli and D. Szász [K-S-Sz(1989)], [K-S-Sz(1990)], [K-S-Sz(1991)], and [K-S-Sz(1992)] we developed several new methods, and proved the ergodicity of more and more complicated semi-dispersive billiards culminating in the proof of ergodicity of four billiard balls in the torus $\mathbb{T}^\nu$ ($\nu \geq 3$), [K-S-Sz(1992)]. Then, in 1992, Bunimovich, Liverani, Pellegrinotti and Sukhov [B-L-P-S(1992)] were able to prove the ergodicity for some systems with an arbitrarily large number of hard balls. The shortcoming of their model, however, is that, on one hand, they restrict the types of all feasible ball-to-ball collisions, on the other hand they introduce some additional scattering effect with the collisions at the strictly concave wall of the container. The only result with an arbitrarily large number of balls in a flat unit torus $\mathbb{T}^\nu$ was achieved in the twin papers of mine [Sim(1992-I-II)], where I managed to prove the ergodicity (actually, the K-mixing property) of $N$ hard balls in $\mathbb{T}^\nu$, provided that $N \leq \nu$. The annoying shortcoming of that result is that the larger the number of balls $N$ is, larger and larger dimension $\nu$ of the ambient container is required by the method of the proof.

On the other hand, if someone considers a hard ball system in an elongated torus which is long in one direction but narrow in the others, so that the balls must keep their cyclic order in the "long direction" (Sinai's "pen-case" model), then the technical difficulties can be handled, thanks to the fact that the collisions of balls are now restricted to neighboring pairs. The hyperbolicity of such models in three dimensions and the ergodicity in dimension four have been proved in [S-Sz(1995)].

The positivity of the metric entropy for several systems of hard balls can be proven relatively easily, as was shown in the paper [W(1988)]. The articles [L-W(1995)] and [W(1990)] are nice surveys describing a general setup leading to the technical problems treated in a series of research papers. For a comprehensive survey of the results and open problems in this field, see [Sz(1996)].

Pesin's theory [P(1977)] on the ergodic properties of non-uniformly hyperbolic, smooth dynamical systems has been generalized substantially to dynamical systems with singularities (and with a relatively mild behavior near the singularities) by A. Katok and J-M. Strelcyn [K-S(1986)]. Since then, the so called Pesin's and Katok-Strelcyn's theories have become part of the folklore in the theory of dynamical systems. They claim that — under some mild regularity conditions, particularly near the singularities — every non-uniformly hyperbolic and ergodic flow enjoys the Kolmogorov-mixing property, shortly the K-mixing property.

Later on it was discovered and proven in [C-H(1996)] and [O-W(1998)] that the above mentioned fully hyperbolic and ergodic flows with singularities turn out to be automatically having the Bernoulli mixing (B-mixing) property. It is worth noting here that almost every semi-dispersive billiard system, especially every hard ball system, enjoys those mild regularity conditions imposed on the systems (as axioms) by [K-S(1986)], [C-H(1996)], and [O-W(1998)]. In other words, for a hard ball flow $(\mathbf{M}, \{S^t\}, \mu)$ the (global) ergodicity of the system actually implies its full hyperbolicity and the B-mixing property, as well.

Finally, in our joint venture with D. Szász [S-Sz(1999)], we prevailed over the difficulty caused by the low value of the dimension $\nu$ by developing a brand new algebraic approach for the study of hard ball systems. That result, however, only establishes complete hyperbolicity (nonzero Lyapunov exponents almost everywhere) for $N$ balls in $\mathbb{T}^\nu$. The ergodicity appeared to be a harder task.

We note, however, that the algebraic method developed in [S-Sz(1999)] is being further developed in this paper in order to obtain ergodicity, not only full hyperbolicity.

Consider the $\nu$-dimensional ($\nu \geq 2$), standard, flat torus $\mathbb{T}^\nu_L = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$ as the vessel containing $N$ ($\geq 2$) hard balls (spheres) $B_1, \dots, B_N$ with positive masses $m_1, \dots, m_N$ and (just for simplicity) common radius $r > 0$. We always assume that the radius $r > 0$ is not too big, so that even the interior of the arising configuration space $\mathbf{Q}$ is connected. Denote the center of the ball $B_i$ by $q_i \in \mathbb{T}^\nu$, and let $v_i = \dot{q}_i$ be the velocity of the $i$-th particle. We investigate the uniform motion of the balls $B_1, \dots, B_N$ inside the container $\mathbb{T}^\nu$ with half a unit of total kinetic energy: $E = \dfrac{1}{2} \sum_{i=1}^N m_i ||v_i||^2 = \dfrac{1}{2}$. We assume that the collisions between balls are perfectly elastic. Since — beside the kinetic energy $E$ — the total momentum $I = \sum_{i=1}^N m_i v_i \in \mathbb{R}^\nu$ is also a trivial first integral of the motion, we make the standard reduction $I = 0$. Due to the apparent translation invariance of the arising dynamical system, we factorize the configuration space with respect to uniform spatial translations as follows: $(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a)$ for all translation vectors $a \in \mathbb{T}^\nu$. The configuration space $\mathbf{Q}$ of the arising flow is then the factor torus $\left( (\mathbb{T}^\nu)^N / \sim \right) \cong \mathbb{T}^{\nu(N-1)}$ minus the cylinders

$$C_{i,j} = \left\{ (q_1, \dots, q_N) \in \mathbb{T}^{\nu(N-1)} \colon \operatorname{dist}(q_i, q_j) < 2r \right\}$$

$(1 \le i < j \le N)$ corresponding to the forbidden overlap between the $i$-th and $j$-th spheres. Then it is easy to see that the compound configuration point

$$q = (q_1, \dots, q_N) \in \mathbf{Q} = \mathbb{T}^{\nu(N-1)} \setminus \bigcup_{1 \le i < j \le N} C_{i,j}$$

moves in $\mathbf{Q}$ uniformly with unit speed and bounces back from the boundaries $\partial C_{i,j}$ of the cylinders $C_{i,j}$ according to the classical law of geometric optics: the angle of reflection equals the angle of incidence. More precisely: the post-collision velocity $v^+$ can be obtained from the pre-collision velocity $v^-$ by the orthogonal reflection across the tangent hyperplane of the boundary $\partial\mathbf{Q}$ at the point of collision. Here we must emphasize that the phrase "orthogonal" should be understood with respect to the natural Riemannian metric (the so called mass metric) $||dq||^2 = \sum_{i=1}^N m_i ||dq_i||^2$ in the configuration space $\mathbf{Q}$. For the normalized Liouville measure $\mu$ of the arising flow $\{S^t\}$ we obviously have $d\mu = \text{const} \cdot dq \cdot dv$, where $dq$ is the Riemannian volume in $\mathbf{Q}$ induced by the above metric and $dv$ is the surface measure (determined by the restriction of the Riemannian metric above) on the sphere of compound velocities

$$\mathbb{S}^{\nu(N-1)-1} = \left\{ (v_1, \dots, v_N) \in (\mathbb{R}^\nu)^N \colon \sum_{i=1}^N m_i v_i = 0 \text{ and } \sum_{i=1}^N m_i ||v_i||^2 = 1 \right\}.$$

The phase space $\mathbf{M}$ of the flow $\{S^t\}$ is the unit tangent bundle $\mathbf{Q} \times \mathbb{S}^{d-1}$ of the configuration space $\mathbf{Q}$. (We will always use the shorthand notation $d = \nu(N-1)$ for the dimension of the billiard table $\mathbf{Q}$.) We must, however, note here that at the boundary $\partial\mathbf{Q}$ of $\mathbf{Q}$ one has to glue together the pre-collision and post-collision velocities in order to form the phase space $\mathbf{M}$, so $\mathbf{M}$ is equal to the unit tangent bundle $\mathbf{Q} \times \mathbb{S}^{d-1}$ modulo this identification.

A bit more detailed definition of hard ball systems with arbitrary masses, as well as their role in the family of cylindric billiards, can be found in §4 of [S-Sz(2000)] and in §1 of [S-Sz(1999)]. We denote the arising flow by $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$.

In the series of articles [K-S-Sz(1989)], [K-S-Sz(1991)], [K-S-Sz(1992)], [Sim(1992-I)], and [Sim(1992-II)] the authors developed a powerful, three-step strategy for proving the (hyperbolic) ergodicity of hard ball systems. First of all, all these proofs are inductions on the number $N$ of balls involved in the problem. Secondly, the induction step itself consists of the following three major steps:

**Step I.** To prove that every non-singular (i. e. smooth) trajectory segment $S^{[a,b]}x_0$ with a "combinatorially rich" (in a well defined sense) symbolic collision sequence is automatically sufficient (or, in other words, "geometrically hyperbolic", see below in §2), provided that the phase point $x_0$ does not belong to a countable union $J$ of smooth sub-manifolds with codimension at least two. (Containing the exceptional phase points.)

The exceptional set $J$ featuring this result is negligible in our dynamical considerations — it is a so called slim set. For the basic properties of slim sets, see §2 below.

**Step II.** Assume the induction hypothesis, i. e. that all hard ball systems with $N'$ balls ($2 \le N' < N$) are (hyperbolic and) ergodic. Prove that then there exists a slim set $S \subset \mathbf{M}$ (see §2) with the following property: For every phase point $x_0 \in \mathbf{M} \setminus S$ the entire trajectory $S^{\mathbb{R}} x_0$ contains at most one singularity and its symbolic collision sequence is combinatorially rich, just as required by the result of Step I.

**Step III.** By using again the induction hypothesis, prove that almost every singular trajectory is sufficient in the time interval $(t_0, +\infty)$, where $t_0$ is the time moment of the singular reflection. (Here the phrase "almost every" refers to the volume defined by the induced Riemannian metric on the singularity manifolds.)

We note here that the almost sure sufficiency of the singular trajectories (featuring Step III) is an essential condition for the proof of the celebrated Theorem on Local Ergodicity for algebraic semi-dispersive billiards proved by Bálint–Chernov–Szász–Tóth in [B-Ch-Sz-T (2001)]. Under this assumption the theorem of [B-Ch-Sz-T (2001)] states that in any algebraic semi-dispersive billiard system (i. e. in a system such that the smooth components of the boundary $\partial \mathbf{Q}$ are algebraic hypersurfaces) a suitable, open neighborhood $U_0$ of any sufficient phase point $x_0 \in \mathbf{M}$ (with at most one singularity on its trajectory) belongs to a single ergodic component of the billiard flow $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$.

In an inductive proof of ergodicity, steps I and II together ensure that there exists an arc-wise connected set $C \subset \mathbf{M}$ with full measure, such that every phase point $x_0 \in C$ is sufficient with at most one singularity on its trajectory. Then the cited Theorem on Local Ergodicity (now taking advantage of the result of Step III) states that for every phase point $x_0 \in C$ an open neighborhood $U_0$ of $x_0$ belongs to one ergodic component of the flow. Finally, the connectedness of the set $C$ and $\mu(\mathbf{M} \setminus C) = 0$ easily imply that the flow $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$ (now with $N$ balls) is indeed ergodic, and actually fully hyperbolic, as well.

The main result of this paper is the

**Theorem.** In the case $\nu \ge 3$ for almost every selection $(m_1, \dots, m_N; L)$ of the outer geometric parameters from the region $m_i > 0$, $L > L_0(r, \nu)$ where the interior of the phase space is connected it is true that the billiard flow $(\mathbf{M}_{\vec{m},L}, \{S^t\}, \mu_{\vec{m},L})$ of the $N$-ball system is ergodic and completely hyperbolic. Then, following from the results of Chernov–Haskell [C-H(1996)] and Ornstein–Weiss [O-W(1998)], such a semi-dispersive billiard system actually enjoys the B-mixing property, as well.

**Remark 1.** We note that the main result of this paper and that of [Sim(2001)] nicely complement each other. They precisely assert the same, almost sure ergodicity of hard ball systems in the cases $\nu \geq 3$ and $\nu = 2$, respectively. It should be noted, however, that the proof of [Sim(2001)] is primarily dynamical–geometric (except the verification of the Chernov-Sinai Ansatz), whereas the novel parts of the present proof are fundamentally algebraic.

**Remark 2.** The above inequality $L > L_0(r, \nu)$ corresponds to physically relevant situations. Indeed, in the case $L < L_0(r, \nu)$ the particles would not have enough room even to freely exchange positions.

The paper is organized as follows: §2 provides all necessary prerequisites and technical tools that will be required by the proof of the theorem. §3 lays down the foundations of our algebraic approach to hard ball systems. To a large extent, that section is a heavily edited adaptation of §3 of [S-Sz(1999)]. The reason why we included that section from an earlier paper is that we wanted to present here an exposition, as self-contained as possible. Based on the results obtained in §3, the subsequent section §4 carries out Step I of the inductive strategy outlined above, but for the case when the outer geometric parameters $(m_1, \dots, m_N;\, L)$ are incorporated in the algebraic process as variables. (Just as the positions and velocities of the particles!) Finally, the closing section §5 utilizes a "Fubini type argument" by proving Step I for almost every (with respect to the Lebesgue measure of the $(m_1, \dots, m_N;\, L)$-space) hard ball system ($N \geq 2$, $\nu \geq 3$). This will finish the inductive proof of the Theorem, for Steps II and III of the induction strategy are easy consequences of some earlier results.

## 2. Prerequisites

**2.1 Cylindric billiards.** Consider the $d$-dimensional ($d \geq 2$) flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathcal{L}$ supplied with the usual Riemannian inner product $\langle\, .\, ,\, .\, \rangle$ inherited from the standard inner product of the universal covering space $\mathbb{R}^d$. Here $\mathcal{L} \subset \mathbb{R}^d$ is assumed to be a lattice, i. e. a discrete subgroup of the additive group $\mathbb{R}^d$ with $\mathrm{rank}(\mathcal{L}) = d$. The reason why we want to allow general lattices, other than just the integer lattice $\mathbb{Z}^d$, is that otherwise the hard ball systems would not be covered. The geometry of the structure lattice $\mathcal{L}$ in the case of a hard ball system is significantly different from the geometry of the standard lattice $\mathbb{Z}^d$ in the standard Euclidean space $\mathbb{R}^d$, see later in this section.

The configuration space of a cylindric billiard is $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \cdots \cup C_k)$, where the cylindric scatterers $C_i$ ($i = 1, \dots, k$) are defined as follows.

Let $A_i \subset \mathbb{R}^d$ be a so called lattice subspace of $\mathbb{R}^d$, which means that $\mathrm{rank}(A_i \cap \mathcal{L}) = \dim A_i$. In this case the factor $A_i/(A_i \cap \mathcal{L})$ is a sub-torus in $\mathbb{T}^d = \mathbb{R}^d/\mathcal{L}$ which will be taken as the generator of the cylinder $C_i \subset \mathbb{T}^d$, $i = 1, \dots, k$. Denote by $L_i = A_i^{\perp}$ the ortho-complement of $A_i$ in $\mathbb{R}^d$. Throughout this paper we will always assume that $\dim L_i \geq 2$. Let, furthermore, the numbers $r_i > 0$ (the radii of the spherical cylinders $C_i$) and some translation vectors $t_i \in \mathbb{T}^d = \mathbb{R}^d/\mathcal{L}$ be given. The translation vectors $t_i$ play a role in positioning the cylinders $C_i$ in the ambient torus $\mathbb{T}^d$. Set

$$C_i = \left\{ x \in \mathbb{T}^d \colon \operatorname{dist}\left(x - t_i, A_i/(A_i \cap \mathcal{L})\right) < r_i \right\}.$$

In order to avoid further unnecessary complications, we always assume that the interior of the configuration space $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \dots \cup C_k)$ is connected. The phase space $\mathbf{M}$ of our cylindric billiard flow will be the unit tangent bundle of $\mathbf{Q}$ (modulo the natural gluing at its boundary), i. e. $\mathbf{M} = \mathbf{Q} \times \mathbb{S}^{d-1}$. (Here $\mathbb{S}^{d-1}$ denotes the unit sphere of $\mathbb{R}^d$.)

The dynamical system $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$, where $S^t$ $(t \in \mathbb{R})$ is the dynamics defined by the uniform motion inside the domain $\mathbf{Q}$ and specular reflections at its boundary (at the scatterers), and $\mu$ is the Liouville measure, is called a cylindric billiard flow.

We note that the cylindric billiards — defined above — belong to the wider class of so called semi-dispersive billiards, which means that the smooth components $\partial \mathbf{Q}_i$ of the boundary $\partial \mathbf{Q}$ of the configuration space $\mathbf{Q}$ are (not necessarily strictly) concave from outside of $\mathbf{Q}$, i. e. they are bending away from the interior of $\mathbf{Q}$. As to the notions and notations in connection with semi-dispersive billiards, the reader is kindly referred to the article [K-S-Sz(1990)].

Throughout this paper we will always assume — without explicitly stating — that the considered semi-dispersive billiard system fulfills the following conditions:

(2.1.1) $\qquad$ $\operatorname{int}\mathbf{Q}$ is connected, and

(2.1.2) $\qquad$ the $d$-dim spatial angle $\alpha(q)$ subtended by $\mathbf{Q}$ at any of its boundary points $q \in \partial \mathbf{Q}$ is uniformly positive.

We note, however, that in the case of hard ball systems with a fixed radius $r$ of the balls (see below) the non-degeneracy condition (2.1.2) only excludes countably many values of the size $L$ of the container torus $\mathbb{T}^{\nu}_L = \mathbb{R}^{\nu}/L \cdot \mathbb{Z}^{\nu}$ from the region $L > L_0(r, \nu)$ where (2.1.1) is true. Therefore, in the sense of our theorem of "almost sure ergodicity", the non-degeneracy condition (2.1.2) does not mean a restriction of generality.

**2.2 Hard ball systems.** Hard ball systems in the flat torus $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$ ($\nu \geq 2$) with positive masses $m_1, \dots, m_N$ are described (for example) in §1 of [S-Sz(1999)]. These are the dynamical systems describing the motion of $N$ ($\geq 2$) hard balls with a common radius $r > 0$ and positive masses $m_1, \dots, m_N$ in the flat torus of size $L$, $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$. (Just for simplicity, we will assume that the radii have the common value $r$.) The center of the $i$-th ball is denoted by $q_i$ ($\in \mathbb{T}_L^\nu$), its time derivative is $v_i = \dot{q}_i$, $i = 1, \dots, N$. One uses the standard reduction of kinetic energy $E = \frac{1}{2} \sum_{i=1}^N m_i ||v_i||^2 = \frac{1}{2}$. The arising configuration space (still without the removal of the scattering cylinders $C_{i,j}$) is the torus

$$\mathbb{T}_L^{\nu N} = \left(\mathbb{T}_L^\nu\right)^N = \{(q_1, \dots, q_N) \colon q_i \in \mathbb{T}_L^\nu,\ i = 1, \dots, N\}$$

supplied with the Riemannian inner product (the so called mass metric)

$$\langle v, v' \rangle = \sum_{i=1}^N m_i \langle v_i, v_i' \rangle \tag{2.2.1}$$

in its common tangent space $\mathbb{R}^{\nu N} = \left(\mathbb{R}^\nu\right)^N$. Now the Euclidean space $\mathbb{R}^{\nu N}$ with the inner product (2.2.1) plays the role of $\mathbb{R}^d$ in the original definition of cylindric billiards, see §2.1 above.

