Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

Website: http://AIMsciences.org
pp. **X–XX**

ALGEBRO-GEOMETRIC METHODS FOR HARD BALL SYSTEMS

Domokos Szász

Budapest University of Technology and Economics
Institute of Mathematics
Budapest, Egry J. u. 1, Hungary H–1111

Dedicated to Yasha Pesin on the occasion of his sixtieth birthday

Abstract. For the study of hard ball systems, the algebro-geometric approach appeared in 1999 — in a sense surprisingly but quite efficiently — for proving the hyperbolicity of typical systems (see [25]). An improvement by Simányi [22] also provided the ergodicity of typical systems, thus an almost complete proof of the Boltzmann–Sinai ergodic hypothesis. More than that, at present, the best form of the local ergodicity theorem for semi-dispersing billiards, [6] also uses algebraic methods (and the algebraicity condition on the scatterers). The goal of the present paper is to discuss the essential steps of the algebro-geometric approach by assuming and using possibly minimum information about hard ball systems. In particular, we also minimize the intersection of the material with the earlier surveys [28] and [20].

1. **Introduction.** In 1994 the scientific world celebrated the 150th anniversary of the birth of Ludwig Boltzmann, whereas in 2006 we commemorated the centenary from his death. My lecture held in 1994 in Vienna (cf. [28] gave me the opportunity to get absorbed by the history of *Boltzmann's ergodic hypothesis*. I followed, in particular, how it led to the birth of ergodic theory, to the first ergodic theorems and to the notion of ergodicity in and around 1930, and then how, in 1963 Sinai's form of Boltzmann's ergodic hypothesis, the so-called *Boltzmann–Sinai ergodic hypothesis*, got formulated for the dynamics of $N \geq 2$ elastic hard balls moving on the ν-torus ($\nu \geq 2$). Finally, I surveyed the major steps reached until 1994 toward establishing the Boltzmann–Sinai ergodic hypothesis. The actuality of that survey was enhanced by the 1987 result of Chernov–Sinai on the ergodicity of the system of 2 hard balls in dimensions higher than 2 (based on their then recent local ergodicity theorem for semi-dispersing billiards) and by the 1991–92 results of the Budapest school on special systems of N balls with $N > 2$. (The ergodicity of planar Sinai-billiards, and consequently that of the system of two discs was proven in Sinai's celebrated 1970 paper [18] and in [7]). These works gave then an impetus to activities in the direction of establishing the Boltzmann–Sinai ergodic hypothesis. Indeed, during

---

2000 *Mathematics Subject Classification.* 37A60, 37D25, 37D50 .

Key words and phrases. Hard ball systems, Boltzmann–Sinai ergodic hypothesis, semi-dispersing billiards, hyperbolicity, ergodicity, algebro-geometric methods.

Research supported by the Hungarian National Foundation for Scientific Research grants No. T046187, NK 63066 and TS 049835, further by Hungarian Science and Technology Foundation grant No. A-9/03.

the more than ten years that have elapsed since 1994, there is an almost complete solution of the problem.

After the study of particular systems reviewed in [28] and [20], our work [25] with Simányi introduced a completely new, algebro-geometric approach. As a result of this work and the forthcoming achievements of Simányi, the Boltzmann–Sinai ergodic hypothesis is, at the time of writing this paper, almost completely solved in a rather general form. Since the algebraic, algebro-geometric methods have become important for the solution of other problems, related to the theory of dispersing or semi-dispersing billiards (cf. [6], [5] and [2]), it has become desirable to explain the method itself, without the technicalities connected to the concrete problem of hyperbolicity/ergodicity of hard ball systems, where the method had appeared. This is actually the basic aim of the present paper. Moreover, we also discuss the main results related to the Boltzmann–Sinai ergodic hypothesis of the period elapsed since 1999.