The generator subspace $A_{i,j} \subset \mathbb{R}^{\nu N}$ ($1 \leq i < j \leq N$) of the cylinder $C_{i,j}$ (describing the collisions between the $i$-th and $j$-th balls) is given by the equation

$$A_{i,j} = \left\{ (q_1, \dots, q_N) \in \left(\mathbb{R}^\nu\right)^N \colon\ q_i = q_j \right\}, \tag{2.2.2}$$

see (4.3) in [S-Sz(2000)]. Its ortho-complement $L_{i,j} \subset \mathbb{R}^{\nu N}$ is then defined by the equation

$$L_{i,j} = \left\{ (q_1, \dots, q_N) \in \left(\mathbb{R}^\nu\right)^N \colon\ q_k = 0 \text{ for } k \neq i, j, \text{ and } m_i q_i + m_j q_j = 0 \right\}, \tag{2.2.3}$$

see (4.4) in [S-Sz(2000)]. Easy calculation shows that the cylinder $C_{i,j}$ (describing the overlap of the $i$-th and $j$-th balls) is indeed spherical and the radius of its base sphere is equal to $r_{i,j} = 2r\sqrt{\frac{m_i m_j}{m_i + m_j}}$, see §4, especially formula (4.6) in [S-Sz(2000)].

The structure lattice $\mathcal{L} \subset \mathbb{R}^{\nu N}$ is clearly the lattice $\mathcal{L} = \left(L \cdot \mathbb{Z}^\nu\right)^N = L \cdot \mathbb{Z}^{N\nu}$.

Due to the presence of an extra invariant quantity $I = \sum_{i=1}^N m_i v_i$, one usually makes the reduction $\sum_{i=1}^N m_i v_i = 0$ and, correspondingly, factorizes the configuration space with respect to uniform spatial translations:

$$(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a), \quad a \in \mathbb{T}_L^\nu. \tag{2.2.4}$$

The natural, common tangent space of this reduced configuration space is then

$$\mathcal{Z}=\left\{(v_1,\dots,v_N)\in(\mathbb{R}^\nu)^N:\ \sum_{i=1}^N m_iv_i=0\right\}=\left(\bigcap_{i<j}A_{i,j}\right)^\perp=(\mathcal{A})^\perp \tag{2.2.5}$$

supplied again with the inner product (2.2.1), see also (4.1) and (4.2) in [S-Sz(2000)]. The base spaces $L_{i,j}$ of (2.2.3) are obviously subspaces of $\mathcal{Z}$, and we take $\tilde{A}_{i,j}=A_{i,j}\cap\mathcal{Z}=P_{\mathcal{Z}}(A_{i,j})$ as the ortho-complement of $L_{i,j}$ in $\mathcal{Z}$. (Here $P_{\mathcal{Z}}$ denotes the orthogonal projection onto the space $\mathcal{Z}$.)

Note that the configuration space of the reduced system (with the identification (2.2.4)) is naturally the torus $\mathbb{R}^{\nu N}/(\mathcal{A}+L\cdot\mathbb{Z}^{\nu N})=\mathcal{Z}/P_{\mathcal{Z}}(L\cdot\mathbb{Z}^{\nu N})$.

**2.3. Collision graphs.** Let $S^{[a,b]}x$ be a nonsingular, finite trajectory segment with the collisions $\sigma_1,\dots,\sigma_n$ listed in time order. (Each $\sigma_k$ is an unordered pair $(i,j)$ of different labels $i,j\in\{1,2,\dots,N\}$.) The graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with vertex set $\mathcal{V}=\{1,2,\dots,N\}$ and set of edges $\mathcal{E}=\{\sigma_1,\dots,\sigma_n\}$ is called the *collision graph* of the orbit segment $S^{[a,b]}x$. For a given positive number $C$, the collision graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ of the orbit segment $S^{[a,b]}x$ will be called *$C$-rich* if $\mathcal{G}$ contains at least $C$ connected, consecutive (i. e. following one after the other in time, according to the time-ordering given by the trajectory segment $S^{[a,b]}x$) subgraphs.

**2.4 Trajectory Branches.** We are going to briefly describe the discontinuity of the flow $\{S^t\}$ caused by a multiple collisions at time $t_0$. Assume first that the pre–collision velocities of the particles are given. What can we say about the possible post–collision velocities? Let us perturb the pre–collision phase point (at time $t_0-0$) infinitesimally, so that the collisions at $\sim t_0$ occur at infinitesimally different moments. By applying the collision laws to the arising finite sequence of collisions, we see that the post-collision velocities are fully determined by the time– ordering of the considered collisions. Therefore, the collection of all possible time-orderings of these collisions gives rise to a finite family of continuations of the trajectory beyond $t_0$. They are called the trajectory branches. It is quite clear that similar statements can be said regarding the evolution of a trajectory through a multiple collision in reverse time. Furthermore, it is also obvious that for any given phase point $x_0\in\mathbf{M}$ there are two, $\omega$-high trees $\mathcal{T}_+$ and $\mathcal{T}_-$ such that $\mathcal{T}_+$ ($\mathcal{T}_-$) describes all the possible continuations of the positive (negative) trajectory $S^{[0,\infty)}x_0$ ($S^{(-\infty,0]}x_0$). (For the definitions of trees and for some of their applications to billiards, cf. the beginning of §5 in [K-S-Sz(1992)].) It is also clear that all possible continuations (branches) of the whole trajectory $S^{(-\infty,\infty)}x_0$ can be uniquely described by all pairs $(B_-,B_+)$ of $\omega$-high branches of the trees $\mathcal{T}_-$ and $\mathcal{T}_+$ ($B_-\subset\mathcal{T}_-, B_+\subset\mathcal{T}_+$).

Finally, we note that the trajectory of the phase point $x_0$ has exactly two branches, provided that $S^tx_0$ hits a singularity for a single value $t=t_0$, and the phase point $S^{t_0}x_0$ does not lie on the intersection of more than one singularity manifolds. In this case we say that the trajectory of $x_0$ has a "simple singularity".

**2.5 Neutral Subspaces, Advance, and Sufficiency.** Consider a nonsingular trajectory segment $S^{[a,b]}x$. Suppose that $a$ and $b$ are not moments of collision.

**Definition 2.5.1.** *The neutral space $\mathcal{N}_0(S^{[a,b]}x)$ of the trajectory segment $S^{[a,b]}x$ at time zero ($a<0<b$) is defined by the following formula:*

$$\begin{aligned}&\mathcal{N}_0(S^{[a,b]}x)=\big\{W\in\mathcal{Z}\colon\ \exists(\delta>0)\ \ \textit{such that}\ \ \forall\alpha\in(-\delta,\delta)\\ &V\left(S^a\left(Q(x)+\alpha W,V(x)\right)\right)=V(S^a x)\ \textit{and}\ V\left(S^b\left(Q(x)+\alpha W,V(x)\right)\right)=V(S^b x)\big\}.\end{aligned}$$

($\mathcal{Z}$ is the common tangent space $\mathcal{T}_q\mathbf{Q}$ of the parallelizable manifold $\mathbf{Q}$ at any of its points $q$, while $V(x)$ is the velocity component of the phase point $x=(Q(x),\,V(x))$.)

It is known (see (3) in §3 of [S-Ch (1987)]) that $\mathcal{N}_0(S^{[a,b]}x)$ is a linear subspace of $\mathcal{Z}$ indeed, and $V(x)\in\mathcal{N}_0(S^{[a,b]}x)$. The neutral space $\mathcal{N}_t(S^{[a,b]}x)$ of the segment $S^{[a,b]}x$ at time $t\in[a,b]$ is defined as follows:

$$\mathcal{N}_t(S^{[a,b]}x)=\mathcal{N}_0\left(S^{[a-t,b-t]}(S^t x)\right).$$

It is clear that the neutral space $\mathcal{N}_t(S^{[a,b]}x)$ can be canonically identified with $\mathcal{N}_0(S^{[a,b]}x)$ by the usual identification of the tangent spaces of $\mathbf{Q}$ along the trajectory $S^{(-\infty,\infty)}x$ (see, for instance, §2 of [K-S-Sz(1990)]).

Our next definition is that of the advance. Consider a non-singular orbit segment $S^{[a,b]}x$ with the symbolic collision sequence $\Sigma=(\sigma_1,\dots,\sigma_n)$ ($n\ge 1$), meaning that $S^{[a,b]}x$ has exactly $n$ collisions with $\partial\mathbf{Q}$, and the $i$-th collision ($1\le i\le n$) takes place at the boundary of the cylinder $C_{\sigma_i}$. For $x=(Q,V)\in\mathbf{M}$ and $W\in\mathcal{Z}$, $\|W\|$ sufficiently small, denote $T_W(Q,V):=(Q+W,V)$.

**Definition 2.5.2.** *For any $1\le k\le n$ and $t\in[a,b]$, the advance*

$$\alpha(\sigma_k)\colon\ \mathcal{N}_t(S^{[a,b]}x)\to\mathbb{R}$$

*of the collision $\sigma_k$ is the unique linear extension of the linear functional $\alpha(\sigma_k)$ defined in a sufficiently small neighborhood of the origin of $\mathcal{N}_t(S^{[a,b]}x)$ in the following way:*

$$\alpha(\sigma_k)(W):=t_k(x)-t_k(S^{-t}T_W S^t x).$$

Here $t_k=t_k(x)$ is the time moment of the $k$-th collision $\sigma_k$ on the trajectory of $x$ after time $t=a$. The above formula and the notion of the advance functional

$$\alpha_k=\alpha(\sigma_k):\ \mathcal{N}_t\left(S^{[a,b]}x\right)\to\mathbb{R}$$

has two important features:

(i) If the spatial translation $(Q, V) \mapsto (Q + W, V)$ is carried out at time $t$, then $t_k$ changes linearly in $W$, and it takes place just $\alpha_k(W)$ units of time earlier. (This is why it is called "advance".)

(ii) If the considered reference time $t$ is somewhere between $t_{k-1}$ and $t_k$, then the neutrality of $W$ with respect to $\sigma_k$ precisely means that

$$W - \alpha_k(W) \cdot V(x) \in A_{\sigma_k},$$

i. e. a neutral (with respect to the collision $\sigma_k$) spatial translation $W$ with the advance $\alpha_k(W) = 0$ means that the vector $W$ belongs to the generator space $A_{\sigma_k}$ of the cylinder $C_{\sigma_k}$.

It is now time to bring up the basic notion of sufficiency (or, sometimes it is also called geometric hyperbolicity) of a trajectory (segment). This is the utmost important necessary condition for the proof of the fundamental theorem for algebraic semi-dispersive billiards, see Theorem 4.4 in [B-Ch-Sz-T(2001)].

**Definition 2.5.3.**

(1) *The nonsingular trajectory segment $S^{[a,b]}x$ ($a$ and $b$ are supposed not to be moments of collision) is said to be sufficient if and only if the dimension of $\mathcal{N}_t(S^{[a,b]}x)$ ($t \in [a, b]$) is minimal, i.e. dim $\mathcal{N}_t(S^{[a,b]}x) = 1$.*

(2) *The trajectory segment $S^{[a,b]}x$ containing exactly one singularity (a so called "simple singularity", see 2.4 above) is said to be sufficient if and only if both branches of this trajectory segment are sufficient.*

**Definition 2.5.4.** *The phase point $x \in \mathbf{M}$ with at most one (simple) singularity is said to be sufficient if and only if its whole trajectory $S^{(-\infty,\infty)}x$ is sufficient, which means, by definition, that some of its bounded segments $S^{[a,b]}x$ are sufficient.*

In the case of an orbit $S^{(-\infty,\infty)}x$ with a simple singularity, sufficiency means that both branches of $S^{(-\infty,\infty)}x$ are sufficient.

**2.6. No accumulation (of collisions) in finite time.** By the results of Vaserstein [V(1979)], Galperin [G(1981)] and Burago-Ferleger-Kononenko [B-F-K(1998)], in a semi-dispersive billiard flow with the property (2.1.2) there can only be finitely many collisions in finite time intervals, see Theorem 1 in [B-F-K(1998)]. Thus, the dynamics is well defined as long as the trajectory does not hit more than one boundary components at the same time.

**2.7. Slim sets.** We are going to summarize the basic properties of codimension-two subsets $A$ of a connected, smooth manifold $M$ with a possible boundary. Since these subsets $A$ are just those negligible in our dynamical discussions, we shall call them slim. As to a broader exposition of the issues, see [E(1978)] or §2 of [K-S-Sz(1991)].

Note that the dimension $\dim A$ of a separable metric space $A$ is one of the three classical notions of topological dimension: the covering (Čech-Lebesgue), the small inductive (Menger-Urysohn), or the large inductive (Brouwer-Čech) dimension. As it is known from general general topology, all of them are the same for separable metric spaces.

**Definition 2.7.1.** A subset $A$ of $M$ is called slim if and only if $A$ can be covered by a countable family of codimension-two (i. e. at least two) closed sets of $\mu$–measure zero, where $\mu$ is a smooth measure on $M$. (Cf. Definition 2.12 of [K-S-Sz(1991)].)

**Property 2.7.2.** The collection of all slim subsets of $M$ is a $\sigma$-ideal, that is, countable unions of slim sets and arbitrary subsets of slim sets are also slim.

**Proposition 2.7.3. (Locality).** A subset $A \subset M$ is slim if and only if for every $x \in A$ there exists an open neighborhood $U$ of $x$ in $M$ such that $U \cap A$ is slim. (Cf. Lemma 2.14 of [K-S-Sz(1991)].)

**Property 2.7.4.** A closed subset $A \subset M$ is slim if and only if $\mu(A) = 0$ and $\dim A \leq \dim M - 2$.

**Property 2.7.5. (Integrability).** If $A \subset M_1 \times M_2$ is a closed subset of the product of two smooth manifolds with possible boundaries, and for every $x \in M_1$ the set

$$A_x = \{y \in M_2 \colon\ (x, y) \in A\}$$

is slim in $M_2$, then $A$ is slim in $M_1 \times M_2$.

The following propositions characterize the codimension-one and codimension-two sets.

**Proposition 2.7.6.** For any closed subset $S \subset M$ the following three conditions are equivalent:

(i) $\dim S \leq \dim M - 2$;

(ii) $\mathrm{int} S = \emptyset$ and for every open connected set $G \subset M$ the difference set $G \setminus S$ is also connected;

(iii) $\mathrm{int} S = \emptyset$ and for every point $x \in M$ and for any open neighborhood $V$ of $x$ in $M$ there exists a smaller open neighborhood $W \subset V$ of the point $x$ such that for every pair of points $y, z \in W \setminus S$ there is a continuous curve $\gamma$ in the set $V \setminus S$ connecting the points $y$ and $z$.

(See Theorem 1.8.13 and Problem 1.8.E of [E(1978)].)

**Proposition 2.7.7.** For any subset $S \subset M$ the condition $\dim S \leq \dim M - 1$ is equivalent to $\mathrm{int} S = \emptyset$. (See Theorem 1.8.10 of [E(1978)].)

We recall an elementary, but important lemma (Lemma 4.15 of [K-S-Sz(1991)]). Let $R_2$ be the set of phase points $x \in \mathbf{M} \setminus \partial \mathbf{M}$ such that the trajectory $S^{(-\infty,\infty)}x$ has more than one singularities.

**Proposition 2.7.8.** The set $R_2$ is a countable union of codimension-two smooth sub-manifolds of $M$ and, being such, it is slim.

The next lemma establishes the most important property of slim sets which gives us the fundamental geometric tool to connect the open ergodic components of billiard flows.

**Proposition 2.7.9.** If $M$ is connected, then the complement $M \setminus A$ of a slim $F_\sigma$ set $A \subset M$ is an arc-wise connected ($G_\delta$) set of full measure. (See Property 3 of §4.1 in [K-S-Sz(1989)]. The $F_\sigma$ sets are, by definition, the countable unions of closed sets, while the $G_\delta$ sets are the countable intersections of open sets.)

**2.8. The subsets $\mathbf{M}^0$ and $\mathbf{M}^\#$.** Denote by $\mathbf{M}^\#$ the set of all phase points $x \in \mathbf{M}$ for which the trajectory of $x$ encounters infinitely many non-tangential collisions in both time directions. The trajectories of the points $x \in \mathbf{M} \setminus \mathbf{M}^\#$ are lines: the motion is linear and uniform, see the appendix of [Sz(1994)]. It is proven in lemmas A.2.1 and A.2.2 of [Sz(1994)] that the closed set $\mathbf{M} \setminus \mathbf{M}^\#$ is a finite union of hyperplanes. It is also proven in [Sz(1994)] that, locally, the two sides of a hyper-planar component of $\mathbf{M} \setminus \mathbf{M}^\#$ can be connected by a positively measured beam of trajectories, hence, from the point of view of ergodicity, in this paper it is enough to show that the connected components of $\mathbf{M}^\#$ entirely belong to one ergodic component. This is what we are going to do in this paper.

Denote by $\mathbf{M}^0$ the set of all phase points $x \in \mathbf{M}^\#$ the trajectory of which does not hit any singularity, and use the notation $\mathbf{M}^1$ for the set of all phase points $x \in \mathbf{M}^\#$ whose orbit contains exactly one, simple singularity. According to Proposition 2.7.8, the set $\mathbf{M}^\# \setminus (\mathbf{M}^0 \cup \mathbf{M}^1)$ is a countable union of smooth, codimension-two ($\geq 2$) submanifolds of $\mathbf{M}$, and, therefore, this set may be discarded in our study of ergodicity, please see also the properties of slim sets above. Thus, we will restrict our attention to the phase points $x \in \mathbf{M}^0 \cup \mathbf{M}^1$.

**2.9. The "Chernov-Sinai Ansatz".** An essential precondition for the Theorem on Local Ergodicity by Bálint–Chernov–Szász–Tóth (Theorem 4.4 of [B-Ch-Sz-T(2001)]) is the so called "Chernov-Sinai Ansatz" which we are going to formulate below. Denote by $\mathcal{SR}^+ \subset \partial\mathbf{M}$ the set of all phase points $x_0 = (q_0, v_0) \in \partial\mathbf{M}$ corresponding to singular reflections (a tangential or a double collision at time zero) supplied with the post-collision (outgoing) velocity $v_0$. It is well known that $\mathcal{SR}^+$ is a compact cell complex with dimension $2d - 3 = \dim\mathbf{M} - 2$. It is also known (see Lemma 4.1 in [K-S-Sz(1990)]) that for $\nu$-almost every phase point $x_0 \in \mathcal{SR}^+$ the forward orbit $S^{(0,\infty)}x_0$ does not hit any further singularity. (Here $\nu$ is the Riemannian volume of $\mathcal{SR}^+$ induced by the restriction of the natural Riemannian metric of $\mathbf{M}$.) The Chernov-Sinai Ansatz postulates that for $\nu$-almost every $x_0 \in \mathcal{SR}^+$ the forward orbit $S^{(0,\infty)}x_0$ is sufficient (geometrically hyperbolic).

**2.10. The Theorem on Local Ergodicity.** The Theorem on Local Ergodicity by Bálint–Chernov–Szász–Tóth (Theorem 4.4 of [B-Ch-Sz-T(2001)]) claims the following: Let $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$ be a semi-dispersive billiard flow with (2.1.1)–(2.1.2) and with the property that the smooth components of the boundary $\partial\mathbf{Q}$ of the configuration space are algebraic hyper-surfaces. (The cylindric billiards automatically fulfill this algebraicity condition.) Assume – further – that the Chernov-Sinai Ansatz holds true, and a phase point $x_0 \in (\mathbf{M} \setminus \partial\mathbf{M}) \cap \mathbf{M}^{\#}$ is given with the properties

(i) $S^{(-\infty,\infty)}x$ has at most one singularity,

and

(ii) $S^{(-\infty,\infty)}x$ is sufficient.

Then some open neighborhood $U_0 \subset \mathbf{M}$ of $x_0$ belongs to a single ergodic component of the flow $(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu)$. (Modulo the zero sets, of course.)

## 3. The Algebraic Approach

In this paper we would like to present a proof for the Theorem as self-contained as possible. For that reason, below we are inserting §3 and the first part of §4 of the paper [S-Sz(1999)]. Only small editions have been performed on the original text in order to make it fully compatible with the present article. At the end of this section we cite all other results proved in [S-Sz(1999)] that are deemed to be necessary for the present proof.