Before doing that we mention two types of sincere criticisms as to the efforts toward proving the Boltzmann–Sinai ergodic hypothesis. The first one relies on the fact that Boltzmann, when formulating his ergodic hypothesis during laying down the foundations of statistical physics, needed to substitute time averages by ensemble averages not for a wide class of measurements, but for very simple functions, only. This is true as concerns to Boltzmann's usage. However, ergodicity is only a qualitative property, and in most physical applications one needs quantitative properties that can be contrasted to measurements, like correlation decay, transport coefficients, etc. Now only the sophisticated methods developed for settling the Boltzmann–Sinai ergodic hypothesis, can provide the basis for describing the deeper statistical properties of the systems of interest. These methods reveal the mechanism behind the ergodic (or random or chaotic) behavior and without their understanding it is, indeed, hopeless to obtain the delicate properties of the dynamics.

Let me make an additional remark about these methods. It is well-known that hard ball systems are isomorphic to certain semi-dispersing billiards. Consequently, the efforts toward establishing the Boltzmann–Sinai ergodic hypothesis were based on the theory of hyperbolic billiards. This theory having been born from the theory of dynamical systems later developed somewhat independently of it. Different is the case with the stochastic properties. Hyperbolic billiards are hyperbolic dynamical systems with singularities, and they often behave similarly to some one-dimensional maps or to the Hénon-map. The approaches toward statistical properties of hyperbolic billiards borrow several ideas from the theories of these systems. Indeed, since 1994, a new branch in the theory of dynamical systems was born: that of hyperbolic dynamical systems with singularities (cf. [30] or the 1996 workshop Hyperbolic Systems with Singularities held at the Erwin Schrödinger Institute in Vienna). It is, however, beyond the scope of the present work to discuss these methods, though their progress in last decade has been most spectacular. As to ideas born in the first part of the last decade we refer to the excellent expository survey [11], whereas several important new ideas and also problems are explained in [9].

In fact, the foundations of the general theory of (nonuniformly) hyperbolic smooth dynamical systems were worked out in the celebrated paper of Pesin [15]. Later the theory was extended to some hyperbolic systems with singularities, most importantly to billiards by Katok and Strelcyn [12]. Though Pesin's original theory has been about smooth systems, many of his ideas and methods have been indeed fundamental in the theory of hyperbolic systems with singularities. Later Yasha Pesin also got interested and also made important contributions directly to the theory of hyperbolic systems with singularities, see for instance [16].

Let us now turn to the second criticism, which is, contrary to the first one, completely right. Namely, it refers to the original formulation of Boltzmann: time averages can be substituted with ensemble averages for *large* systems of particles in equilibrium. One main advantage of Sinai's version of Boltzmann's hypothesis is that it refers to systems of fixed size, i.e. to systems, which ergodic theory (and ergodicity) is originally about. Consequently, methods of ergodic theory and that of dynamical systems are applicable. On the other hand, typical hamiltonian systems are not expected to be ergodic, and the second main advantage of Sinai's formulation was that it had found the interaction, which, nonetheless, leads to ergodic systems. Once the Boltzmann–Sinai ergodic hypothesis gets established or at least it seems to be within reach, it is time to return to Boltzmann's original conjecture. Unfortunately, it is not formulated in a mathematically rigorous way, but this is not a serious problem. Most experts agree that it is realistic and at the same time physically relevant to expect that given a typical interaction, in the phase space of larger and larger systems there arises a dominant ergodic component, whose measure tends to 1 as the size of the system tends to infinity. This is, however, a much harder question than the Boltzmann–Sinai ergodic hypothesis, which itself has not been that easy either. There exist no rigorous results at all in this generality. In the Boltzmann–Sinai ergodic hypothesis, ergodicity of the systems arises as the result of the — at least — partial hyperbolicity of the hard ball interaction. For large mechanical systems with a typical hamiltonian interaction, the appearance of a large ergodic component should, as Boltzmann did assume, be the result of the large size of the system. At present, there are no tools at hand to grasp this phenomenon. Probably one should copy here the approach of equilibrium statistical mechanics: first one has to understand the behavior of infinite systems.

As aid, the goal of this paper to summarize and explain the developments of the last decade toward proving the Boltzmann–Sinai ergodic hypothesis. We also note that it would be desirable to discuss what is known or done in understanding Boltzmann's ergodic hypothesis for infinite systems, and for large ones as well, but this lies beyond the scope of the present work.