The aim of this section is to properly understand the algebraic relationship between the kinetic data of the billiard flow measured at different moments of time.

From now on we shall investigate orbit segments $S^{[0,T]}x_0$ $(T > 0)$ of the standard billiard ball flow $(\mathbf{M}, \{S^t\}, \mu, \vec{m}, L)$. (Here $L > 0$ is the size of the container torus $\mathbb{T}^\nu_L = \mathbb{R}^\nu / (L \cdot \mathbb{Z}^\nu)$ in which the balls of radius $r$ are moving.) We note here that $\mathbf{M} = \mathbf{Q} \times \mathbb{S}^{d-1}$, where the configuration space $\mathbf{Q}$ was defined in §2 and $\mathbb{S}^{d-1}$ is the velocity sphere

$$\mathbb{S}^{d-1} = \left\{ (v_1, \dots, v_N) \in \mathbb{R}^{\nu N} \middle| \sum_{i=1}^N m_i v_i = 0 \ \& \ \sum_{i=1}^N m_i \|v_i\|^2 = 1 \right\}$$

introduced in §2. Also note that in the geometric-algebraic considerations of the upcoming sections we do not use the factorization with respect to uniform spatial translations introduced in §2. Later on even the conditions $\sum_{i=1}^N m_i v_i = 0$ and $\sum_{i=1}^N m_i \|v_i\|^2 = 1$ will be dropped, cf. Remark 3.14 below.

The symbolic collision sequence of $S^{[0,T]}x_0$ is denoted by $\Sigma\left(S^{[0,T]}x_0\right) = (\sigma_1, \sigma_2, \dots, \sigma_n)$. The symbol $v_i^k = v_i^+(t_k) \in \mathbb{R}^\nu$ shall denote the velocity $\dot{q}_i$ of the $i$-th ball right *after* the $k$-th collision $\sigma_k = \{i_k, j_k\}$ $(1 \le i_k < j_k \le N)$ occurring at time $t_k$, $k = 1, \dots, n$. Of course, there is no need to deal with the velocities right *before* collisions, since $v_i^-(t_k) = v_i^+(t_{k-1}) = v_i^{k-1}$. As usual, $q_i^k \in \mathbb{T}^\nu_L = \mathbb{R}^\nu / (L \cdot \mathbb{Z}^\nu)$

denotes the position of the center of the $i$-th ball at the moment $t_k$ of the $k$-th collision.

Let us now fix the symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ $(n \geq 1)$, and explore the algebraic relationship between the data

$$\left\{q_i^{k-1},\, v_i^{k-1} \,|\, i = 1, \dots, N\right\}$$

and

$$\left\{q_i^{k},\, v_i^{k} \,|\, i = 1, \dots, N\right\},$$

$k = 1, \dots, n$. By definition, the data $\left\{q_i^0,\, v_i^0 \,|\, i = 1, \dots, N\right\}$ correspond to the initial (non-collision) phase point $x_0$. We also set $t_0 = 0$.

Since we would like to carry out arithmetic operations on these data, the periodic positions $q_i^k \in \mathbb{T}_L^\nu$ are not suitable for this purpose. Therefore, instead of studying the genuine orbit segments $S^{[0,T]}x_0$, we will deal with their Euclidean lifting as follows:

**Proposition 3.1.** Let $S^{[0,T]}x_0 = \{q_i(t),\, v_i(t) \,|\, 0 \leq t \leq T\}$ be an orbit segment as above, and assume that certain pre-images (Euclidean lifting) $\tilde{q}_i(0) = \tilde{q}_i^0 \in \mathbb{R}^\nu$ of the positions $q_i(0) \in \mathbb{T}_L^\nu = \mathbb{R}^\nu / (L \cdot \mathbb{Z}^\nu)$ are given. Then there is a uniquely defined, continuous, Euclidean lifting $\{\tilde{q}_i(t) \in \mathbb{R}^\nu \,|\, 0 \leq t \leq T\}$ of the given orbit segment that is an extension of the initial lifting $\left\{\tilde{q}_i^0 \,|\, i = 1, \dots, N\right\}$. Moreover, for every collision $\sigma_k$ there exists a uniquely defined integer vector $a_k \in \mathbb{Z}^\nu$ – named the adjustment vector of $\sigma_k$ – such that

$$\left\|\tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k\right\|^2 - 4r^2 = 0. \tag{3.2}$$

We shall call the orbit segment $\omega = \left\{\tilde{q}_i^k,\, v_i^k \,|\, i = 1, \dots, N;\, k = 0, \dots, n\right\}$ a *lifted orbit segment with the system of adjustment vectors* $\mathcal{A} = (a_1, \dots, a_n) \in \mathbb{Z}^{n\nu}$.
The proof of this proposition is straightforward. □

The next result establishes the already mentioned polynomial relationships between the kinetic data $\left\{\tilde{q}_i^{k-1},\, v_i^{k-1} \,|\, i = 1, \dots, N\right\}$ and $\left\{\tilde{q}_i^k,\, v_i^k \,|\, i = 1, \dots, N\right\}$, $k = 1, \dots, n$.

**Proposition 3.3.** Using all notations and notions from above, one has the following polynomial relations between the kinetic data at $\sigma_{k-1}$ and $\sigma_k$, $k = 1, \dots, n$. In order to simplify the notations, we write down these equations as if $\sigma_k$ were $\{1, 2\}$:

$$v_i^k = v_i^{k-1}, \quad i \notin \{1, 2\}, \tag{3.4}$$

$$m_1 v_1^k + m_2 v_2^k = m_1 v_1^{k-1} + m_2 v_2^{k-1} \tag{3.5}$$

(conservation of the momentum),

(3.6)
$$v_1^k - v_2^k = v_1^{k-1} - v_2^{k-1} - \frac{1}{2r^2}\left\langle v_1^{k-1} - v_2^{k-1};\, \tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right\rangle \cdot \left(\tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right)$$

(reflection of the relative velocity determined by the elastic collision),

$$\tilde{q}_i^k = \tilde{q}_i^{k-1} + \tau_k v_i^{k-1}, \quad i = 1, \ldots, N, \tag{3.7}$$

where the time slot $\tau_k = t_k - t_{k-1}$ $(t_0 := 0)$ is determined by the quadratic equation

$$\left\|\tilde{q}_1^{k-1} - \tilde{q}_2^{k-1} + \tau_k\left(v_1^{k-1} - v_2^{k-1}\right) - L\cdot a_k\right\|^2 = 4r^2. \tag{3.8}$$

The proof of this proposition is also obvious. □

**Remark.** Observe that the new velocities $v_1^k$ and $v_2^k$ can be computed from (3.5)-(3.6) as follows:

(3.9)
$$\begin{aligned} v_1^k &= v_1^{k-1} - \frac{m_2}{2r^2(m_1+m_2)}\left\langle v_1^{k-1} - v_2^{k-1}; \tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right\rangle \cdot \left(\tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right), \\ v_2^k &= v_2^{k-1} + \frac{m_1}{2r^2(m_1+m_2)}\left\langle v_1^{k-1} - v_2^{k-1}; \tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right\rangle \cdot \left(\tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right). \end{aligned}$$

We also note that the equations above even extend analytically to the case when one of the two masses, say $m_1$, is equal to zero:

$$\begin{aligned} v_1^k &= v_1^{k-1} - \frac{1}{2r^2}\left\langle v_1^{k-1} - v_2^{k-1}; \tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right\rangle \cdot \left(\tilde{q}_1^k - \tilde{q}_2^k - L\cdot a_k\right), \\ v_2^k &= v_2^{k-1}. \end{aligned} \tag{3.10}$$

We note that in our understanding the symbol $\langle\, .\, ;\, .\,\rangle$ denotes the Euclidean inner product of $\nu$-dimensional real vectors and $\|\, .\, \|$ is the corresponding norm.

## The Complexification of the Billiard Map

Given the pair $(\Sigma, \mathcal{A}) = (\sigma_1, \ldots, \sigma_n; a_1, \ldots, a_n)$, the equations (3.4)-(3.8) make it possible to iteratively compute the kinetic data $\left\{\tilde{q}_i^k,\, v_i^k \,|\, i = 1, \ldots, N\right\}$ by using the preceding data $\left\{\tilde{q}_i^{k-1},\, v_i^{k-1} \,|\, i = 1, \ldots, N\right\}$. Throughout these computations one only uses the field operations and square roots. (The latter one is merely used when computing $\tau_k$ as the root of the quadratic equation (3.8). Obviously, since – at the moment – we are dealing with genuine, real orbit segments, in this dynamically realistic situation the equations (3.8) have two distinct, positive real roots, and $\tau_k$ is the smaller one.) Therefore, the data $\left\{\tilde{q}_i^k,\, v_i^k \,|\, i = 1, \ldots, N\right\}$ can be expressed as certain algebraic functions of the initial data $\left\{\tilde{q}_i^0,\, v_i^0 \,|\, i = 1, \ldots, N\right\}$, and the arising algebraic functions merely contain field operations, square roots, the radius $r$ of the balls, the size $L > 0$ of the torus and, finally, the masses $m_i$ as constants.

Since these algebraic functions make full sense over the complex field $\mathbb{C}$ and, after all, our proof of the theorem requires the complexification, we are now going to complexify the whole system by considering the kinetic variables, the size $L$, and the masses as complex ones and retaining the polynomial equations (3.4)-(3.8). The fixed radius $r > 0$ plays the role of a constant, not a variable. However, due to the ambiguity of selecting a root of (3.8) out of the two, it proves to be important to explore first the algebraic framework of the relations (3.4)-(3.8) connecting the studied variables. This is what we do now.

### The Field Extension Associated With the Pair $(\Sigma, \mathcal{A})$

To avoid misunderstanding, we immediately stress that the field extensions $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A})$ to be defined below will also depend on a sequence $\vec{\tau} = (\tau_1, \dots, \tau_n)$ of field elements to be introduced iteratively in Definition 3.11 below. If we also want to emphasize the dependence of $\mathbb{K}$ on $\vec{\tau}$, then we will write $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A}) = \mathbb{K}_n(\Sigma, \mathcal{A}, \vec{\tau})$.

We are going to define the function field $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A})$ generated by all functions

$$\left\{(\tilde{q}_i^k)_j,\, (v_i^k)_j \,|\, i = 1, \dots, N;\; k = 0, \dots, n;\; j = 1, \dots, \nu\right\}$$

of the lifted orbit segments corresponding to the fixed parameters

$$(\Sigma, \mathcal{A}) = (\sigma_1, \dots, \sigma_n; a_1, \dots, a_n)$$

in such a way that the field $\mathbb{K}_n(\Sigma; \mathcal{A})$ incorporates all algebraic relations among these variables that are consequences of the equations (3.4)-(3.8). (Here the subscript $j$ denotes the $j$-th component of a $\nu$-vector.) In our setup the ground field (the coefficient field) of allowed coefficients is, by definition, the complex field $\mathbb{C}$. The precise definition of $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A})$ is

**Definition 3.11.** For $n = 0$ the field $\mathbb{K}_0 = \mathbb{K}_0(\emptyset; \emptyset)$ is the transcendental extension $\mathbb{C}(\mathcal{B})$ of the coefficient field $\mathbb{C}$ by the algebraically independent formal initial variables

$$\mathcal{B} = \left\{(\tilde{q}_i^0)_j,\, (v_i^0)_j,\, m_i,\, L \,|\, i = 1, \dots, N;\; j = 1, \dots, \nu\right\}.$$

Suppose now that $n > 0$ and the commutative field $\mathbb{K}_{n-1} = \mathbb{K}_{n-1}(\Sigma'; \mathcal{A}')$ has already been defined, where $\Sigma' = (\sigma_1, \dots, \sigma_{n-1})$, $\mathcal{A}' = (a_1, \dots, a_{n-1})$. Then we consider the quadratic equation $b_n \tau_n^2 + c_n \tau_n + d_n = 0$ in (3.8) with $k = n$ as a polynomial equation defining a new field element $\tau_n$ to be adjoined to the field $\mathbb{K}_{n-1} = \mathbb{K}_{n-1}(\Sigma'; \mathcal{A}')$. (Recall that the coefficients $b_n$, $c_n$, $d_n$ come from the field $\mathbb{K}_{n-1}$.)

There are two possibilities:

(i) The quadratic polynomial $b_n x^2 + c_n x + d_n$ is reducible over the field $\mathbb{K}_{n-1}$. Then we take $\mathbb{K}_n = \mathbb{K}_{n-1}$.

(ii) The polynomial $b_n x^2 + c_n x + d_n$ is irreducible over the field $\mathbb{K}_{n-1}$. Then we define $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A})$ as the extension of $\mathbb{K}_{n-1} = \mathbb{K}_{n-1}(\Sigma'; \mathcal{A}')$ by the root $\tau_n$ of this irreducible polynomial.

**Remark 3.12.** The importance of the field $\mathbb{K}_n$ is underscored by the fact that this field encodes all algebraic relations among the kinetic data that follow from the polynomial equations (3.4)-(3.8). Furthermore, Proposition 3.3 gives us an iterative computation rule for successively obtaining the kinetic variables

$$\left\{\tilde{q}_i^k,\, v_i^k \,|\, i = 1, \dots, N\right\},$$

$k = 0, \dots, n$. The field $\mathbb{K}_n = \mathbb{K}_n(\Sigma, \mathcal{A})$ is, after all, the algebraic framework of such computations. We note that – everywhere in this paper – the symbols $\langle\, .\,;\, .\,\rangle$ and $\|\,.\,\|^2$ do not refer to an Hermitian inner product but, rather, $\langle x; y\rangle = \sum_{j=1}^{\nu} x_j y_j$ and $\|x\|^2 = \sum_{j=1}^{\nu} x_j^2$, so that these expressions retain their polynomial form.

**Remark 3.13.** Later on it will be necessary to express each kinetic variable $\left(\tilde{q}_i^k\right)_j$ and $\left(v_i^k\right)_j$ as an algebraic function of the initial variables $\mathcal{B}$. This raises, however, an important question: Which one of the two roots $\tau_k$ of (3.8) should be considered during these computations? This is no problem if the polynomial $b_k x^2 + c_k x + d_k$, defining $\tau_k$, is irreducible over the field $\mathbb{K}_{k-1}$, because the two roots of this polynomial are algebraically equivalent over $\mathbb{K}_{k-1}$, and any of them can be chosen as $\tau_k$. However, if the polynomial $b_k x^2 + c_k x + d_k$ is reducible over the field $\mathbb{K}_{k-1}$, then it is necessary to make a decision and assign to one of the two roots (both in the field $\mathbb{K}_{k-1}$ now) the role of $\tau_k$. This is what we do. The result is the field $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A})$ endowed with a distinguished n-tuple $(\tau_1, \dots, \tau_n)$ of its elements. The field $\mathbb{K}_n$ with the distinguished $n$-tuple $\vec{\tau}$ is denoted by $\mathbb{K}_n = \mathbb{K}_n(\Sigma; \mathcal{A}; \vec{\tau})$. Then the algebraic object $\mathbb{K}_n(\Sigma; \mathcal{A}; \vec{\tau})$ completely controls the whole process of the iterative computation of the kinetic variables.

Note that in the genuine, real case the two roots of $b_k x^2 + c_k x + d_k$ are distinct, positive real numbers, and – by the nature of the billiard dynamics – the chosen value is always the smaller one.

Finally, we mention here a simple case when the polynomial $b_k x^2 + c_k x + d_k$ is reducible over the field $\mathbb{K}_{k-1}$. Namely, this takes place whenever $\sigma_{k-1} = \sigma_k$ and $a_{k-1} = a_k$. In that case the value $\tau_k = 0$ is clearly a solution of the above polynomial $b_k x^2 + c_k x + d_k$, and, being so, that polynomial is reducible over the field $\mathbb{K}_{k-1}$, see also 3.31–3.32 below.

**Remark 3.14.** The reader may wonder why we did not postulate the algebraic dependencies $\sum_{i=1}^{N} m_i v_i^k = 0$, $\sum_{i=1}^{N} m_i \|v_i^k\|^2 = 1$? The answer is the following: The definition of the neutral linear space together with its characterization via the

Connecting Path Formula (CPF, see Lemma 2.9 of [Sim(1992)-II] or Proposition 2.19 in [S-Sz(1999)]), i. e. the partial linearity of the billiard flow, are invariant under

(1) all uniform velocity translations (adding the same vector to all velocities) and
(2) time rescaling.

**Remark 3.15.** All fields occurring in this paper are only defined up to an isomorphism over the coefficient field $\mathbb{C}$. Therefore, the statements like "the field $\mathbb{K}_2$ is an extension of $\mathbb{K}_1$" should be understood as follows: $\mathbb{K}_1$ is a subfield of $\mathbb{K}_2$ after the natural identification of the generating variables bearing the same name. For the necessary notions, properties, and results from the theory of field extensions and Galois theory, the reader is kindly recommended to look up, for instance, the book by I. Stewart, [St(1973)].

**Remark 3.16.** As said before (see Definition 3.11), the collection $\mathcal{B}$ of elements of the field $\mathbb{K}_n(\Sigma;\mathcal{A};\vec{\tau})$ is a base of transcendence in that field over the subfield $\mathbb{C}$, and the degree of the extension $|\mathbb{K}_n(\Sigma;\mathcal{A};\vec{\tau}):\mathbb{C}(\mathcal{B})| = |\mathbb{K}_n:\mathbb{K}_0|$ is a power of two. Moreover, each of the following sets is a generator for the field $\mathbb{K}_n$:

(a) $\mathbb{C}\cup\{m_1,\dots,m_N,L\}\cup\big\{(\tilde{q}_i^0)_j,\,(v_i^0)_j \,|\, i=1,\dots,N;\ j=1,\dots,\nu\big\}\cup\{\tau_1,\dots,\tau_n\}$;
(b) $\mathbb{C}\cup\{m_1,\dots,m_N,L\}\cup\big\{(\tilde{q}_i^k)_j,\,(v_i^k)_j \,|\, i=1,\dots,N;\ j=1,\dots,\nu;\ k=0,\dots,n\big\}$.

**Remark 3.17.** The above procedure of constructing field extensions is closely related to the classical theory of geometric constructions by a straight edge and a compass, and this is not for surprise: The billiard trajectory is constructed by intersecting a straight line with a sphere and then reflecting it across the tangent hyperplane of the sphere. This is a sort of classical geometric construction in $\nu$ dimensions. For the details see, for instance, §57 and §60 of [VDW (1970)], or Chapter 5 of [St(1973)].

Let us fix the pair $(\Sigma;\mathcal{A})$ and the $n$-tuple $\vec{\tau}=(\tau_1,\dots,\tau_n)$ of elements of the field $\mathbb{K}_n(\Sigma,\mathcal{A})$. We are now defining a $(2\nu+1)N+1$-dimensional complex analytic manifold $\Omega=\Omega\,(\Sigma,\mathcal{A},\vec{\tau})$ and certain holomorphic functions $\tilde{q}_i^k,\,v_i^k:\Omega\to\mathbb{C}^\nu$, $m_i:\Omega\to\mathbb{C}$, $L:\Omega\to\mathbb{C}$, $\tau_k:\Omega\to\mathbb{C}$, $i=1,\dots,N$, $k=0,\dots,n$, ($\tau_0$ is not defined). But first we introduce

**Definition 3.18.** Define the domain $D=D\,(\Sigma,\mathcal{A},\vec{\tau})\subset\mathbb{C}^{(2\nu+1)N+1}$ as the set of all complex $(2\nu+1)N+1$-tuples

$$\big\{(\tilde{q}_i^0)_j,\,(v_i^0)_j,\,m_i,\,L \,|\, i=1,\dots,N;\ j=1,\dots,\nu\big\}$$

for which

(a) the leading coefficient $\left\|v_{i_k}^{k-1}-v_{j_k}^{k-1}\right\|^2$ and the discriminant of the quadratic equation (3.8) is never zero, $k=1,\dots,n$, (The latter condition is equivalent to

$$\left\langle v_{i_k}^{k-1}-v_{j_k}^{k-1};\,\tilde{q}_{i_k}^k-\tilde{q}_{j_k}^k-L\cdot a_k\right\rangle\neq 0,$$

see also Remark 3.29), and

(b) $m_{i_k} + m_{j_k} \neq 0$ for $k = 1, \ldots, n$,
provided that we carry out the iterative computation of the kinetic variables

$$\left\{ (\tilde{q}_i^k)_j,\, (v_i^k)_j \,|\, i = 1, \ldots, N;\; j = 1, \ldots, \nu;\; k = 0, \ldots, n \right\}$$

as governed by the polynomial equations (3.4)-(3.8) and by the fixed selection of elements $\vec{\tau}$.