In order to keep the presentation within limits, first of all, I do not repeat details, in particular those about the history of the Boltzmann and the Boltzmann–Sinai Ergodic Hypotheses, and those about that of ergodic theory, which were contained in my earlier paper [28]. To this end, there exists, moreover, the excellent survey by Simányi [20], complementing my earlier work in a successful way and also containing new developments after the appearance of mine. Also, the exposition concentrates on explaining the main flow of ideas and, consequently, avoids the discussion of several technical issues. The importance of algebraic methods is motivated by the fact, that beside proving hyperbolicity or ergodicity, the algebraic approach has recently been used, and is so far, the only way available for the verification of the fundamental "local ergodicity theorem" for higher dimensional semi-dispersing billiards (cf. [6]). Moreover, in the problem of establishing the complexity condition in the form of [30], basic for demonstrating exponential correlation decay for higher dimensional dispersing billiards (cf. [5]) it seems for me that the most promising way is to use the algebraic approach.

2. **Sinai's version of Boltzmann's Ergodic Hypothesis. Billiards.**

The Boltzmann–Sinai Ergodic Hypothesis, [17]. *The system of N hard balls given on $\mathbb{T}^\nu$; $\nu \geq 2$ is ergodic for any $N \geq 2$.*

In fact, Sinai formulated the conjecture for the physical dimensions $\nu = 2, 3$, but the case $\nu = 3$ seems to reveal all the difficulties of the general case.

The reader familiar with the basic definitions about billiards can skip the following brief reminder. We note that the balls have no rotational degrees of freedom!

Billiards. A billiard is a dynamical system describing the motion of a point particle in a connected, compact domain $Q \subset \mathbb{T}^d$ called the *configuration space*. The boundary of the domain is assumed to be a finite union of C^3-smooth pieces. Inside Q the motion is uniform while the reflection at the boundary ∂Q is elastic. As the absolute value of the velocity is a first integral of motion, the phase space of the billiard flow is fixed as $\mathbf{M} = Q \times S^{d-1}$ – in other words, every phase point x is of the form $x = (q, v)$ with $q \in Q$ and $v \in \mathbb{R}^d$, $|v| = 1$. The Liouville probability measure μ on $\mathbf{M}$ is essentially the product of the Lebesgue measures, i.e. $d\mu = \text{const.}\, dqdv$. The resulting dynamical system $(\mathbf{M}, S^{\mathbb{R}}, \mu)$ is the (toric) *billiard flow*.

Let $n(q)$ denote the unit normal vector of a smooth component of the boundary ∂Q at the point q, directed inwards Q. Throughout the paper we restrict our attention to *semi-dispersing billiards* (and sometimes even to *dispersing* ones: we require for every $q \in \partial Q$ the second fundamental form $K(q)$ of the boundary component to be non-negative (or positive resp.).

The boundary ∂Q defines a natural cross-section for the billiard flow. Namely consider

$$\partial \mathbf{M} = \{(q, v) \mid q \in \partial Q, \langle v, n(q) \rangle \geq 0\}.$$

This set actually has a natural bundle structure (cf. [6]). The Poincaré section map T, also called the *billiard map* is defined as the first return map on $\partial \mathbf{M}$. The invariant measure for the map is denoted by μ_1, and we have $d\mu_1 = c_{\mu_1} |\langle v, n(q) \rangle| \, dqdv$, where c_{μ_1} is the normalizing constant $\left(\int_{\partial M} |\langle v, n(q) \rangle| \, dqdv \right)^{-1}$. Throughout the paper we work with this discrete time dynamical system $(\partial \mathbf{M}, T, \mu_1)$. Recall the usual notation: for $(q, v) \in \mathbf{M}$ one denotes by $\pi(q, v) = q$ the natural projection.

The boundary of this phase space consists of singular collisions denoted by $\mathcal{S}_0$ (they are either tangencies, i.e. the orbit is tangent or multiple collisions, i.e. the collision occurs at intersection of smooth boundary components). The dynamics resp. the inverse dynamics is non-continuous on backward resp. forward images of this set. We will denote $\mathcal{S}_i = T^i \mathcal{S}_0$; $i \in \mathbb{Z}$.