**Remark 1.** When inverting the dynamics, a useful remark is that

$$\left\| v_{i_k}^{k-1} - v_{j_k}^{k-1} \right\|^2 = \left\| v_{i_k}^{k} - v_{j_k}^{k} \right\|^2,$$

and

$$\left\langle v_{i_k}^{k-1} - v_{j_k}^{k-1};\, \tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k \right\rangle = - \left\langle v_{i_k}^{k} - v_{j_k}^{k};\, \tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k \right\rangle.$$

**Remark 2.** Note that we require the validity of (a) (of 3.18) for any branch of the square root function, when computing the $\tau_k$'s with irreducible defining polynomials (3.8).

**Lemma 3.19.** The complement set $\mathbb{C}^{(2\nu+1)N+1} \setminus D$ is a proper, closed, algebraic subset of $\mathbb{C}^{(2\nu+1)N+1}$, especially a finite union of complex analytic submanifolds with codimension at least one, so the open set $D \subset \mathbb{C}^{(2\nu+1)N+1}$ is *connected and dense.*

**Proof.** We have the polynomial equations (3.4)-(3.8) making it possible to set up an algorithm for iteratively computing the kinetic variables measured at different times. The point is not just this iterative algorithm, but also its invertibility (time reversibility). The inverse process has similar algebraic properties, for it just means time reversal. One can easily prove by an induction on the length $n$ of $\Sigma$ that the iteratively defined dynamics determines a several-to-several mapping with maximum rank between a nonempty Zariski open set of the kinetic variables with superscript 1 and a nonempty Zariski open set of the kinetic variables measured right after the last reflection (with superscript $n$). (A Zariski open set is the complement of a closed algebraic set defined as the simultaneous zero set of finitely many polynomials.) Then one can go ahead with the induction from $n$ to $n+1$, because all possible obstructions to extending the process only occur on proper algebraic submanifolds of the Zariski open set of kinetic variables with superscript $n$. $\square$

**Definition 3.20.** We define $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$ as the set of all complex $(2\nu(n+1)N + N + n + 1)$-tuples

$$\omega =$$

$$\left( \left(\tilde{q}_i^k\right)_j,\, (v_i^k)_j,\, m_i,\, L,\, \tau_l \,\middle|\, i = 1, \ldots, N;\; k = 0, \ldots, n;\; j = 1, \ldots, \nu;\; l = 1, \ldots, n \right)$$

for which these coordinates

(1) are interrelated by the equations (3.4)–(3.8),
(2) respect the choices of $\tau_k$ prescribed by $\vec{\tau}$ (cf 3.11, and also 3.13),

and the vector

$$\vec{x}(\omega) = \Big( \left( \tilde{q}_i^0 \right)_j, \, (v_i^0)_j, \, m_i, \, L \Big| \, i = 1, \dots, N; \, j = 1, \dots, \nu \Big)$$

of initial data belongs to the set $D\,(\Sigma, \mathcal{A}, \vec{\tau})$ defined above.

It is clear now that $\tilde{q}_i^k, \, v_i^k : \Omega \to \mathbb{C}^\nu, \, m_i : \Omega \to \mathbb{C}, \, L : \Omega \to \mathbb{C}$, and $\tau_k : \Omega \to \mathbb{C}$ $(i = 1, \dots, N, \; k = 0, \dots, n)$ are holomorphic functions on $\Omega = \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$. The complex analytic manifold $\Omega$ endowed with the above holomorphic functions can justifiably considered as the complexification of the $(\Sigma, \mathcal{A}, \vec{\tau})$-iterated billiard map.

**Remark.** The careful reader has certainly done the following observation: The $\tau_k$'s figure in two different roles. They first denote successively chosen field elements of the extensions introduced in Definition 3.11, and secondly they also denote complex-valued functions on $\Omega$ (or multi-valued complex functions on $D$, cf. Definitions 3.20 and 3.18).

**Remark.** It is very likely that the manifold $\Omega$ is connected. Nevertheless, in our proof we do not need this connectedness, and are not going to pursue the goal of proving it.

## The Complex Neutral Space $\mathcal{N}(\omega)$

Fix a base point $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$. The tangent space $\mathcal{T}_\omega \Omega$ of $\Omega$ at $\omega$ consists of all complex vectors

$$x = \left( \delta \tilde{q}_i^0, \, \delta v_i^0, \, \delta m_i, \, \delta L \, | \, i = 1, \dots, N \right) \in \mathbb{C}^{(2\nu+1)N+1},$$

thus $\mathcal{T}_\omega \Omega$ can be naturally identified with the complex vector space $\mathbb{C}^{(2\nu+1)N+1}$. Set

$$\begin{aligned} \mathcal{N}(\omega) = \Big\{ x = \left( \delta \tilde{q}_i^0, \, \delta v_i^0, \, \delta m_i, \, \delta L \, | \, i = 1, \dots, N \right) \in \mathcal{T}_\omega \Omega \, \Big| \\ D_x(v_i^k) = D_x(m_i) = D_x(L) = 0 \text{ for } i = 1, \dots, N; \, k = 0, \dots, n \Big\}, \end{aligned} \tag{3.21}$$

where $D_x(\,.\,)$ denotes the directional differentiation in the direction of $x$. Clearly, the complex linear subspace $\mathcal{N}(\omega)$ of $\mathbb{C}^{(2\nu+1)N+1}$ is the proper complex analog of the neutral space of a genuine, real orbit segment, see §2.

**Remark.** Since $\delta v_i^0$, $\delta m_i$, and $\delta L$ must be equal to zero for a neutral vector $x \in \mathcal{N}(\omega)$, we shall simply write $x = (\delta\tilde{q}_1^0, \dots, \delta\tilde{q}_N^0)$ instead of indicating the zero entries.

The proof of the following proposition is completely analogous to that one for the real case, and therefore we omit it.

**Proposition 3.22.** For every tangent vector $x \in \mathcal{N}(\omega)$ $(\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau}))$ and $1 \le k \le n$ we have

$$D_x\left(\tilde{q}_{i_k}^{k-1} - \tilde{q}_{j_k}^{k-1}\right) = (\alpha_k(x) - \alpha_{k-1}(x)) \cdot \left(v_{i_k}^{k-1} - v_{j_k}^{k-1}\right), \tag{3.23}$$

$$D_x(t_k) = -\alpha_k(x), \tag{3.24}$$

and

$$D_x(\tau_k) = D_x(t_k - t_{k-1}) = \alpha_{k-1}(x) - \alpha_k(x), \tag{3.25}$$

where, by definition, $t_0 = 0 = \alpha_0(x)$. The functions $\alpha_k : \mathcal{N}(\omega) \to \mathbb{C}$ $(k = 1, \dots, n)$ are linear functionals. The name of the functional $\alpha_k$ is the *advance of the $k$-th collision* $\sigma_k = \{i_k, j_k\}$, see also Section 2.

This result markedly shows that, indeed, the vector space $\mathcal{N}(\omega)$ describes the linearity of the $(\Sigma, \mathcal{A}, \vec{\tau})$-iterated billiard map.

**Remark 3.26.** It is a matter of simple computation to convince ourselves that all assertions of Section 2 of the present paper and [Sim(1992)-II] pertaining to the neutral space $\mathcal{N}(\omega)$ remain valid for the complexified dynamics. Here we only point out the three most important statements out of those results:

(1) The vector space $\mathcal{N}(\omega)$ measures the ambiguity in recovering the orbit segment $\omega$ purely from its velocity history $\left\{v_i^k(\omega) : \ i = 1, \dots, N; \ k = 0, \dots, n\right\}$ and the outer geometric data $(\vec{m}, L)$. In other words, this means that if, locally in the phase space $\Omega$ (in a small open set), two phase points $\omega_1$ and $\omega_2$ have the same velocities $(v_i^k(\omega_1) = v_i^k(\omega_2), \ i = 1, \dots, N; \ k = 0, \dots, n)$ and geometric parameters $(\vec{m}, L)$, then the initial data of $\omega_1$ and $\omega_2$ can only differ by a spatial translation by a vector from the space $\mathcal{N}(\omega_1) = \mathcal{N}(\omega_2)$, and this statement is obviously reversible.

(2) The reflection laws for the neutral vectors are exactly the same as for velocities.

(3) The Connecting Path Formula (CPF, i. e. Lemma 2.9 of [Sim(1992)-II] or Proposition 2.19 in [S-Sz(1999)]) is applicable, and it enables us to compute the neutral space $\mathcal{N}(\omega)$ via solving a homogeneous system of linear equations (now over the field $\mathbb{C}$)

$$\sum_{k=1}^{n} \alpha_k \Gamma_{ik} = 0, \ i = 1, \dots, n + P_\Sigma - N, \tag{3.27}$$

where each equation in (3.27) is a $\nu$-dimensional complex vector equation, the coefficients $\Gamma_{ik} = \Gamma_{ik}(\omega) \in \mathbb{C}^\nu$ are certain linear combinations of relative velocities $v_i^k(\omega) - v_j^k(\omega)$ (the coefficients of those linear combinations are just fractional linear expressions of the masses, see 2.17 and 2.18 in [S-Sz(1999)]) and, finally, $P_\Sigma$ denotes the number of connected components of the collision graph of $\Sigma$. For a bit more detailed explanation of (3.27), see Remark 3.41 below.

**Remark 3.28.** We note here that the ordering of the moments of collisions $t(f_i)$ (which plays a significant role in the CPF) is no longer meaningful over the unordered field $\mathbb{C}$. However, the use of the corresponding inequalities in the CPF is purely technical/notational. Those inequalities only serve to introduce the combinatorial ordering of the collisions $f_i$ (which are $\sigma$'s) prescribed by the indexes of the symbolic collision sequence $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$.

**Remark 3.29.** As a straightforward consequence of the equation (3.6), we obtain that $v_i^{k-1} = v_i^k$ for $i = 1, \dots, N$ (i. e. the $k$-th reflection does not change the compound velocity) if and only if

$$\left\langle v_{i_k}^{k-1} - v_{j_k}^{k-1};\, \tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k \right\rangle = 0,$$

and this is just the case of a tangential collision. Easy calculation shows that the mentioned tangentiality occurs if and only if the two roots $\tau_k$ of (3.8) coincide.

It is obvious that $\dim_{\mathbb{C}} \mathcal{N}(\omega)$ is at least $\nu + 1$ (as long as not all velocities are the same, which we can assume, for $v_1^k(\omega) = \cdots = v_N^k(\omega)$ occurs on a set of codimension $\geq 2$), because the flow direction and the uniform spatial translations are necessarily contained in $\mathcal{N}(\omega)$.

**Definition 3.30. (Definition of sufficiency or "geometric hyperbolicity").**
The orbit segment $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ is said to be sufficient if and only if $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$.

Finally, the main result of [S-Sz(1999)] (Key Lemma 3.33 below) will use the following

**Definition 3.31.** We say that the triple $(\Sigma, \mathcal{A}, \vec{\tau})$ has Property (A) if and only if the following assertion holds:

For every pair of indices $1 \leq k < l \leq n$ for which $\sigma_k = \sigma_l$, $a_k = a_l$, and $\sigma_j \cap \sigma_k = \emptyset$ for $j = k+1, \dots, l-1$, we have that $\tau_l \not\equiv -\sum_{j=k+1}^{l-1} \tau_j$ in the algebraic selection $\vec{\tau}$ of the time slots. (Note that in this case the quadratic polynomial $b_l x^2 + c_l x + d_l$ is reducible over the field $\mathbb{K}_{l-1}$, because the time slot $\tau_l \equiv -\sum_{j=k+1}^{l-1} \tau_j \in \mathbb{K}_{l-1}$ is automatically a solution of (3.8).)

**Remark 3.32.** It is obvious that the schemes $(\Sigma, \mathcal{A}, \vec{\tau})$ corresponding to genuine, real billiard trajectory segments enjoy Property (A). Indeed, in those cases even the equations $\sigma_k = \sigma_l$, $a_k = a_l$, and $\sigma_j \cap \sigma_k = \emptyset$ $(j = k+1, \dots, l-1)$ cannot hold simultaneously.

## The Notion of (combinatorial) Richness. Richness Implies Sufficiency

The main result proved in [S-Sz(1999)] is

**Key Lemma 3.33 (Key Lemma 4.1 in [S-Sz(1999)]).** There exists a positive number $C(N)$ (depending merely on the number of balls $N \geq 2$) with the following property: If a symbolic sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ is $C(N)$-rich and the triplet $(\Sigma, \mathcal{A}, \vec{\tau})$ has Property (A) (see Definition 3.31), then $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$ (i. e. the trajectory is sufficient) for almost every $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$. (Or, equivalently, we prove that at least one of the polynomials $P_i(\vec{x})$ in (3.37) is not the zero polynomial, see also the "Dichotomy Corollary" 3.42.) The real version of this result is also valid. (For the definition of a $C$-rich symbolic sequence, see §2.)

**Remark 3.34.** Plainly, the exceptional zero measure subset of $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ mentioned in this lemma is actually a countable union of proper, analytic submanifolds.

**Remark 3.35.** In the course of proving the key lemma it was shown in [S-Sz(1999)] that the thresholds $C(N)$ $(N \geq 2)$ can actually be defined by the recursion $C(2) = 1$, $C(N) = \frac{N}{2} \cdot \max\{C(N-1), 3\}$ $(N \geq 3)$, that is, $C(2) = 1$, $C(N) = \dfrac{3N!}{2^{N-1}}$ for $N \geq 3$. We note that, as it has been proved in [K-S-Sz(1991)], for $N = 3$ the value $C(3) = 2$ already does the job. For the stronger results of the present paper we will need another, faster increasing threshold $C(N)$, see Lemma 3.51 and the remark following it.

## The Dichotomy. Algebraic Characterization of Sufficiency

**Lemma 3.36.** Consider the open, connected and dense subset $D$ of $\mathbb{C}^{(2\nu+1)N+1}$ formed by all possible complex $(2\nu+1)N+1$-tuples

$$\vec{x}(\omega) = \left( \left(\tilde{q}_i^0(\omega)\right)_j, \left(v_i^0(\omega)\right)_j, m_i(\omega), L(\omega) \middle| \, i = 1, \dots, N;\ j = 1, \dots, \nu \right), \ (\omega \in \Omega)$$

of initial data of the complex $(\Sigma, \mathcal{A}, \vec{\tau})$-dynamics, see Definitions 3.18 and 3.20. We claim that there are finitely many complex polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ $(\vec{x} \in \mathbb{C}^{(2\nu+1)N+1})$, determined by the discrete parameters $(\Sigma, \mathcal{A}, \vec{\tau})$, such that for every $\vec{x} \in D\,(\Sigma, \mathcal{A}, \vec{\tau})$ the following implication holds true:

$$\begin{aligned} &(\exists\, \omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau}) \text{ such that } \vec{x} = \vec{x}(\omega) \,\&\, \dim_{\mathbb{C}} \mathcal{N}(\omega) > \nu + 1) \\ &\qquad\qquad \Longrightarrow P_1(\vec{x}) = P_2(\vec{x}) = \dots = P_s(\vec{x}) = 0. \end{aligned} \tag{3.37}$$

We note here that the plynomials in (3.37) are actually real polynomials, see Proposition 3.41/c below. This note gains a particular importance when – in §4 – we switch from the complex case back to the real one.

**Remark 3.38.** The left-hand-side in the above implication (i. e. the premise) precisely says that *some* branch of the multiple-valued $(\Sigma, \mathcal{A}, \vec{\tau})$-dynamics with initial data $\vec{x}$ is not sufficient, see also Remark 2 to Definition 3.18.

**Proof.** This proof is a typical application of the Connecting Path Formula (see §2 here or Lemma 2.9 of [Sim(1992)-II]) and Proposition 3.4 of [Sim(1992)-II]. Namely, Proposition 3.4 of [Sim(1992)-II] asserts that

$$\dim_{\mathbb{C}}\mathcal{N}(\omega) = \nu \cdot P_\Sigma + \dim_{\mathbb{C}}\{\alpha_1, \dots, \alpha_n\}, \tag{3.39}$$

where, as said before, $P_\Sigma$ denotes the number of connected components of the collision graph of $\Sigma$, and $\{\alpha_1, \dots, \alpha_n\}$ is a shorthand notation for the complex linear space of all possible $n$-tuples $(\alpha_1(x), \dots, \alpha_n(x))$ of advances, $x \in \mathcal{N}(\omega)$.

**Remark 3.40.** It is worth noting here that our present formula (3.39) differs from its counterpart in Proposition 3.4 of [Sim(1992)-II] by an additional term $\nu$. This is, however, due to the fact that in the present approach we no longer have the reduction equation $\sum_{i=1}^N m_i \delta \tilde{q}_i = 0$. It follows easily from (3.39) that the sufficiency is equivalent to $\dim_{\mathbb{C}}\{\alpha_1, \dots, \alpha_n\} = 1$, i. e. sufficiency means that all advances are equal to the same functional. Note that $\dim_{\mathbb{C}}\{\alpha_1, \dots, \alpha_n\} = 1$ obviously implies $P_\Sigma = 1$.

A simple, but important consequence of the Connecting Path Formula is that the linear space $\{\alpha_1, \dots, \alpha_n\}$ is the solution set of a homogeneous system of linear equations (3.27).

**Remark 3.41.** Note that, when applying Lemma 2.9 of [Sim(1992-II)], the left-hand-side of the Connecting Path Formula has to be written as the relative velocity of the colliding balls multiplied by the advance of that collision. It follows immediately from the exposition of [Sim(1992)-II] that the equations of this sort (arising from all CPF's) are the *only* constraints on the advance functionals. The reason why this is true is that the fulfillment of all CPF's precisely means that the relative displacement (variation of position) of every pair of particles right before their new collision is parallel to the relative incoming velocity of these particles (see also Proposition 3.22), and this fact guarantees that the variation of the newly formed relative outgoing velocity will also be zero, just as required by (3.21).

Also note that the number $n + P_\Sigma - N$ of equations in (3.27) is equal to the number of all collisions $\sigma_k = \{i_k, j_k\}$ for which $i_k$ and $j_k$ are in the same connected component of the collision graph of $\{\sigma_1, \sigma_2, \dots, \sigma_{k-1}\}$.