Step 1. The first step follows the way of thinking of physics: we replace the N particles with just one high-dimensional one. I.e., we unite the configuration vectors of all centers of balls into a high dimensional configuration vector, and we do the same with the velocity vectors. The interesting, absolutely fundamental and at the same time easy observation is that then the original system of elastic hard balls becomes isomorphic to a high dimensional billiard. To be more definite, the dimension of the configuration vector of the big system is $N\nu$, and — because of energy conservation — that of the velocity vector is $N\nu - 1$. However, momentum is also invariant, and if we assume that it is equal to 0 (if it is not, then to the motion of the system with zero momentum a quasi-periodic motion should be added), then the center of mass is also invariant, so we assume, in addition, that it is also zero. Consequently, the dimension of the configuration vector of the big system is $d = (N-1)\nu$ and that of its velocity is $d-1$.

The scatterers of this high-dimensional billiard are determined by the hard core condition: the centers of any pair of balls can not get closer than $2r$. It is easy to see that these conditions determine spherical scatterers if $N = 2$, and cylindrical ones, if $N \geq 3$.

Consequently, the isomorphic billiard is a dispersing one (i.e. the scatterers are strictly convex) if $N = 2$, and is a semi-dispersing one (i.e. the scatterers are just convex) if $N \geq 3$. (In the forthcoming discussion we will assume that $N \geq 3$, since the case $N = 2$ is an easy consequence of the local ergodicity theorem of [24].)

Fortunately, an analogous reduction also works if the balls have not necessarily identical masses, as is explained in detail in [25], [20], [22].

Step 2. After the reduction of Step 1, our object is the study of certain semi-dispersing billiards, where the scatterers are on the one hand cylinders, and on the other hand their position also reflects the symmetrical role played by the balls. Let us first recall the local ergodicity theorem for semi-dispersing billiards. To this end we have to introduce the fundamental definition of sufficiency of a phase point.

Let us consider a nonsingular finite trajectory segment for the flow: $S^{[a,b]}x$, where $a < 0 < b$ and a, b, 0 are not moments of collision. $\mathcal{N}_0(S^{[a,b]}x)$, the *neutral subspace* at time 0 for the segment $S^{[a,b]}x$ is defined as follows:

$$\begin{aligned}\mathcal{N}_0(S^{[a,b]}x) := \{w \in \mathbb{R}^d : \exists(\delta > 0) \text{ s.t. } \forall\alpha \in (-\delta, \delta)\\ v(S^a(q(x) + \alpha w, v(x))) = v(S^a x)\&\\ v(S^b(q(x) + \alpha w, v(x))) = v(S^b x)\}.\end{aligned}$$

Observe that $v(x) \in \mathcal{N}_0(S^{[a,b]}x)$ is always true, the neutral subspace is at least 1 dimensional. Neutral subspaces at time moments different from 0 are defined by $\mathcal{N}_t(S^{[a,b]}x) := \mathcal{N}_0(S^{[a-t,b-t]}(S^t x))$, thus they are naturally isomorphic to the one at 0.

The non-singular trajectory segment $S^{[a,b]}x$ is *sufficient* if for some (and in that case for any) $t \in [a,b]$: $\dim \mathcal{N}_t(S^{[a,b]}x) = 1$. Denote by M^0 the set of orbits containing no singular collision at all, and by M^1 the set of those containing exactly one singular collision. A point $x \in M^0$ is said to be sufficient if its entire trajectory $S^{(-\infty,\infty)}x$ contains a finite sufficient segment. Singular points are treated by the help of trajectory branches (see [13]): a point $x \in M^1$ (this precisely means that the entire trajectory contains one singular reflection) is sufficient if both of its trajectory branches are sufficient.

All these concepts have their natural counterparts for the billiard map phase space ∂M.

3. **Local ergodicity for semi-dispersing billiards.** Before formulating the local ergodicity theorem let us start with some remarks. Among the assumptions of the theorem there are some about the geometry of the billiard table. For brevity, we do not repeat them here, except for one, namely the algebraicity condition,

and suggest the reader to consult any of the works [24], [13], or [6].

Condition 1 (Algebraicity Condition)**.** *The smooth pieces of the boundary ∂Q of the billiard table are algebraic submanifolds.*

In fact, the authors of [24] and [13] have overlooked a gap in their proofs and, as shown in [6], the algebraicity condition saves the proof for algebraic billiards, at least. Fortunately, so far the interesting multi-dimensional applications of the local ergodicity theorem, like hard ball systems, and the Bunimovich–Rehacek stadia [4], were algebraic, but on the other hand, all experts expect that this requirement is not really needed. Though there exist partial results for resolving this insufficiency of the theory (see [1] with its companion paper [3] and [2]), no satisfactory results have yet been obtained.

We recall, however, an additional substantial condition:

Condition 2 (Chernov–Sinai Ansatz, Condition 3.1 from [13])**.** *For $\nu_{\mathcal{S}_0}$-almost every point $x \in \mathcal{S}_0$ we have $x \in \partial M^* := M^0 \cup M^1$ and, moreover, the positive semi-trajectory of the point x is sufficient. (Here $\nu_{\mathcal{S}_0}$ denotes the Riemannian measure on $\mathcal{S}_0$.)*

Step 3. Local ergodicity theorem. Also we are going to provide a "soft" formulation of the **local ergodicity theorem** which is amply sufficient for our present purpose. A more technical formulation requiring several additional definitions can be found in the aforementioned works.

Theorem 3.1. *Under the Chernov–Sinai Ansatz, Condition 2 plus the Algebraicity Condition 1 and some other geometric conditions (see* [24],[13], [6]*) the following is true: Assume that the orbit of the sufficient phase point $x \in \mathbf{M}$ does not hit the singularity set $\mathcal{S}_0$ more than once. Then x has an open neighborhood which belongs to one ergodic component of the system.*

Assume that the conditions of the theorem hold. Then, if we were so lucky that all points were sufficient, then, of course, the whole phase space would make just one ergodic component. In general, this is not the case and the right way to show ergodicity is to demonstrate that the set of non-sufficient phase points has measure zero and has topological codimension at least two. If we can only show that this codimension is at least one, then we can still claim that for a. e. phase points the relevant Lyapunov exponents are nonzero, and consequently the system is hyperbolic. Then its phase space can be decomposed into a countable number of ergodic components of positive measure.

From now on we will only consider hard ball systems.

Since for semi-dispersing billiards there are *no trajectories at all with a finite accumulation point of collision moments*, for an arbitrary non-singular orbit segment $S^{[a,b]}x$ of the standard billiard ball flow, there is a uniquely defined maximal sequence $a \le t_1 < t_2 < \cdots < t_n \le b : n \ge 0$ of collision times and a uniquely defined sequence $\sigma_1 < \sigma_2 < \cdots < \sigma_n$ of "colliding pairs", i.e. $\sigma_k = \{i_k, j_k\}$ whenever $Q(t_k) = \pi(S^{t_k}x) \in \partial\tilde{C}_{i_k,j_k}$. The sequence $\Sigma := \Sigma(S^{[a,b]}x) := (\sigma_1, \sigma_2, \ldots, \sigma_n)$ is called the *symbolic collision sequence* of the trajectory segment $S^{[a,b]}x$.

Definition 3.2. We say that the symbolic collision sequence $\Sigma = (\sigma_1, \ldots, \sigma_n)$ is *connected* if the collision graph of this sequence:

$$\mathcal{G}_\Sigma := (\mathcal{V} = \{1, 2, \ldots, N\}, \mathcal{E}_\Sigma := \{\{i_k, j_k\} : \text{ where } \sigma_k = \{i_k, j_k\},\ 1 \le k \le n\})$$

is connected. (The vertices of the collision graph are the balls whereas its edges are determined by the colliding pairs of balls during the time interval under investigation.)

Definition 3.3. We say that the symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ is *C-rich*, with C being a natural number, if it can be decomposed into at least C consecutive, disjoint collision sequences in such a way that each of them is connected.

Let us note that for hard ball systems the conditions of the local ergodicity theorem are satisfied almost trivially except for the Chernov–Sinai Ansatz, whose proof is a hard work. Since in this paper our main aim is to explain the algebraic approach to billiards, we will not go into details about the proof of the Chernov–Sinai Ansatz and will rather assume it throughout. Then, as explained in [28] and in [20], the proof of hyperbolicity or ergodicity consists of two completely different parts and is based on a suitable selection of the constants $C = C(N)$ used in the definition of richness (N is the number of balls).