The determinant $D(M)$ of every minor $M$ of the coefficient matrix of (3.27) is a homogeneous velocity polynomial and, therefore, $D(M)$ is a holomorphic function on $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$. Thus, there are finitely many velocity polynomials (actually, determinants) $R_1(\omega), \dots, R_s(\omega)$ such that their simultaneous vanishing follows from the non-sufficiency of the orbit segment $\omega$. By using the equations (3.4)-(3.8) (which define the complex dynamics recursively), each of these velocity polynomials $R_i(\omega)$ can be written as an algebraic function $f_i\left(\vec{x}(\omega)\right) = f_i(\vec{x})$ of the initial data $\vec{x}$, and these algebraic functions $f_i(\vec{x})$ only contain (finitely many) field operations and square roots. By using the canonical method of successive elimination of the square roots from the equation $f_i(\vec{x}) = 0$, one obtains a complex polynomial $P_i(\vec{x})$ such that

$$P_i(\vec{x}) = 0 \Longleftrightarrow f_i(\vec{x}) = 0.$$

We emphasize here that

(1) we understand the relation $f_i(\vec{x}) = 0$ in such a way that $f_i(\vec{x})$ becomes zero on *some* branch of the square root function when evaluating the multiple-valued algebraic function $f_i(\vec{x})$;

(2) the equivalence is claimed to only hold on $D$, cf. the formulation of Lemma 3.36.

The process of eliminating the square roots from an equation $f(\vec{x}) = 0$ (of the above type), however, requires a bit of clarification. The algebraic function $f$ has a natural representing element $\hat{f}$ in the field $\mathbb{K}_n$, see Definition 3.11. Consider the normal hull $\overline{\mathbb{K}}_n : \mathbb{K}_0$ of the field extension $\mathbb{K}_n : \mathbb{K}_0$. In other words, let $\overline{\mathbb{K}}_n$ be the splitting field (over $\mathbb{K}_0$) of the minimal polynomial $\phi(x) = x^k + \sum_{i=0}^{k-1} a_i x^i$ of a generator $\alpha \in \mathbb{K}_n$ of the extension $\mathbb{K}_n$ over the field $\mathbb{K}_0 = \mathbb{C}(\vec{x})$. (We note here that, since the extension $\mathbb{K}_n : \mathbb{K}_0$ is a simple algebraic extension, there exists such an element $\alpha \in \mathbb{K}_n$ with $\mathbb{K}_n = \mathbb{K}_0(\alpha)$, i. e. the field $\mathbb{K}_n$ can be obtained by a simple adjunction of the element $\alpha$ to the field $\mathbb{K}_0$.) The field $\overline{\mathbb{K}}_n$ contains all conjugates $\hat{f}^{(1)}, \ldots, \hat{f}^{(k)}$ of the element $\hat{f} = \hat{f}^{(1)}$ over the field $\mathbb{K}_0$. Let

$$N(\hat{f}) = \prod_{i=1}^{k} \hat{f}^{(i)} = a_0 \in \mathbb{K}_0 = \mathbb{C}(\vec{x}) \tag{3.41/a}$$

be the norm of the element $\hat{f}$ over $\mathbb{K}$. (For the elementary concepts and facts from Galois theory, the reader is kindly referred to the books [St(1973)], [VDW(1970)].) The element $N(\hat{f})$ has the form

$$N(\hat{f}) = \frac{P(\vec{x})}{Q(\vec{x})} \in \mathbb{K}_0 \tag{3.41/b}$$

with uniquely determined (up to scalar multipliers) relatively prime complex polynomials $P(\vec{x})$ and $Q(\vec{x})$.

Actually, as it directly follows from Proposition 3.3 and from the method of eliminating the square roots (presented above), in each denominator $Q_l(\vec{x})$ ($1 \leq l \leq s$) associated to the algebraic function $f_l(\vec{x})$ there are only factors of the type

$m_i + m_j$ or $\|v_i^k - v_j^k\|^2$, $1 \le i < j \le N$, $0 \le k \le n$. Therefore, none of the denominators $Q_l(\vec{x})$ vanishes on $D$. This explains property (2) formulated above.

It follows easily from the method of eliminating the square roots and from the actual form of the equations (3.4)–(3.8), that

**Proposition 3.41/c.** The coefficients of the polynomials $P_i(\vec{x})$ (in (3.37)) are actually real.

**Proof.** Let the automorphism $J_0 : \mathbb{K}_0 \to \mathbb{K}_0$ ($\mathbb{K}_0 = \mathbb{C}(\vec{x})$) be the natural extension of the conjugation automorphism of the ground field $\mathbb{C}$ to $\mathbb{K}_0$ by taking $J(x_i) = x_i$ for any of the initial variables $x_i$ in $\vec{x}$. Consider the tower of fields $\mathbb{C}(\vec{x}) = \mathbb{K}_0 \subset \mathbb{K}_1 \subset \cdots \subset \mathbb{K}_n$ constructed in 3.11, where $\mathbb{K}_n = \mathbb{K}_n(\Sigma, \mathcal{A}) = \mathbb{K}_0(\tau_1, \dots, \tau_n)$, $\mathbb{K}_m = \mathbb{K}_0(\tau_1, \dots, \tau_m)$ for $m = 0, 1, \dots, n$. Since the defining quadratic polynomial (3.8) of every $\tau_m$ (over the field $\mathbb{K}_{m-1}$) has only real coefficients (i. e. the coefficients come from the field $\mathbb{R}(\vec{x}; \tau_1, \dots, \tau_{m-1})$), we can extend the automorphism $J_0 : \mathbb{K}_0 \to \mathbb{K}_0$ step-by-step to an automorphism $J_n : \overline{\mathbb{K}}_n \to \overline{\mathbb{K}}_n$ by taking $J_n(\tau_m) = \tau_m$, $m = 1, \dots, n$. (Recall that $\overline{\mathbb{K}}_n$ denotes the normal hull of the extension $\mathbb{K}_n : \mathbb{K}_0$.) Since the algebraic function $f_j(\vec{x})$ ($j = 1, \dots, s$) only contains real constants, we get that every conjugate $\hat{f}_j^{(i)}$ of the representing element $\hat{f}_j \in \mathbb{K}_n$ of $f_j(\vec{x})$ is left fixed by the automorphism $J_n$. Thus the norm $N\left(\hat{f}_j\right) \in \mathbb{K}_0$ of (3.41/a–b) is also left fixed by $J_0$, meaning that the rational function

$$N\left(\hat{f}_j\right) = \frac{P_j(\vec{x})}{Q_j(\vec{x})}$$

has real coefficients. $\square$

So far we have seen that for each velocity determinant $R_i(\omega)$ ($i = 1, \dots, s$) one can find a *canonically determined* (up-to a nonzero constant multiplier, of course) real polynomial $P_i(\vec{x})$ with the following property: For every vector $\vec{x} \in D(\Sigma, \mathcal{A}, \vec{\tau})$ the equation $P_i(\vec{x}) = 0$ is equivalent to the equation $f_i(\vec{x}) = 0$ *on some branch of the square root function when evaluating the algebraic function* $f_i(\vec{x})$. Hence Lemma 3.36 follows. $\square$

**Corollary 3.42 (Dichotomy Corollary).** As a direct consequence of the definition of $N(\hat{f})$, we obtain that if $P \equiv 0$, then $\hat{f} = 0$, thus $f(\vec{x}) = 0$ for every $\vec{x} \in \mathbb{C}^{(2\nu+1)N+1}$ and on every branch of the square root function when evaluating the multiple-valued algebraic function $f(\vec{x})$. This also means that if all polynomials $P_1, \dots, P_s$ in Lemma 3.36 happen to be identically zero, then no orbit segment $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$ is sufficient.

Therefore, thanks to the algebraic feature of the dynamics, we have got a nice dichotomy here: *Either every* $(\Sigma, \mathcal{A}, \vec{\tau})$*-orbit segment is non-sufficient, or almost every such trajectory segment is sufficient.* $\square$

According to Lemma 3.36, in order to prove Key Lemma 3.33, it is enough to show that at least one of the polynomials $P_i(\vec{x})$ is nonzero. This is exactly what was done in [S-Sz(1999)] by an induction on the number $N \geq 2$.

**Lemma 3.43 (Lemma 4.8 in [S-Sz(1999)]).**
Pertaining to the polynomials $P_i(\vec{x})$ associated with $R_i(\omega) = f_i(\vec{x}(\omega))$ in Lemma 3.36, we assert the following:
If the algebraic function $f_i(\vec{x})$ is identically zero on a non-empty, open subset $B$ of $\mathbb{R}^{(2\nu+1)N+1}$, then $f_i$ is algebraically trivial, i. e. $\hat{f}_i = 0$ or, equivalently, $P_i(\vec{x}) \equiv 0$.

**Proof.** Denote the minimal polynomial of $\hat{f}_i$ over the base field $\mathbb{K}_0 = \mathbb{C}(\vec{x})$ by $M(x) = \sum_{j=0}^{k} c_j x^j$, $c_j \in \mathbb{K}_0$. Suppose, on the contrary, that $f_i$ yet vanishes on an open ball $B$ of $\mathbb{R}^{(2\nu+1)N+1}$. Since $f_i$ fulfills the identity $\sum_{j=0}^{k} c_j (f_i)^j \equiv 0$, the vanishing of $f_i$ on the ball $B$ implies that the element $c_0$ of the rational function field $\mathbb{K}_0$ vanishes on $B$, as well. Therefore, $c_0 = 0$ in $\mathbb{K}_0$, so $M(x)$ is divisible by $x$. The irreducibility of $M$ yields $M(x) = x$, i. e. $\hat{f}_i = 0$. Hence the lemma follows. □

## The Substitution $m_N = 0$. Derived Schemes

The next lemma describes the $(\Sigma, \mathcal{A}, \vec{\tau})$-orbit segments in the case when the $N$-th particle is "infinitely light" compared to the others, i. e. when $m_N = 0$ but $m_1 \cdot \ldots \cdot m_{N-1} \neq 0$.

**Key Lemma 3.44 (Lemma 4.9 in [S-Sz(1999)]).** Suppose that $m_N = 0$ and $m_1 \cdot \ldots \cdot m_{N-1} \neq 0$. Assume further that there exist indices $1 \leq p < q \leq n$ such that $N \in \sigma_p$, $N \in \sigma_q$ and $N \notin \sigma_j$ for $j = p+1, \ldots, q-1$. Then an orbit segment $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$ is sufficient (i. e. $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$) if

(1) the $\{1, 2, \ldots, N-1\}$-part

$$\left\{ \tilde{q}_i^k(\omega),\, v_i^k(\omega),\, m_i(\omega),\, L(\omega) \,\middle|\, i = 1, \ldots, N-1;\ k \in I_N \right\}$$

of the orbit segment is sufficient as an orbit segment of the particles $1, \ldots, N-1$ and

(2) the relative velocities $v_N^p(\omega) - v_{i_p}^p(\omega)$ and $v_N^{q-1}(\omega) - v_{i_q}^{q-1}(\omega)$ are not parallel.

(Here, as usual, $i_p$ ($i_q$) is the index of the ball colliding with the $N$-th particle at $\sigma_p$ ($\sigma_q$), and the index set $I_N \subset \{0, 1, \ldots, n\}$ contains 0 and those indices $i > 0$ for which $N \notin \sigma_i$.)

**Remark.** It is worth noting that, as it follows easily from the conditions, $v_N^p(\omega) = v_N^{q-1}(\omega)$.

**Proof.** First of all, we observe that the motion of the $N$-th ball with zero mass has absolutely no effect on the evolution of the $\{1,\dots,N-1\}$-part of the trajectory segment, and this statement is also valid for the time evolution of the $\{1,\dots,N-1\}$-part $\left\{\delta\tilde{q}_i^k \,|\, i=1,\dots,N-1\right\}$ of a neutral vector $\left(\delta\tilde{q}_1^0,\dots,\delta\tilde{q}_N^0\right)\in\mathcal{N}(\omega)$, see also the equations (3.4), (3.7)–(3.8) and (3.10).

Consider now an arbitrary neutral vector $\delta Q=\left(\delta\tilde{q}_1^0,\dots,\delta\tilde{q}_N^0\right)\in\mathcal{N}(\omega)$. According to (1) and the above principle, we can modify the neutral vector $\delta Q$ by a scalar multiple of the flow direction and by a uniform spatial displacement in such a way that $\delta\tilde{q}_i^0=0$ for $i=1,\dots,N-1$, and then $\delta\tilde{q}_i^k=0$ remains true for the whole orbit segment, $i=1,\dots,N-1$, $k=0,\dots,n$. The neutrality of $\delta Q$ with respect to the entire orbit segment means, however, that

$$(3.45)\qquad \delta\tilde{q}_N^p=\alpha_p\cdot\left(v_N^p(\omega)-v_{i_p}^p(\omega)\right)\ \text{and}\ \delta\tilde{q}_N^{q-1}=\alpha_q\cdot\left(v_N^{q-1}(\omega)-v_{i_q}^{q-1}(\omega)\right).$$

According to our hypothesis (2), the consequence of (3.45) and of the obvious equation $\delta\tilde{q}_N^p=\delta\tilde{q}_N^{q-1}$ is that $\alpha_p=\alpha_q=0$, and thus $\delta Q=0$. Hence the lemma follows. $\square$

The crucial part of the inductive proof of Key Lemma 3.33, i. e. the substitution $m_N=0$, requires some preparatory thoughts and lemmas. Assume that a combinatorial-algebraic scheme $(\Sigma,\mathcal{A},\vec{\tau})$ is given for the $N$-ball system $\{1,2,\dots,N\}$ so that Property (A) holds (see Definition 3.31). In the proof of Key Lemma 3.33 we want to use some combinatorial-algebraic $(N-1)$-schemes $(\Sigma',\mathcal{A}',\vec{\tau}')$ that

(i) govern the time evolution of the $\{1,2,\dots,N-1\}$-part of *some* $(\Sigma,\mathcal{A},\vec{\tau})$-orbit segments with an infinitely light $N$-th ball, i. e. $m_N=0$ and

(ii) enjoy Property (A).

In the construction of such $(N-1)$-schemes $(\Sigma',\mathcal{A}',\vec{\tau}')$, the so called *derived schemes*, we want to follow the guiding principles below, which also serve as the definition of the *derived schemes:*

**Definition 3.46.** We say that the $(N-1)$-scheme $(\Sigma',\mathcal{A}',\vec{\tau}')$ is a scheme derived from $(\Sigma,\mathcal{A},\vec{\tau})$ by putting $m_N=0$ if it is obtained as follows:

(1) First of all, we discard all symbols $\sigma_j$, $a_j$, and $\tau_j$ from $(\Sigma,\mathcal{A},\vec{\tau})$ that correspond to a $\sigma_j$ containing the label $N$. This also means that we retain the other symbols $\sigma_j$, $a_j$ without change: they only get re-indexed, due to the dropping of the other symbols;

(2) As far as the selection of $\vec{\tau}'$ is concerned, we want to retain *all algebraic relations* among the variables that are encoded in the original scheme $(\Sigma,\mathcal{A},\vec{\tau})$, i. e. all algebraic relations that follow from the scheme $(\Sigma,\mathcal{A},\vec{\tau})$ and from $m_N=0$;

(3) We want the derived scheme $(\Sigma',\mathcal{A}',\vec{\tau}')$ to enjoy Property (A), see 3.31.

It is straightforward that, following just the instructions in (1)-(2) above, one can easily construct such schemes $(\Sigma,\mathcal{A},\vec{\tau})$ which, perhaps, do not have Property (A).

**Remark 3.47.** Part (2) of the above definition means the following: Initially, the algebra of $(\Sigma, \mathcal{A}, \vec{\tau})$ consists

(1) of the variables

$$\left\{ \left(\tilde{q}_i^k\right)_j, \, (v_i^k)_j, \, m_i, \, L, \, \tau_l \,\middle|\, i = 1, \ldots, N; \; k = 0, \ldots, n; \; j = 1, \ldots, \nu; \; l = 1, \ldots, n \right\};$$

(2) of the equations (3.4-8), and finally

(3) of the choices of the "signs of the square roots" in the solutions of the quadratic equations (3.8) whenever the equations are reducible (cf. Remark 3.13).

To obtain the algebra of the derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$, we put first $m_N = 0$ and then $\tau_l' = \sum_{j=k+1}^{l} \tau_j$ whenever for some $1 \le k < l \le n$ we have $N \notin \sigma_k \cup \sigma_l$, $N \in \bigcap_{j=k+1}^{l-1} \sigma_j$ (here the indexing of the time slots $\tau_l'$ of $(\Sigma', \mathcal{A}', \vec{\tau}')$ is lacunary). After trivial substitutions, the quadratic equation for $\tau_l$ leads to the natural quadratic equation for $\tau_l'$, and then, finally, we can cancel all the variables $\left\{(\tilde{q}_N^k)_j, (v_N^k)_j, m_N \middle| \; 0 \le k \le n, 1 \le j \le \nu\right\}$, and all the equations they figure in as well. The collection of the remaining variables and equations with the corresponding vector $\vec{\tau}'$ make $(\Sigma', \mathcal{A}', \vec{\tau}')$.

The main point is the following one: Once $\tau_l$ is the root of a reducible polynomial, it is selected "artificially", following the instructions encoded in the scheme $\vec{\tau}$. Now we want to avoid the selection of the other root when defining the derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$. This is why (2) is a requirement, and not a corollary, and this is why, in Remark 3.47, we use the wording "the time slot $\tau_l'$ should be expressed".

The fact that the reducibility of the polynomial defining $\tau_l$ is inherited during a substitution $m_N = 0$ follows from a general principle: If some algebraic function in $\mathbb{K}_{l-1}$ (the discriminant of the quadratic polynomial defining $\tau_l$) is the square of another such algebraic expression, then this square relation remains even true if one annihilates a variable, i. e. considers that square relation on a smaller set. It is now the nature of our model that $m_N = 0$ immediately kills all variables involving the motion of the N-th particle, since the considered new discriminant of the quadratic equation defining $\tau_l'$ is obviously independent of those variables.

**Remark 3.48.** Note that the construction of such schemes $(\Sigma', \mathcal{A}', \vec{\tau}')$, by just requiring (1)-(2), is far from unique: It is well possible that, originally, for some $k$ we have $|\mathbb{K}_k : \mathbb{K}_{k-1}| = 2$ (i. e. the polynomial (3.8) determining $\tau_k$ is irreducible), but after the substitution $m_N = 0$ the corresponding field extension collapses to a trivial, degree-one extension. If this phenomenon of "reducible splitting" takes place, then we have to make our choice: Which one of the two (now distinguishable!) roots of the reducible (3.8) to include in $\vec{\tau}'$?

The requirement (A) for the derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$, however, ties our hands when selecting this root in the above scenario of "reducible splitting", and it actually implies the uniqueness of the derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$.

The phenomenon of this reducible splitting can also be well represented by the following, simple paradigm: Let $a \in \mathbb{C}$ be a complex parameter (playing the role of $m_N$). Consider the Riemann-leaf of the function $y = f(z) = \sqrt{z^2 + a}$, i. e. the algebraic variety defined by the equation $y^2 - z^2 - a = 0$. This is irreducible (indecomposable) for $a \neq 0$, and it splits into two components ($y - z = 0$ and $y + z = 0$) for $a = 0$. However, given a sequence $a_k \to 0$ ($a_k \neq 0$), every point $(z, y)$ on the curve $y^2 - z^2 = 0$ turns out to be a limit of a sequence $(z_k, y_k)$ with $y_k^2 - z_k^2 - a_k = 0$.