Part I. Show that the subset of non $C(N)$-rich phase points has measure zero and topological codimension at least 2. The proof of this statement was given in the most general case in [19] for $C =?$ implying the statement sufficient for our purposes. We note that these points form a Cantor-type set. On the other hand one can see that, for any C, the subset of the C-rich yet insufficient orbits is a countable union of algebraic submanifolds (the scatterers are quadratically defined!), and so the next task is:

Step 4. Task of the algebro-geometric part.

Part $\mathbf{II_h}$. For demonstrating hyperbolicity: Show that the (algebraic=topological) codimension of $C(N)$-rich, yet insufficient phase points is at least 1.

Part $\mathbf{II_e}$. For demonstrating ergodicity: Show that the (algebraic=topological) codimension of $C(N)$-rich, yet insufficient phase points is at least 2.

The goal of the forthcoming part of the work is, on the one hand, to explain the basic moments of the algebro-geometric approach of [25] for executing the task of Part $\mathbf{II_h}$, and, on the other hand, to hint to the main improvement of [22], for completing Part $\mathbf{II_e}$.

The starting point is the masterly connecting Path Formula (CPF) of Simányi, [19], whose extension to systems of hard balls with different masses is straightforward and was done in [25] (see also [20]).

Step 5. Connecting path formula.

Remark. The essential content of the CPF is a *characterization of the neutral subspace of an orbit segment* through a system of linear equations in terms of the advance functionals of collisions appearing in the symbolic collision sequence of the orbit segment; the coefficients for the advance functional of a given collision are linear functions of the velocity jumps of the particles participating in the collision in question.

4. **Connecting path formula.** The goal of this section is the formulation of Proposition 1 requiring a series of definitions. In a sense, all these definitions are not essential for the application of the CPF. Therefore in the above Remark we have distilled those properties of the CPF which are necessary for understanding the forthcoming arguments. The reader not interested in the CPF itself can jump over the definitions and concentrate on grasping the Proposition and the Remark.

The CPF gives an explicit characterization of the neutral subspace. The reader is reminded that for non-sufficient points the neutral subspace should contain a nontrivial direction apart from the trivial flow direction.

First we introduce the definition of the *advance*. Consider a non-singular orbit segment $S^{[a,b]}x$ with symbolic collision sequence $\Sigma = (\sigma_1, \dots, \sigma_n)$ $(n \ge 1)$. For $x = (Q,V) \in \mathbf{M}$ and $W \in \mathcal{Z}$, $\|W\|$ sufficiently small, denote $T_W(Q,V) := (Q+W,V)$.

Definition 4.1. For any $1 \le k \le n$ and $t \in [a,b]$, the advance

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

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

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

Remark. It is worth observing that exactly for the perturbations in the neutral subspace the orbits remain parallel, and perturbations of their collision points at the consecutive scatterers move along lines. (Roughly speaking it is true that in the neutral subspace perturbations always see a linear structure.) This fact explains why the advance $\alpha(\sigma_k)$ (and similarly why $\Delta q_i(t)$ below) is a linear functional.

Consider a phase point $x_0 \in \mathbf{M}$ whose trajectory segment $S^{[-T,0]}x_0$ is not singular, $T > 0$. In the forthcoming discussion the phase point x_0 and the positive number T will be fixed. All the velocities $v_i(t) \in \mathbb{R}^\nu$ $i \in \{1,2,\dots,N\}$, $-T \le t \le 0$ appearing in the considerations are velocities of certain balls at specified moments t and always with the starting phase point x_0. ($v_i(t)$ is the velocity of the i-th ball at time t.) We suppose that the moments 0 and $-T$ are not moments of collision. We label the balls by the natural numbers $1,2,\dots,N$ (so the set $\{1,2,\dots,N\}$ is always the vertex set of the collision graph) and we denote by $e_1, e_2, \dots, e_n$ the collisions of the trajectory segment $S^{[-T,0]}x_0$ (i.e. the edges of the collision graph) so that the time order of these collisions is just the opposite of the order given by the indices. A few more definitions and notations:

1. $t_i = t(e_i)$ denotes the time of the collision e_i, so $0 > t_1 > t_2 > \cdots > t_n > -T$.
2. If $t \in \mathbb{R}$ is not a moment of collision $(-T \le t \le 0)$, then

   $$\Delta q_i(t):\ \mathcal{N}_0(S^{[-T,0]}x_0) \to \mathbb{R}^\nu$$

   is a linear mapping assigning to every element $W \in \mathcal{N}_0(S^{[-T,0]}x_0)$ the displacement of the i-th ball at time t, provided that the configuration displacement at time zero is given by W. Originally, this linear mapping is only defined for vectors $W \in \mathcal{N}_0(S^{[-T,0]}x_0)$ close enough to the origin, but it can be uniquely extended to the whole space $\mathcal{N}_0(S^{[-T,0]}x_0)$ by preserving linearity.
3. $\alpha(e_i)$ denotes the advance of the collision e_i, thus

   $$\alpha(e_i):\ \mathcal{N}_0(S^{[-T,0]}x_0) \to \mathbb{R}$$

   is a linear mapping $(i = 1,2,\dots,n)$.
4. The integers $1 = k(1) < k(2) < \cdots < k(l_0) \le n$ are defined by the requirement that for every j $(1 \le j \le l_0)$ the graph $\{e_1, e_2, \dots, e_{k(j)}\}$ consists of $N-j$ connected components (on the vertex set $\{1,2,\dots,N\}$, as always!) while the graph $\{e_1, e_2, \dots, e_{k(j)-1}\}$ consists of $N-j+1$ connected components and, moreover, we require that the number of connected components of the whole

graph $\{e_1, e_2, \dots, e_n\}$ be equal to $N - l_0$. It is clear from this definition that the graph

$$\mathcal{T} = \{e_{k(1)}, e_{k(2)}, \dots, e_{k(l_0)}\}$$

does not contain any loop, especially $l_0 \le N - 1$.

Here we make two remarks commenting the above notions.

Remark. We often do not indicate the variable $W \in \mathcal{N}_0(S^{[-T,0]}x_0)$ of the linear mappings $\Delta q_i(t)$ and $\alpha(e_i)$, for we will not be dealing with specific neutral tangent vectors W but, instead, we think of W as a typical (running) element of $\mathcal{N}_0(S^{[-T,0]}x_0)$ and $\Delta q_i(t)$, $\alpha(e_i)$ as linear mappings defined on $\mathcal{N}_0(S^{[-T,0]}x_0)$ in order to obtain an appropriate description of the neutral space $\mathcal{N}_0(S^{[-T,0]}x_0)$.

Remark. If $W \in \mathcal{N}_0(S^{[-T,0]}x_0)$ has the property $\Delta q_i(0)[W] = \lambda v_i(0)$ for some $\lambda \in \mathbb{R}$ and for all $i \in \{1, 2, \dots, N\}$ (here $v_i(0)$ is the velocity of the i-th ball at time zero), then $\alpha(e_k)[W] = \lambda$ for all $k = 1, 2, \dots, n$. This particular W corresponds to the direction of the flow. In the sequel we shall often refer to this remark.

Let us fix two *distinct* balls α, $\omega \in \{1, 2, \dots, N\}$ that *are in the same connected component* of the collision graph $\mathcal{G}_n = \{e_1, e_2, \dots, e_n\}$. The CPF expresses the relative displacement $\Delta q_\alpha(0) - \Delta q_\omega(0)$ in terms of the advances $\alpha(e_i)$ and the relative velocities occuring at these collisions e_i. In order to be able to formulate the CPF we need to define some graph-theoretic notions concerning the pair of vertices (α, ω).