**Remark 3.49.** The paradigm just given also makes it possible to illustrate why (2) in Definition 3.46 is a requirement and not a corollary. Indeed, consider the field $\mathbb{K}_0 = \mathbb{C}(z, m_N)$, i. e. $\mathbb{K}_0$ is the complex field extended by two transcendental elements $z$ and $m_N$. Let $\mathbb{K}_1 := \mathbb{K}_0(\tau_1) = \mathbb{K}_0(\sqrt{z^2 + 1 + m_N})$ be the extension of $\mathbb{K}_0$ by the solution $\tau_1$ of the irreducible equation $\tau_1^2 - z^2 - 1 - m_N = 0$. Let us now extend $\mathbb{K}_1$ by the root $\tau_2$ of the quadratic polynomial $x^2 - 5\tau_1 x + 6\tau_1^2$. This polynomial is reducible over the field $\mathbb{K}_1$: its two roots are $2\tau_1$ and $3\tau_1$. Now the extended field $\mathbb{K}_2$ will be just equal to $\mathbb{K}_1$, and the selection of a designated root (as $\tau_2$) of the above polynomial amounts to selecting either $2\tau_1$ or $3\tau_1$. Observe that the root of the considered reducible polynomial could have been an arbitrary rational expression of the elements $z$, and $\tau_1$. In a general reducible step, such a rational expression shows the unique way of computing the actual $\tau_n$ from the initial data and the $\tau_i$'s with $i = 1, \dots, n-1$. However, when assigning actual complex values to the field element $\tau_1 = \sqrt{z^2 + 1 + m_N}$, we have two possibilities. Summing everything up, our selection does not mean giving values to these $\tau$'s but, rather, telling if we take (formally) $2\tau_1$ or $3\tau_1$ as $\tau_2$. (This selection determines the discrete algebraic structure, which becomes transparent if we think about repeated $\sigma$'s with identical adjustment vectors.) When defining the derived scheme this means that once we have chosen $3\sqrt{z^2 + 1 + m_N}$ (instead of $2\sqrt{z^2 + 1 + m_N}$) we still want to choose $3\sqrt{z^2 + 1}$ after the substitution $m_N = 0$, and do not want to switch suddenly to the other root $2\sqrt{z^2 + 1}$.

Finally, the last but one essential result of §4 of [S-Sz(1999)] that we will need here (and cite without proof) is

**Proposition 3.50 (Corollary 4.35 in [S-Sz(1999)]).** For every triple $(\Sigma, \mathcal{A}, \vec{\tau})$ with Property (A) and $N \geq 3$, there exists at least one derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$.

The closing part of this section contains a slightly modified version of Lemma 4.43 from [S-Sz(1999)]. The reason why we had to modify the recursion for the sequence $C(N)$ (from $C(N) = (N/2) \cdot \max\{C(N-1), 3\}$ to $C(N) = (N/2) \cdot \{2C(N-1) + 1\}$) is that our Corollary 4.4 (below) requires $(2C(N) + 1)$-richness instead of the usual $C(N)$-richness. In the present paper the sequence $C(N)$ always denotes the one defined by the recursion in Lemma 3.51 instead of the one defined in Lemma 4.43 of [S-Sz(1999)]. This should not cause any confusion.

We note that the upcoming lemma is purely combinatorial.

**Lemma 3.51.** Define the sequence of positive numbers $C(N)$ recursively by taking $C(2) = 1$ and $C(N) = (N/2)\cdot\{2C(N-1)+1\}$ for $N \geq 3$. Let $N \geq 3$, and suppose that the symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ for $N$ particles is $C(N)$-rich. Then we can find a particle, say the one with label $N$, and two indices $1 \leq p < q \leq n$ such that

(i) $N \in \sigma_p \cap \sigma_q$,

(ii) $N \notin \bigcup_{j=p+1}^{q-1} \sigma_j$,

(iii) $\sigma_p = \sigma_q \implies (\exists j)\ \ (p < j < q \,\&\, \sigma_p \cap \sigma_j \neq \emptyset)$, and

(iv) $\Sigma'$ is $(2C(N-1)+1)$-rich on the vertex set $\{1, \dots, N-1\}$.

(Here, just as in the case of derived schemes, we denote by $\Sigma'$ the symbolic sequence that can be obtained from $\Sigma$ by discarding all edges containing $N$.)

**Proof.** The hypothesis on $\Sigma$ implies that there exist subsequences $\Sigma_1, \dots, \Sigma_r$ of $\Sigma$ with the following properties:

(1) For $1 \leq i < j \leq r$ every collision of $\Sigma_i$ precedes every collision of $\Sigma_j$,

(2) the graph of $\Sigma_i$ $(1 \leq i \leq r)$ is a tree (a connected graph without loop) on the vertex set $\{1, \dots, N\}$, and

(3) $r \geq C(N)$.

Since every tree contains at least two vertices with degree one and $C(N) = (N/2)\cdot\{2C(N-1)+1\}$, there is a vertex, say the one labeled by $N$, such that $N$ is a degree-one vertex of $\Sigma_{i(1)}, \dots, \Sigma_{i(t)}$, where $1 \leq i(1) < \dots < i(t) \leq r$ and $t \geq 2C(N-1)+1$. Thus (iv) obviously holds.

Let $\sigma_{p'}$ the edge of $\Sigma_{i(1)}$ that contains $N$ and, similarly, let $\sigma_{q'}$ be the edge of $\Sigma_{i(t)}$ containing the vertex $N$. Then the fact $t \geq 3$ ensures that the following properties hold:

(i)' $N \in \sigma_{p'} \cap \sigma_{q'}$,

(iii)' $\sigma_{p'} = \sigma_{q'} \implies \exists j\ p' < j < q' \,\&\, \sigma_{p'} \cap \sigma_j \neq \emptyset,\ \sigma_j \neq \sigma_{p'}$.

Let $\sigma_p$, $\sigma_q$ $(1 \leq p < q \leq n)$ be a pair of edges $\sigma_{p'}$, $\sigma_{q'}$ $(1 \leq p' < q' \leq n)$ fulfilling (i)' and (iii)' and having the minimum possible value of $q' - p'$. Elementary inspection shows that then (ii) must also hold for $\sigma_p$, $\sigma_q$. Lemma 3.51 is now proved. $\square$

**Remark.** One easily checks that the exact solution of the above recursion is

$$C(N) = \frac{N!}{2} \cdot \sum_{k=1}^{N-1} \frac{1}{k!} \sim \frac{e-1}{2} N!.$$

However, we will not need this estimation in the future.

## 4. Non-sufficiency Occurs on a Codimension-two Set The Case $\nu \geq 3$

Let us fix a triplet $(\Sigma, \mathcal{A}, \vec{\tau})$ of the discrete (combinatorial) orbit structure with Property (A) (just as in the previous section), and assume that $\Sigma = (\sigma_1, \dots, \sigma_n)$ is $C(N)$-rich, i. e. it contains at least $C(N)$ consecutive, connected collision graphs. We also consider the complex analytic manifold $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ of all complex $(\Sigma, \mathcal{A}, \vec{\tau})$-orbits $\omega$ (Definition 3.20) and the open, dense, connected domain $D\,(\Sigma, \mathcal{A}, \vec{\tau}) \subset \mathbb{C}^{(2\nu+1)N+1}$ of all allowable initial data $\vec{x} = \vec{x}(\omega)$, see Definition 3.18. Let, finally, $Q(\vec{x})$ be a common *irreducible* divisor of the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ from (3.37). (If such a common divisor exists.) In this section we will need several results about such common irreducible divisors $Q(\vec{x})$ of the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$.

The first of them, as it is classically known from algebraic geometry (see, for example, [M(1976)]), is that the solution set $V = \{Q(\vec{x}) = 0\}$ of the equation $Q(\vec{x}) = 0$ is a so called irreducible (or, indecomposable) complex algebraic variety of codimension 1 in $\mathbb{C}^{(2\nu+1)N+1}$, which means that $V$ is not the union of two, proper algebraic sub-varieties. Secondly, the smooth part $S$ of $V$ turns out to be a connected complex analytic manifold, while the non-smooth part $V \setminus S$ of $V$ is a complex algebraic variety of dimension strictly less that $\dim V = (2\nu + 1)N$, see [M(1976)]. Finally, if a polynomial $R(\vec{x})$ vanishes on $V$, then $Q(\vec{x})$ must be a divisor of $R(\vec{x})$. (The last statement is a direct consequence of Hilbert's Theorem on Zeroes, see again [M(1976)].)

The first result, specific to our current dynamical situation, is

**Proposition 4.1.** The polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ of (3.37) are homogeneous in the masses $m_1, \dots, m_N$ and, consequently, any common divisor $Q(\vec{x})$ of these polynomials is also homogeneous in the masses.

**Proof.** It is clear that the complex dynamics encoded in the orbits $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ only depends on the ratios of masses $m_2/m_1$, $m_3/m_1$, $\dots$, $m_N/m_1$. Consequently, all algebraic functions $f_i(\vec{x})$ $(i = 1, \dots, s)$ featuring the proof of Lemma 3.36 above are homogeneous of degree 0 in the masses. Since the rational function

$$\frac{P_i(\vec{x})}{Q_i(\vec{x})} \in \mathbb{K}_0 = \mathbb{C}(\vec{x})$$

is the product $\alpha$ of all conjugates of $f_i(\vec{x})$ (see the proof of Lemma 3.36), we get that $\frac{P_i(\vec{x})}{Q_i(\vec{x})}$ is also homogeneous of degree 0 in the variables $m_1, \dots, m_N$. Then elementary algebra yields that both $P_i(\vec{x})$ and $Q_i(\vec{x})$ are homogeneous (of the same degree) in the masses. Since any factor of a homogeneous polynomial is easily seen to be also homogeneous, we get that the common divisor $Q(\vec{x})$ of $P_1(\vec{x})$, $\dots$, $P_s(\vec{x})$ is also homogeneous in the variables $m_1, \dots, m_N$. $\square$

Our next result, specific to our dynamics, that will be needed later is

**Proposition 4.2.** Let $\nu \geq 3$, $(\Sigma, \mathcal{A}, \vec{\tau})$ be a discrete orbit structure with Property (A) and a $C(N)$-rich symbolic collision sequence $\Sigma = (\sigma_1, \ldots, \sigma_n)$. Denote by $P_1(\vec{x}), \ldots, P_s(\vec{x})$ the polynomials of (3.37) just as before, and let $Q(\vec{x})$ be a common irreducible divisor of $P_1(\vec{x}), \ldots, P_s(\vec{x})$. (If such a common divisor exists.) Let, finally, $\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$ be an extended discrete orbit structure with Property (A) and an extended collision sequence $\overline{\Sigma} = (\sigma_0, \sigma_1, \ldots, \sigma_n)$. We claim that the irreducible (indecomposable) solution set $V$ of the equation $Q(\vec{x}) = 0$ can not even locally coincide with any of the following singularity manifolds $C$ defined by one of the following equations:

(1) $\|v_{i_0}^0 - v_{j_0}^0\|^2 = 0$,

(2) $\langle v_{i_0}^0 - v_{j_0}^0;\, \tilde{q}_{i_0}^0 - \tilde{q}_{j_0}^0 - L \cdot a_0 \rangle = 0$,

(3) $m_{i_0} + m_{j_0} = 0$,

i. e. the irreducible polynomial $Q(\vec{x})$ is not equal to any of the (irreducible) polynomials on the left-hand-sides of (1)—(3). (These equations feature Definition 3.18 of the domain $D\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$.) Consequently, the open subset $V \cap D\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$ of $V$ is connected and dense in $\mathbb{C}^{(2\nu+1)N+1}$.

**Remark.** The first point where we (implicitly) use the condition $\nu \geq 3$ is the irreducibility of the polynomial $\|v_{i_0}^0 - v_{j_0}^0\|^2$ on the left-hand-side of (1). Indeed, in the case $\nu = 2$ this polynomial splits as

$$\begin{aligned}\|v_{i_0}^0 - v_{j_0}^0\|^2 \equiv &\left[\left(v_{i_0}^0\right)_1 - \left(v_{j_0}^0\right)_1 + \sqrt{-1}\left(\left(v_{i_0}^0\right)_2 - \left(v_{j_0}^0\right)_2\right)\right] \\ &\cdot \left[\left(v_{i_0}^0\right)_1 - \left(v_{j_0}^0\right)_1 - \sqrt{-1}\left(\left(v_{i_0}^0\right)_2 - \left(v_{j_0}^0\right)_2\right)\right].\end{aligned}$$

However, it is easy to see that, in the case $\nu \geq 3$, all polynomials on the left of (1)–(3) are indeed irreducible.

**Proof.** First of all, we slightly reformulate the negation of the statement of the proposition.

Fix one of the three equations of (1)–(3) above, and denote the irreducible polynomial on its left-hand-side by $R(\vec{x})$. By using the quadratic (or linear) equation $R(\vec{x}) = 0$, we eliminate one variable $x_j$ out of $\vec{x}$ by expressing it as an algebraic function $x_j = g(\vec{y})$ of the remaining variables $\vec{y}$ of $\vec{x}$, so that the algebraic function $g$ only contains finitely many field operations and (at most one) square root. After this elimination $x_j = g(\vec{y})$, the meaning of $R(\vec{x}) \equiv Q(\vec{x})$ (i. e. the negation of the assertion of the proposition) is that all algebraic functions $f_i(\vec{x}) \equiv \tilde{f}_i(\vec{y})$ of (3.37) ($i = 1, \ldots, s$, constructed for $(\Sigma, \mathcal{A}, \vec{\tau})$, not for $\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$) are identically zero in terms of $\vec{y}$, meaning that every complex orbit segment $\omega \in \Omega\left(\Sigma, \mathcal{A}, \vec{\tau}\right)$, with the initial data $\vec{x}(\omega)$ in the solution set of $R(\vec{x}) = 0$, is non-sufficient, see also the "Dichotomy Corollary" 3.42. Thus, the negation of the proposition means that no orbit segment $\omega \in \Omega\left(\Sigma, \mathcal{A}, \vec{\tau}\right)$ in the considered singularity is sufficient.

Now we carry out an induction on the number of balls $N$ quite in the spirit of the proof of Key Lemma 4.1 of [S-Sz(1999)]. Indeed, the statement of the proposition is obviously true in the case $N = 2$, for in that case there are no non-sufficient (complex) trajectories $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$, i. e. the greatest common divisor of the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ is 1.

Assume now that $N \geq 3$, $\nu \geq 3$, and the statement of Proposition 4.2 has been proven for all values $N' < N$. Suppose, however, that the statement is false for some $(\Sigma, \mathcal{A}, \vec{\tau})$ and extension $\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$ with $N$ balls and Property (A), i. e. that there exists a common irreducible divisor $Q(\vec{x})$ of all the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$, and $Q(\vec{x})$ happens to be one of the irreducible polynomials on the left-hand-side of (1), (2), or (3). By using the $C(N)$-richness of $\Sigma = (\sigma_1, \dots, \sigma_n)$, we select a suitable label $k_0 \in \{1, 2, \dots, N\}$ for the substitution $m_{k_0} = 0$ along the lines of Lemma 3.51 above, by also ensuring the existence of the derived schemes $(\Sigma', \mathcal{A}', \vec{\tau}')$ and $\left(\overline{\Sigma}', \overline{\mathcal{A}}', \vec{\rho}'\right)$ for the $(N-1)$-ball-system $\{1, \dots, N\} \setminus \{k_0\}$ and properties (1)–(4) (of Lemma 3.51) for $\Sigma'$, see Proposition 3.50 and Lemma 3.51 above. Denote by $\tilde{Q}(\vec{x})$ the polynomial obtained from $Q(\vec{x})$ after the substitution $m_{k_0} = 0$.

**Lemma 4.3.** The polynomial $\tilde{Q}(\vec{x})$ is not constant.

**Proof.** Assume that $\tilde{Q}(\vec{x}) \equiv c \in \mathbb{C}$. The case $c = 0$ means that $m_{k_0}$ is a divisor of $Q(\vec{x})$, thus $m_{k_0} \equiv Q(\vec{x})$ (for $Q(\vec{x})$ is irreducible), which is impossible, since $Q(\vec{x})$ has to be one of the polynomials on the left-hand-side of (1), (2), or (3).

If, however, $c \neq 0$, then we have that $Q(\vec{x}) \equiv c + m_{k_0} S(\vec{x})$ with some nonzero polynomial $S(\vec{x})$. ($S(\vec{x})$ has to be non-zero, otherwise $Q(\vec{x})$ would be a constant, not an irreducible polynomial.) However, this contradicts to the proved homogeneity of the polynomial $Q(\vec{x})$ in the masses, see Proposition 4.1 above. This finishes the proof of the lemma. $\square$

**Remark.** If one takes a look at the equations (1), (2), (3), he immediately realizes that either $\tilde{Q}(\vec{x}) \equiv Q(\vec{x})$ (when $Q(\vec{x})$ is the polynomial on the left-hand-side of (1) or (2), or $Q(\vec{x}) \equiv m_{i_0} + m_{j_0}$ and $k_0 \notin \{i_0, j_0\}$), or $\tilde{Q}(\vec{x}) \equiv m_{i_0}$ when $Q(\vec{x}) \equiv m_{i_0} + m_{j_0}$ and $k_0 = j_0$. In this way one can directly and easily verify Lemma 4.3 without the above "involved" algebraic proof. The reason why we still included the above proof is that later on in this section (in the proof of Sub-lemma 4.6) we will need the idea of the presented proof.

The next lemma will use

**Definition 4.4.** Suppose that two indices $1 \leq p < q \leq n$ and two labels of balls $i, j \in \{1, \dots, N\}$ are given with the additional requirement that if $i = j$, then $i \in \bigcup_{l=p+1}^{q-1} \sigma_l$. Following the proof of Lemma 3.36, denote by $Q_1(\vec{x}), Q_2(\vec{x}), \dots, Q_\nu(\vec{x})$ ($\vec{x} \in \mathbb{C}^{(2\nu+1)N+1}$) the polynomials with the property that for every vector of initial data $\vec{x} \in D(\Sigma, \mathcal{A}, \vec{\tau})$ and for every $k$, $k = 1, \dots, \nu$, the following equivalence holds true:

$$\left(\exists\, \omega \in \Omega \text{ such that } \vec{x}(\omega) = \vec{x} \;\&\; \left(v_i^p(\omega)\right)_k = \left(v_j^{q-1}(\omega)\right)_k\right)$$
$$\iff Q_k(\vec{x}) = 0.$$

Our next lemma is a strengthened version of Lemma 4.39 of [S-Sz(1999)]:

**Lemma 4.5.** Assume that the combinatorial-algebraic scheme $(\Sigma, \mathcal{A}, \vec{\tau})$ has Property (A), and use the assumptions and notations of the above definition.

We claim that the polynomials $Q_1(\vec{x})$, $Q_2(\vec{x})$, $\dots$, $Q_\nu(\vec{x})$ do not have any nonconstant common divisor. As a consequence, the equality $v_i^p(\omega) = v_j^{q-1}(\omega)$ only takes place on an algebraic variety with at least two codimensions.

**Remark.** Lemma 4.39 of [S-Sz(1999)] just asserted that at least one of the above polynomials $Q_k(\vec{x})$ is nonzero. Then, by the permutation symmetry of the components $k \in \{1, \dots, \nu\}$, all of these polynomials are actually nonzero.

**Proof.** Induction on the number $N \geq 2$.

1. Base of the induction, $N = 2$: First of all, by performing the substitution $L = 0$, we can annihilate all adjustment vectors, see the proofs of parts 4/a–4/b in Main Lemma 4.21 of [S-Sz(1999)], in accordance with statement (VI) of that lemma. Then an elementary inspection shows that for any selection of *positive real* masses $(m_1, m_2)$, indeed, the equality $v_i^p(\omega) = v_j^{q-1}(\omega)$ only occurs on a manifold with $\nu - 1$ $(\geq 2)$ codimensions in the section $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau}, \vec{m})$ of $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ corresponding to the selected masses, since any trajectory segment of a two-particle system with positive masses and $\mathcal{A} = 0$ has a very nice, totally real (and essentially $\nu$-dimensional) representation in the relative coordinates of the particles: The consecutive, elastic bounces of a point particle moving uniformly inside a ball of radius $2r$ of $\mathbb{R}^\nu$. Therefore, the statement of the lemma is true for $N = 2$.

Assume now that $N \geq 3$, and the lemma has been successfully proven for all smaller numbers of balls. By re-labeling the particles, if necessary, we can achieve that

(i) $N \neq i$, $N \neq j$ and

(ii) if $i = j$, then the ball $i$ has at least one collision between $\sigma_p$ and $\sigma_q$ with a particle different from $N$.

For the fixed combinatorial scheme $(\Sigma, \mathcal{A}, \vec{\tau})$, select a derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$ corresponding to the substitution $m_N = 0$, see Definition 3.46 and Proposition 3.50 above.

Our induction step is going to be a proof by contradiction. Assume, therefore, that the nonzero polynomials $Q_1(\vec{x})$, $Q_2(\vec{x})$, $\dots$, $Q_\nu(\vec{x})$ do have a common irreducible divisor $R(\vec{x})$. According to Proposition 4.1 above, the (irreducible) polynomial $R(\vec{x})$ is homogeneous in the variables $m_1, \dots, m_N$. Denote by $\tilde{R}(\vec{x})$ the polynomial that we obtain from $R(\vec{x})$ after the substitution $m_N = 0$. Similarly to Lemma 4.3 above, we claim

**Sub-lemma 4.6.** The polynomial $\tilde{R}(\vec{x})$ is not constant.

**Remark.** The reason why we cannot simply apply Lemma 4.3 is that in the proof of that lemma we used the assumption that the irreducible polynomial $Q(\vec{x})$ was one of the polynomials on the left-hand-side of (1), (2), or (3) of Proposition 4.2. Right here we do not have such an assumption.

**Proof.** Suppose that $\tilde{R}(\vec{x}) \equiv c$, where $c \in \mathbb{C}$ is a constant, i. e. $R(\vec{x}) \equiv c + m_N \cdot S(\vec{x})$. In the case $c = 0$ the polynomial $m_N \equiv R(\vec{x})$ would be a common divisor of all the polynomials $Q_1(\vec{x})$, $Q_2(\vec{x})$, $\dots$, $Q_\nu(\vec{x})$, meaning that in the considered $N$-ball system $\Omega\left(\Sigma, \mathcal{A}, \vec{\tau}\right)$ the equation $m_N(\omega) = 0$ implies the equality $v_i^p(\omega) = v_j^{q-1}(\omega)$. This, in turn, means that in the $(N-1)$-ball system $\{1, \dots, N-1\}$ (with the discrete algebraic scheme $(\Sigma', \mathcal{A}', \vec{\tau}'))$ the equality $v_i^p(\omega) = v_j^{q-1}(\omega)$ is an identity, thus contradicting to the induction hypothesis.

Therefore $c \neq 0$, and in the expansion $R(\vec{x}) \equiv c + m_N \cdot S(\vec{x})$ of the irreducible polynomial $R(\vec{x})$ we certainly have $S(\vec{x}) \not\equiv 0$, and this means that $R(\vec{x})$ is not homogeneous in the mass variables, thus contradicting to Proposition 4.1. This finishes the proof of the sub-lemma. $\square$

**Finishing the proof of Lemma 4.5.** Denote by $\tilde{Q}_k(\vec{x})$ the polynomial that we obtain from $Q_k(\vec{x})$ after the substitution $m_N = 0$ $(k = 1, \dots, \nu)$, and by $T_k(\vec{x})$ the polynomial constructed for the $(N-1)$-ball system $\{1, \dots, N-1\}$ (with the discrete algebraic scheme $(\Sigma', \mathcal{A}', \vec{\tau}'))$ along the lines of Lemma 3.36, describing the event $v_i^p(\omega) = v_j^{q-1}(\omega)$ in this subsystem $(k = 1, \dots, \nu)$. It follows from the induction hypothesis that the zero set $W_k$ of the polynomial $\tilde{Q}_k(\vec{x})$ (in the phase space of the $(N-1)$-ball system $(\Sigma', \mathcal{A}', \vec{\tau}'))$ has a codimension-two intersection with the singularities of the $(\Sigma', \mathcal{A}', \vec{\tau}')$ system. Indeed, otherwise we would have $(v_i^p)_k \equiv \left(v_j^{q-1}\right)_k$ on some (irreducible) singularity manifold of the $(\Sigma', \mathcal{A}', \vec{\tau}')$ subsystem. Then, by the symmetry with respect to the co-ordinates $k = 1, 2, \dots, \nu$, we would have $v_i^p \equiv v_j^{q-1}$ on a codimension-one singularity of the subsystem $(\Sigma', \mathcal{A}', \vec{\tau}')$, contradicting to the induction hypothesis. Therefore, in the rational expression

$$N(\hat{f}_k) = \frac{Q_k(\vec{x})}{S_k(\vec{x})}$$

of (3.41/b) producing $Q_k(\vec{x})$ in its numerator, no irreducible factor of $\tilde{Q}_k(\vec{x})$ will get canceled when transforming the fraction in (3.41/b) to its shortest form after the substitution $m_N = 0$, see also the important observation on the structure of the denominators $S_k(\vec{x})$ right after (3.41/b). This means that the polynomial $T_k(\vec{x})$ vanishes on the zero set $W_k$ of $\tilde{Q}_k(\vec{x})$, so the non-constant common divisor $\tilde{R}(\vec{x})$

of $\tilde{Q}_k(\vec{x})$ is a common divisor of $T_1(\vec{x}), \dots, T_\nu(\vec{x})$, contradicting to the induction hypothesis. This finishes the proof of Lemma 4.5. $\square$

**Continuing the proof of Proposition 4.2.** Denote by $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$ the "$P_i$ polynomials" of the $N$-ball system $(\Sigma, \mathcal{A}, \vec{\tau})$ with the constraint $m_{k_0} = 0$ constructed the same way as the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ in (3.37) for the general case $m_{k_0} \in \mathbb{C}$, see also (3.41/a) and (3.41/b). It follows from the algebraic construction of these polynomials that $\tilde{Q}(\vec{x})$ is a common divisor of $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$. (See the analogous divisibility argument above, in the closing part of the proof of Lemma 4.5.) Recall that, according to our indirect assumption made right before Lemma 4.3, $Q(\vec{x})$ is a common, irreducible divisor of the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ and, at the same time, $Q(\vec{x})$ is one of the polynomials on the left-hand-side of (1), (2), or (3) in Proposition 4.2. The polynomial $\tilde{Q}(\vec{x})$ was obtained from $Q(\vec{x})$ after the substitution $m_{k_0} = 0$.

Let us focus now on Key Lemma 3.44 of the previous section. The non-sufficiency of the $N$-ball system $(\Sigma, \mathcal{A}, \vec{\tau})$ with the side condition $m_{k_0} = 0$ comes from two sources: Either from the parallelity of the relative velocities in (2) of Key Lemma 3.44, or from the non-sufficiency of the $(N-1)$-ball part of the orbit segment $\omega$ with the combinatorial scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$. The first case takes place on a complex algebraic set of $\nu - 1$ codimensions, and this codimension is at least two, thanks to our original assumption $\nu \geq 3$ and Lemma 4.5 above. Concerning the application of the "non-equality" Lemma 4.5 above, we note here that once the velocities $v^p_{i_p}(\omega)$ and $v^{q-1}_{i_q}(\omega)$ are not equal, the relative velocities $v^p_{k_0}(\omega) - v^p_{i_p}(\omega)$ and $v^{q-1}_{k_0}(\omega) - v^{q-1}_{i_q}(\omega) = v^p_{k_0}(\omega) - v^{q-1}_{i_q}(\omega)$ are not parallel, unless the common velocity $v^p_{k_0}(\omega) = v^{q-1}_{k_0}(\omega)$ belongs to the complex line connecting the different velocities $v^p_{i_p}(\omega)$ and $v^{q-1}_{i_q}(\omega)$, which is a codimension-$(\nu-1)$ condition on the velocity $v^p_{k_0}(\omega)$. Therefore, the non-constant common divisor $\tilde{Q}(\vec{x})$ of the polynomials $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$ can only describe the non-sufficiency of the $(\Sigma', \mathcal{A}', \vec{\tau}')$-part of the system, thus lacking the kinetic and mass variables corresponding to the ball with label $k_0$, as the following sub-lemma states.

**Sub-lemma 4.7.** The non-constant common divisor $\tilde{Q}(\vec{x})$ of the polynomials

$$\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$$

does not contain the variables with label $k_0$.

**Proof.** Let $D = D\left(\Sigma, \mathcal{A}, \vec{\tau} \middle| m_{k_0} = 0\right) \subset \mathbb{C}^{(2\nu+1)N}$ be the open, connected and dense domain in $\mathbb{C}^{(2\nu+1)N}$ defined analogously to Definition 3.18 but incorporating the constraint $m_{k_0} = 0$, see also Lemma 3.19. Let, further, $N \subset D$ be a small, complex analytic submanifold of $D$ with complex dimension $(2\nu+1)N - 1$, holomorphic to the unit open ball of $\mathbb{C}^{(2\nu+1)N-1}$, and such that the polynomial $\tilde{Q}(\vec{x})$ $(\vec{x} \in D)$ vanishes on $N$. We split the vectors $\vec{x} \in D$ as $\vec{x} = (\vec{y}, \vec{z})$, so that the $\vec{z}$-part precisely contains the variables bearing the ball label $k_0$. We may assume that $N \subset B_1 \times B_2$, where $B_1$ and $B_2$ are small, open balls in the spaces of the components $\vec{y}$ and $\vec{z}$, respectively.

Assume, to the contrary of the statement of the sub-lemma, that the polynomial $\tilde{Q}(\vec{x}) \equiv \tilde{Q}(\vec{y}, \vec{z})$ does depend on the component $\vec{z}$. Then, for typical but fixed values $\vec{y}_0$ of $\vec{y}$, the "slice" $\{\vec{y}_0\} \times B_2$ intersects the manifold $N$ in a set of complex codimension one. However, this fact clearly contradicts our earlier observation that the non-sufficiency of the orbit segments $\omega \in D = D\left(\Sigma, \mathcal{A}, \vec{\tau}\middle| m_{k_0} = 0\right)$ imposes a codimension-$(\nu - 1)$ (which is at least two) condition on the coordinates $\vec{z}$ bearing the label $k_0$. This contradiction finishes the proof of the sub-lemma. $\square$

**Finishing the proof of Proposition 4.2.**

If $Q(\vec{x})$ is the left-hand-side of (1) or (2) in 4.2, then we arrive at the conclusion that the irreducible polynomial $\tilde{Q}(\vec{x}) \equiv Q(\vec{x})$ divides $\tilde{P}_1(\vec{x}), \ldots, \tilde{P}_t(\vec{x})$, and $k_0 \neq i_0$, $k_0 \neq j_0$ by Sub-lemma 4.7. This means, however, that the statement of the proposition is false for the $(N-1)$-ball system with the discrete algebraic scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$, contradicting to our induction hypothesis.

If, however, the polynomial $Q(\vec{x})$ is $m_{i_0} + m_{j_0}$, then in the case if $k_0 \notin \{i_0, j_0\}$ we arrive at a contradiction just the same way as above. If $k_0 \in \{i_0, j_0\}$, say $k_0 = j_0$, then $\tilde{Q}(\vec{x}) \equiv m_{i_0}$, and $m_{i_0}$ is a common divisor of all polynomials $\tilde{P}_1(\vec{x}), \ldots, \tilde{P}_t(\vec{x})$ describing the non-sufficiency of the $(\Sigma', \mathcal{A}', \vec{\tau}')$ subsystem with the $N-1$ balls $\{1, 2, \ldots, N\} \setminus \{k_0\}$. This means that the above $(\Sigma', \mathcal{A}', \vec{\tau}')$ subsystem is always non-sufficient, provided that $m_{i_0} = 0$. In the case $N \geq 4$ it follows from Key Lemma 3.33 above (applied to the $(N-2)$-ball system $\{1, 2, \ldots, N\} \setminus \{i_0, k_0\}$) and from the "non-equality" Lemma 4.5 that almost every $(\Sigma', \mathcal{A}', \vec{\tau}')$-orbit with $m_{i_0} = 0$ is in fact sufficient. One easily checks that, in the case $N = 3$, actually every orbit of the 2-ball system $\{1, 2, 3\} \setminus \{k_0\}$ with the side condition $m_{i_0} = 0$ is hyperbolic (sufficient). The obtained contradiction finishes the inductive proof of Proposition 4.2. $\square$

**Corollary 4.8.** Keep all the notations and assumptions of Proposition 4.2, except that we assume now that $\Sigma = (\sigma_1, \ldots, \sigma_n)$ is $(2C(N)+1)$-rich and the singularity manifold $C$ is defined by one of the following equations:

(1)' $\|v_{i_k}^k - v_{j_k}^k\|^2 = 0$,
(2)' $\langle v_{i_k}^k - v_{j_k}^k;\, \tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k \rangle = 0$,
(3)' $m_{i_k} + m_{j_k} = 0$

with some $k$, $1 \leq k \leq n$. Let $Q(\vec{x})$ be a common irreducible divisor of the polynomials $P_1(\vec{x}), \ldots, P_s(\vec{x})$ in (3.37) constructed for the entire discrete structure $(\Sigma, \mathcal{A}, \vec{\tau})$ as above.

We again claim the same thing: The manifold $C$ and the solution set of $Q(\vec{x}) = 0$ can not locally coincide. Consequently, an open, dense, and connected part of the irreducible variety $V = \{Q(\vec{x}) = 0\}$ belongs to the domain $D\left(\Sigma, \mathcal{A}, \vec{\tau}\right)$ of the allowable initial data.

**Proof.** We write $\Sigma = (\sigma_1, \dots, \sigma_n)$ in the form $\Sigma = (\Sigma_1,\, \sigma_k,\, \Sigma_2)$, where $\Sigma_1 = (\sigma_1, \dots, \sigma_{k-1})$, $\Sigma_2 = (\sigma_{k+1}, \dots, \sigma_n)$. Then either $\Sigma_1$ or $\Sigma_2$ is $C(N)$-rich. If $\Sigma_2$ turns out to be $C(N)$-rich, then we can directly apply the proposition after a simple time shift $0 \longmapsto k$. In the other case, when we only know that $\Sigma_1$ is $C(N)$-rich, beside the time shift $0 \longmapsto k$ an additional time-reversal is also necessary to facilitate the applicability of the proposition. $\square$

Another consequence of Proposition 4.2 is

**Corollary 4.9.** Let $\nu \geq 3$, $(\Sigma, \mathcal{A}, \vec{\tau})$ be a discrete orbit structure with Property (A) and a $C(N)$-rich symbolic collision sequence $\Sigma = (\sigma_1,\, \dots,\, \sigma_n)$. Denote by $P_1(\vec{x}),\, \dots,\, P_s(\vec{x})$ the polynomials of (3.37) just as before, and let $Q(\vec{x})$ be a common irreducible divisor of $P_1(\vec{x}),\, \dots,\, P_s(\vec{x})$. (If such a common divisor exists.) Let, finally, $\left(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho}\right)$ be an extended discrete orbit structure with Property (A) and an extended collision sequence $\overline{\Sigma} = (\sigma_0,\, \sigma_1,\, \dots,\, \sigma_n)$. According to Lemma 3.51, we can find a particle, say the one with label $N$, and two indices $1 \leq p < q \leq n$ such that

(i) $N \in \sigma_p \cap \sigma_q$,
(ii) $N \notin \bigcup_{j=p+1}^{q-1} \sigma_j$,
(iii) $\sigma_p = \sigma_q \Longrightarrow (\exists j)\ (p < j < q \,\&\, \sigma_p \cap \sigma_j \neq \emptyset)$, and
(iv) $\Sigma'$ is $(2C(N-1)+1)$-rich on the vertex set $\{1, \dots, N-1\}$.

(Here, just as in the case of derived schemes, we denote by $\Sigma'$ the symbolic sequence that can be obtained from $\Sigma$ by discarding all edges containing $N$.) Denote by $\tilde{Q}(\vec{x})$ the polynomial that we obtain from $Q(\vec{x})$ after the substitution $m_N = 0$. We claim that none of the irreducible polynomials on the left-hand-side of

(1) $\|v_{i_0}^0 - v_{j_0}^0\|^2 = 0$,

(2) $\langle v_{i_0}^0 - v_{j_0}^0;\, \tilde{q}_{i_0}^0 - \tilde{q}_{j_0}^0 - L \cdot a_0 \rangle = 0$,

(3) $m_{i_0} + m_{j_0} = 0$

is a divisor of $\tilde{Q}(\vec{x})$.

**Proof.** Consider and fix an irreducible factor $R(\vec{x})$ of $\tilde{Q}(\vec{x})$. According to Sub-lemma 4.7, the polynomial $R(\vec{x}) \equiv R(\vec{y},\, \vec{z})$ does not contain any variable bearing the label $N$, i. e. $R(\vec{x}) \equiv R(\vec{y})$.

Assume, to the contrary of the statement that we want to prove, that the irreducible polynomial $R(\vec{x}) \equiv R(\vec{y})$ is identical to one of the irreducible polynomials on the left-hand-side of (1), (2), or (3). In particular, we have that $N \neq i_0$, $N \neq j_0$. As we saw in the course of the proof of Sub-lemma 4.7, the algebraic variety $V = \left\{\vec{y} \in \mathbb{C}^{(2\nu+1)(N-1)+1} \,\middle|\, R(\vec{y}) = 0\right\}$, defined by one of the equations (1), (2), or (3), describes the non-sufficiency of the derived system $\Omega\left(\Sigma', \mathcal{A}', \vec{\tau}'\right)$ that one obtains from the original $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ by taking $m_N = 0$, i. e. for any point $\vec{y} \in D\,(\Sigma', \mathcal{A}', \vec{\tau}')$ it is true that $\vec{y} \in V$ if and only if there is some non-sufficient complex orbit segment $\omega \in \Omega\,(\Sigma', \mathcal{A}', \vec{\tau}')$ with $\vec{y}(\omega) = \vec{y}$. However, this statement contradicts to the assertion of Proposition 4.2. $\square$

The main result of this section is

**Key Lemma 4.10.** Keep all the notations and notions of this section. Assume that $\nu \geq 3$ and the symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ of the discrete algebraic frame $(\Sigma, \mathcal{A}, \vec{\tau})$ (with Property (A)) is $C(N)$-rich.

We claim that all orbit segments $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ are sufficient apart from an algebraic variety of codimension-two (at least two, that is), i. e. the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ of (3.37) do not have a non-constant common divisor.

**Proof.** The inductive proof employs many of the ideas of the proof of Proposition 4.2 and it will use the statement of the proposition itself. (More precisely, the statement of Corollary 4.8 is to be used.)

Indeed, the assertion of this key lemma is trivially true in the case $N = 2$, for in that case there are no non-sufficient, complex orbit segments $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ at all.

Assume that $N \geq 3$, and the statement of the key lemma has been successfully proven for all smaller values $(2 \leq)\ N' < N$. Our induction step is going to be a proof by contradiction. Suppose, therefore, that the polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ do have a common irreducible divisor $Q(\vec{x})$. Following the assertion of Lemma 3.51, select a suitable label $k_0 \in \{1, \dots, N\}$ for the substitution $m_{k_0} = 0$ so that a derived scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$ (with Property (A)) exists for the arising $(N-1)$-ball system $\{1, \dots, N\} \setminus \{k_0\}$ with a symbolic sequence $\Sigma'$, possessing the properties (1)–(4) of Lemma 3.51, the same way as we did in the proof of Proposition 4.2.

Consider now the original system $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ with the constraint $m_{k_0} = 0$. After the substitution $m_{k_0} = 0$ the polynomial $Q(\vec{x})$ becomes a new, non-constant polynomial $\tilde{Q}(\vec{x})$, see the proof of Sub-lemma 4.6 above. Let $S(\vec{x})$ be an irreducible divisor of $\tilde{Q}(\vec{x})$. The (indecomposable) algebraic variety $V = \{S(\vec{x}) = 0\}$ has one codimension, for every $\vec{x} \in V$ there exists a non-hyperbolic complex orbit segment $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ with $\vec{x}(\omega) = \vec{x}$ and $m_{k_0}(\omega) = 0$. As far as the non-sufficiency of the orbits $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ with $m_{k_0}(\omega) = 0$ is concerned, we again take a close look at Key Lemma 3.44 above. We saw earlier (in the proof of Proposition 4.2) that the parallelity of the relative velocities $v^p_{k_0}(\omega) - v^p_{i_p}(\omega)$ and $v^{q-1}_{k_0}(\omega) - v^{q-1}_{i_q}(\omega)$ takes place on a manifold with codimension at least two in our case of $\nu \geq 3$. Therefore, according to Key Lemma 3.44, the "codimension-one event" $\vec{x}(\omega) \in V$ ($\Longleftrightarrow S(\vec{x}(\omega)) = 0$) for orbits with $m_{k_0}(\omega) = 0$ can only be equivalent to the non-sufficiency of the $\{1, \dots, N\} \setminus \{k_0\}$-part $\mathrm{trunc}(\omega) \in \Omega\,(\Sigma', \mathcal{A}', \vec{\tau}')$ of the system. In this way it follows from Key Lemma 3.44 that the irreducible polynomial $S(\vec{x})$ lacks all variables bearing the label $k_0$, see also the assertion and the proof of Sub-lemma 4.7. We conclude that for every $\vec{x} \in V$ (i. e. with $S(\vec{x}) = 0$) there exists an orbit segment $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ with $m_{k_0}(\omega) = 0$ and a non-sufficient truncated segment

$\omega' = \text{trunc}(\omega) \in \Omega\,(\Sigma', \mathcal{A}', \vec{\tau}')$. According to Proposition 4.2 above (applied to the $(N-1)$-ball system $\{1, \dots, N\} \setminus \{k_0\}$ with the algebraic scheme $(\Sigma', \mathcal{A}', \vec{\tau}')$), the variety $\{S(\vec{x}) = 0\}$ does not even locally coincide with the singularities of the complex dynamics $\Omega\,(\Sigma', \mathcal{A}', \vec{\tau}')$. This means that a codimension-one family of complex orbit segments $\omega' = \text{trunc}(\omega) \in \Omega\,(\Sigma', \mathcal{A}', \vec{\tau}')$ is not sufficient. This, in turn, contradicts the induction hypothesis of the proof of Key Lemma 4.10 by actually finishing it. $\square$

## 5. Finishing the Proof of Ergodicity

**From $\mathbb{C}$ back to $\mathbb{R}$.**

First of all, we transfer the main result of the previous section (Key Lemma 4.10) from the complex set-up back to the real case. This result will be an almost immediate consequence of Key Lemma 4.10.

Fix a discrete algebraic scheme $(\Sigma, \mathcal{A}, \vec{\tau})$ for $N$ balls with Property (A) (see Definition 3.31) and a $C(N)$-rich symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$. (The definition of the threshold $C(N)$ can be found in Lemma 3.51.) Denote by $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}\,(\Sigma, \mathcal{A}, \vec{\tau})$ the set of all elements $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ for which

(1) all kinetic functions $\left(\tilde{q}_i^k(\omega)\right)_j$, $\left(v_i^k(\omega)\right)_j$, $m_i(\omega)$, and $L(\omega)$ take real values, $i = 1, \dots, N$; $k = 0, 1, \dots, n$; $j = 1, \dots, \nu$;

(2) $\tau_k(\omega) = t_k(\omega) - t_{k-1}(\omega) > 0$ for $k = 1, \dots, n$;

(3) out of the two real roots of (3.8) the root $\tau_k$ is always selected as the smaller one, $k = 1, \dots, n$.

It is clear that either $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}\,(\Sigma, \mathcal{A}, \vec{\tau})$ is a $((2\nu+1)N+1)$-dimensional, real analytic submanifold of $\Omega = \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$, or $\Omega_{\mathbb{R}} = \emptyset$. Of course, we will never investigate the case $\Omega_{\mathbb{R}} = \emptyset$.

Consider the corresponding polynomials $P_1(\vec{x}), \dots, P_s(\vec{x})$ of (3.37) describing the non-sufficiency of the complex orbit segments $\omega \in \Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ along the lines of Lemma 3.36 in terms of the kinetic data $\vec{x} = \vec{x}(\omega)$. According to Proposition 3.41/c, these polynomials $P_i(\vec{x})$ admit real coefficients. By Key Lemma 4.10, the greatest common divisor of $P_1(\vec{x}), \dots, P_s(\vec{x})$ is 1, hence the common zero set

$$\left\{ \vec{x} \in \mathbb{R}^{(2\nu+1)N+1} \,\middle|\, P_1(\vec{x}) = P_2(\vec{x}) = \cdots = P_s(\vec{x}) = 0 \right\}$$

of these polynomials does not contain any smooth real submanifold of (real) dimension $(2\nu+1)N$. In this way we obtained

**Proposition 5.1.** Use all the notions, notations and assumptions from above. There exists no smooth, real submanifold $\mathcal{M}$ of $\Omega_{\mathbb{R}}$ with $\dim_{\mathbb{R}}\mathcal{M} = \dim_{\mathbb{R}}\Omega_{\mathbb{R}} - 1\,(=(2\nu+1)N)$ and with the property that all orbit segments $\omega \in \mathcal{M}$ are non-sufficient. (For the concept of non-sufficiency, please see §2.) $\square$

**The "Fubini-type" Argument.**

Our dynamics $\Omega\,(\Sigma, \mathcal{A}, \vec{\tau})$ has the obvious feature that the variables $m_i = m_i(\omega)$ $(i = 1, \dots, N)$ and $L(\omega)$ (the so called outer geometric parameters) remain unchanged during the time-evolution. Quite naturally, we do not need Proposition 5.1 directly but, rather, we need to use its analog for (almost) every fixed $(N+1)$-tuple $(m_1, \dots, m_N; L) \in \mathbb{R}^{N+1}$. This will be easily achieved by a classical "Fubini-type" product argument. The result is

**Proposition 5.2.** Use all the notions, notations and assumptions from above. Denote by

$$NS = NS\,(\Sigma, \mathcal{A}, \vec{\tau}) = \left\{\omega \in \Omega_{\mathbb{R}}\,(\Sigma, \mathcal{A}, \vec{\tau}) \,\middle|\, \dim_{\mathbb{C}}\mathcal{N}(\omega) > \nu + 1\right\}$$

the set of all non-sufficient, real orbit segments $\omega \in \Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}\,(\Sigma, \mathcal{A}, \vec{\tau})$. (For the definition of the complex neutral space $\mathcal{N}(\omega)$, please see (3.21).) Finally, we use the notation

$$\Omega_{\mathbb{R}}(\vec{m},\, L) = \left\{\omega \in \Omega_{\mathbb{R}} \,\middle|\, \vec{m}(\omega) = \vec{m},\ \text{and } L(\omega) = L\right\}$$

for any given $(N+1)$-tuple $(\vec{m},\, L) = (m_1, \dots, m_N,\, L) \in \mathbb{R}^{N+1}$. We claim that for almost every $(\vec{m},\, L) \in \mathbb{R}^{N+1}$ (according to the Lebesgue measure) the intersection $NS \cap \Omega_{\mathbb{R}}(\vec{m},\, L)$ has at least 2 codimensions in $\Omega_{\mathbb{R}}(\vec{m},\, L)$.

**Remark 5.3.** As it is always the case with such algebraic systems, the exceptional zero-measure set of the parameters $(\vec{m},\, L)$ turns out to be a countable union of smooth, proper submanifolds of $\mathbb{R}^{N+1}$.

**Proof of 5.2.** It is clear that the statement of the proposition is a local one, therefore it is enough to prove that for any small, open subset $U_0 \subset \Omega_{\mathbb{R}}$ of $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}\,(\Sigma, \mathcal{A}, \vec{\tau})$ the set

$$\left\{(\vec{m},\, L) \in \mathbb{R}^{N+1} \,\middle|\, \dim_{\mathbb{R}}\left(NS \cap \Omega_{\mathbb{R}}(\vec{m},\, L) \cap U_0\right) \geq 2\nu N - 1\right\}$$

of the "bad points" $(\vec{m},\, L)$ has zero Lebesgue measure. The points $\omega \in U_0$ can be identified locally (in $U_0$) with the vector $\vec{x} = \vec{x}(\omega) \in D_{\mathbb{R}} = D\,(\Sigma, \mathcal{A}, \vec{\tau}) \cap \Omega_{\mathbb{R}}$ of their coordinates. After this identification the small open set $U_0 \subset \Omega_{\mathbb{R}}$ naturally becomes an open subset $U_0 \subset D_{\mathbb{R}}$. Furthermore, we split the points $\vec{x} \in U_0$ as $\vec{x} = ((\vec{m},\, L),\, \vec{y})$, where $\vec{y}$ contains all variables other than $m_1,\, \dots,\, m_N,\, L$. In this way we may assume that $U_0$ has a product structure $U_0 = B_0 \times B_1$ of two small open balls, so that $B_0 \subset \mathbb{R}^{N+1}$, while the open ball $B_1 \subset \mathbb{R}^{2\nu N}$ contains the $\vec{y}$-parts of the points $\vec{x} = ((\vec{m}, L), \vec{y}) \in U_0$.

Assume that the statement of the proposition is false. Then there exists a small open set $U_0 = B_0 \times B_1 \subset \mathbb{R}^{N+1} \times \mathbb{R}^{2\nu N}$ (with the above splitting) and there is a positive number $\epsilon_0$ such that the set

$$\begin{aligned} A_0 = \Big\{ (\vec{m}, L) \in B_0 \Big| \; & ((\vec{m}, L) \times B_1) \cap NS \text{ contains a} \\ & (2\nu N - 1)\text{-dimensional, smooth, real submanifold with inner radius } > \epsilon_0 \Big\} \end{aligned}$$

has a positive Lebesgue measure in $B_0$. Then one can find an orthogonal projection $P: \mathbb{R}^{2\nu N} \to H$ onto a hyperplane $H$ of $\mathbb{R}^{2\nu N}$ such that, by taking $\Pi(\vec{x}) = \Pi\left((\vec{m}, L), \vec{y}\right) = P(\vec{y})$ ($\Pi: \mathbb{R}^{(2\nu+1)N+1} \to H$), the set

$$\begin{aligned} A_1 = \Big\{ (\vec{m}, L) \in B_0 \Big| \; & \Pi\left[((\vec{m}, L) \times B_1) \cap NS\right] \text{ contains} \\ & \text{an open ball of radius } > \epsilon_0/2 \text{ in } H \Big\} \end{aligned}$$

has positive Lebesgue measure in $B_0$. By the Fubini Theorem the set

$$\tilde{\Pi}\left[NS \cap (B_0 \times B_1)\right]$$

has positive Lebesgue measure in $B_0 \times H$, where

$$\tilde{\Pi}(\vec{x}) = \tilde{\Pi}\left((\vec{m}, L), \vec{y}\right) = \left((\vec{m}, L), P(\vec{y})\right) \in B_0 \times H$$

for $\vec{x} \in B_0 \times B_1$. However, $\dim_{\mathbb{R}}(B_0 \times H) = (2\nu + 1)N$, and $\dim_{\mathbb{R}}\left(NS \cap \Omega_{\mathbb{R}}\right) \leq (2\nu + 1)N - 1$ (according to Proposition 5.1). Thus, we obtained that the real algebraic set

$$\tilde{\Pi}\left(NS \cap (B_0 \times B_1)\right) \subset B_0 \times H$$

has dimension strictly less than $\dim_{\mathbb{R}}(B_0 \times H) = (2\nu + 1)N$, yet it has positive Lebesgue measure in the space $B_0 \times H$. The obtained contradiction finishes the proof of Proposition 5.2. $\square$

## Finishing the Proof of the Theorem

We will carry out an induction with respect to the number of balls $N$ ($\geq 2$). For $N = 2$ the system is well known to be a strictly dispersive billiard flow (after the obvious reductions $m_1 v_1 + m_2 v_2 = 0$, $m_1 ||v_1||^2 + m_2 ||v_2||^2 = 1$ ($m_1, m_2 > 0$), and after the factorization with respect to the uniform spatial translations, as usual) and, as such, it is proved to be ergodic by Sinai in [Sin(1970)], see also the paper [S-W(1989)] about the case of different masses.

Assume now that $N \geq 3$, $\nu \geq 3$, and the theorem has been successfully proven for all smaller numbers of balls $N' < N$. Suppose that a billiard flow

$$\left(\mathbf{M}, \{S^t\}_{t\in\mathbb{R}}, \mu\right) = \left(\mathbf{M}_{\vec{m},L},\, \left\{S^t_{\vec{m},L}\right\},\, \mu_{\vec{m},L}\right)$$

is given for $N$ balls and outer geometric parameters $(\vec{m},\, L) = (m_1,\, \dots,\, m_N,\, L)$ $(m_i > 0,\, L > 0)$ in such a way that, besides the always assumed properties (2.1.1)–(2.1.2),

(*) the vector $(\vec{m},\, L)$ of geometric parameters is such that for any proper subsystem $1 \leq i_1 < i_2 < \cdots < i_{N'} \leq N$ $(2 \leq N' < N)$ and for any $C(N')$-rich discrete algebraic scheme $(\Sigma, \mathcal{A}, \vec{\tau})$ (with Property (A)) for this subsystem $\left(m_{i_1},\, \dots,\, m_{i_{N'}},\, L\right)$ it is true that the parameter vector $\left(m_{i_1},\, \dots,\, m_{i_{N'}},\, L\right)$ does not belong to the zero-measured exceptional set of parameters featuring Proposition 5.2.

According to Lemma 4.1 of [K-S-Sz(1990)], the set $R_2 \subset \mathbf{M}$ of the phase points with at least two singularities on their trajectories is a countable union of smooth submanifolds of $\mathbf{M}$ with codimension two, so this set $R_2$ can be safely discarded in the proof, for it is slim, see also §2 about the slim sets. Secondly, by the induction hypothesis and by Theorem 5.1 of [Sim(1992-I)] (adapted to the case of different masses) there is a slim subset $S_1 \subset \mathbf{M}$ such that for every phase point $x \in \mathbf{M} \setminus S_1$

(i) $S^{(-\infty,\infty)}x$ contains at most one singularity, and

(ii) $S^{(-\infty,\infty)}x$ contains an arbitrarily large number of consecutive, connected collision graphs.

(In the case of a singular trajectory $S^{(-\infty,\infty)}x$ we require that both branches contain an arbitrarily large number of consecutive, connected collision graphs.) Then, by Proposition 5.2 just proved, there is another slim subset $S_2 \supset S_1$ of $\mathbf{M}$ such that

(H) for every $x \in \mathbf{M}\setminus S_2$ the trajectory $S^{(-\infty,\infty)}x$ contains at most one singularity and it is sufficient (or, geometrically hyperbolic).

According to Theorem 6.1 of [Sim(1992-I)] (easily adapted to the case of different masses) and Proposition 5.2, the so called Chernov-Sinai Ansatz (see §2) holds true, i. e. for almost every singular phase point $x \in \mathcal{SR}^+$ the positive semi-trajectory $S^{(0,\infty)}x$ is non-singular and sufficient.

This is the point where the Fundamental Theorem for Algebraic Semi-dispersive Billiards (Theorem 4.4 in [B-Ch-Sz-T(2001)]) comes to play! According to that theorem, by also using the crucial conditions (H) and the Ansatz above, it is true that for every phase point $x \in (\mathrm{int}\mathbf{M}) \setminus S_2$ some open neighborhood $U_x$ of $x$ in $\mathbf{M}$ belongs to a single ergodic component of the considered billiard flow $\left\{S^t_{\vec{m},L}\right\}$. Since the set $(\mathrm{int}\mathbf{M}) \setminus S_2$ contains an arc-wise connected set $C$ with full $\mu$-measure (see Proposition 2.7.9 above), we get that the entire set $C$ belongs to a single ergodic component of the flow $\left\{S^t_{\vec{m},L}\right\}$. This finishes the proof of the ergodicity theorem. $\square$

## References


[B-Ch-Sz-T(2001)] P. Bálint, N. Chernov, D. Szász, I.P. Tóth, *Multi-dimensional Semi-dispersing Billiards: Singularities and the Fundamental Theorem*, To appear in Annales Henri Poincaré.

[B-F-K(1998)] D. Burago, S. Ferleger, A. Kononenko, *Uniform estimates on the number of collisions in semi-dispersing billiards*, Annals of Mathematics **147 (1998)**, 695-708.

[B-L-P-S(1992)] L. Bunimovich, C. Liverani, A. Pellegrinotti, Yu. Sukhov, *Special Systems of Hard Balls that Are Ergodic*, Commun. Math. Phys. **146**, 357-396.

[B-S(1973)] L. A. Bunimovich, Ya. G. Sinai, *The fundamental theorem of the theory of scattering billiards*, Math. USSR-Sb. **19**, 407-423 (1973).

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

[E(1978)] R. Engelking, *Dimension Theory*, North Holland.

[G(1981)] G. Galperin, *On systems of locally interacting and repelling particles moving in space*, Trudy MMO **43**, 142-196 (1981).

[K(1942)] N. S. Krylov, *The Processes of Relaxation of Statistical Systems and the Criterion of Mechanical Instability*, Thesis, Moscow, (1942); *Republished in English by Princeton University Press*, Princeton N. J. (1979).

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

[K-S-Sz(1989)] A. Krámli, N. Simányi, D. Szász, *Ergodic Properties of Semi–Dispersing Billiards I. Two Cylindric Scatterers in the 3–D Torus*, Nonlinearity **2**, 311–326 (1989).

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

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

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

[L-W(1995)] C. Liverani, M. Wojtkowski, *Ergodicity in Hamiltonian systems*, Dynamics Reported **4**, 130-202, arXiv:math.DS/9210229..

[M(1976)] D. Mumford, *Algebraic Geometry I. Complex Projective Varieties*, Springer Verlag, Berlin Heidelberg, 1976.

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

[P(1977)] Ya. Pesin, *Characteristic Exponents and Smooth Ergodic Theory*, Russian Math. surveys **32**, 55-114 (1977).

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

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

[Sim(2001)] N. Simányi, *Proof of the Boltzmann-Sinai Ergodic Hypothesis for Typical Hard Disk Systems*, Submitted for publication, 2001.

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

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

[St(1973)] I. Stewart, *Galois Theory*, Chapman and Hall, London (1973).

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

[S-Sz(1995)] N. Simányi, D. Szász, *The K-property of Hamiltonian systems with restricted hard ball interactions*, Mathematical Research Letters **2, No. 6**, 751-770 (1995).

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

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

[S-W(1989)] N. Simányi, M. Wojtkowski, *Two-particle billiard system with arbitrary mass ratio*, Ergod. Th. & Dynam. Sys. **9**, 165-171 (1989).

[Sz(1994)] D. Szász, *The K-property of 'Orthogonal' Cylindric Billiards*, Commun. Math. Phys. **160**, 581-597 (1994).

[Sz(1996)] D. Szász, *Boltzmann's Ergodic Hypothesis, a Conjecture for Centuries?*, Studia Sci. Math. Hung **31**, 299-322 (1996).

[V(1979)] L. N. Vaserstein, *On Systems of Particles with Finite Range and/or Repulsive Interactions*, Commun. Math. Phys. **69**, 31-56 (1979).

[VDW(1970)] B. L. van der Waerden, *Algebra I*, Frederick Ungar Publ. Co., 1970.

[W(1988)] M. Wojtkowski, *Measure theoretic entropy of the system of hard spheres*, Ergod. Th. & Dynam. Sys. **8**, 133-153.

[W(1990)] M. Wojtkowski, *Linearly stable orbits in 3-dimensional billiards*, Commun. Math. Phys. **129, no. 2**, 319-327.