Definition 4.2. Since the graph $\mathcal{T} = \{e_{k(1)}, e_{k(2)}, \dots, e_{k(l_0)}\}$ contains no loop and the vertices α, ω belong to the same connected component of $\mathcal{T}$, there is a unique path $\Pi(\alpha, \omega) = \{f_1, f_2, \dots, f_h\}$ in the graph $\mathcal{T}$ connecting the vertices α and ω. The edges $f_i \in \mathcal{T}$ $(i = 1, 2, \dots, h)$ are listed up successively along this path $\Pi(\alpha, \omega)$ starting from α and ending at ω. The vertices of the path $\Pi(\alpha, \omega)$ are denoted by $\alpha = B_0, B_1, B_2, \dots, B_h = \omega$ indexed along this path going from α to ω, so the edge f_i connects the vertices B_{i-1} and B_i $(i = 1, 2, \dots, h)$.

When trying to compute $\Delta q_\alpha(0) - \Delta q_\omega(0)$ by using the advances $\alpha(e_i)$ and the relative velocities at these collisions, it turns out that not only the collisions f_i $(i = 1, 2, \dots, h)$ make an impact on $\Delta q_\alpha(0) - \Delta q_\omega(0)$, but some other adjacent edges too. This motivates the following definition:

Definition 4.3. Let $i \in \{1, 2, \dots, h-1\}$ be an integer. We define the set $\mathcal{A}_i$ of *adjacent edges at the vertex* B_i as follows:

$$\begin{aligned}\mathcal{A}_i = \{e_j :\ & j \in \{1, 2, \dots, n\} \,\&\, (t(e_j) - t(f_i)) \cdot (t(e_j) - t(f_{i+1})) < 0 \,\&\\ & B_i \text{ is a vertex of } e_j\}.\end{aligned}$$

We adopt a similar definition to the sets $\mathcal{A}_0$, $\mathcal{A}_h$ of adjacent edges at the vertices B_0 and B_h, respectively:

Definition 4.4.

$$\mathcal{A}_0 = \{e_j :\ 1 \le j \le n \,\&\, t(e_j) > t(f_1) \,\&\, B_0 \text{ is a vertex of } e_j\}\,;$$
$$\mathcal{A}_h = \{e_j :\ 1 \le j \le n \,\&\, t(e_j) > t(f_h) \,\&\, B_h \text{ is a vertex of } e_j\}\,.$$

We note that the sets $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_h$ are not necessarily mutually disjoint.

Finally, we need to define the "contribution" of the collision e_j to $\Delta q_\alpha(0) - \Delta q_\omega(0)$ which is composed from the relative velocities just before and after the moment $t(e_j)$ of the collision e_j.

Definition 4.5. For $i \in \{1, 2, \ldots, h\}$ the "contribution" $\Gamma(f_i)$ of the edge $f_i \in \Pi(\alpha, \omega)$ is given by the formula

$$\Gamma(f_i) = \begin{cases} v^-_{B_{i-1}}(t(f_i)) - v^-_{B_i}(t(f_i)), \\ \text{if } t(f_{i-1}) < t(f_i) \ \&\ t(f_{i+1}) < t(f_i); \\ v^+_{B_{i-1}}(t(f_i)) - v^+_{B_i}(t(f_i)), \\ \text{if } t(f_{i-1}) > t(f_i) \ \&\ t(f_{i+1}) > t(f_i); \\ \dfrac{1}{m_{B_{i-1}} + m_{B_i}} \Big[m_{B_{i-1}} \left(v^-_{B_{i-1}}(t(f_i)) - v^-_{B_i}(t(f_i)) \right) \\ +m_{B_i} \left(v^+_{B_{i-1}}(t(f_i)) - v^+_{B_i}(t(f_i)) \right) \Big], \\ \text{if } t(f_{i+1}) < t(f_i) < t(f_{i-1}) \\ \dfrac{1}{m_{B_{i-1}} + m_{B_i}} \Big[m_{B_{i-1}} \left(v^+_{B_{i-1}}(t(f_i)) - v^+_{B_i}(t(f_i)) \right) \\ +m_{B_i} \left(v^-_{B_{i-1}}(t(f_i)) - v^-_{B_i}(t(f_i)) \right) \Big], \\ \text{if } t(f_{i-1}) < t(f_i) < t(f_{i+1}). \end{cases}$$

Here $v^-_{B_i}(t(f_i))$ denotes the velocity of the B_i-th particle just before the collision f_i (occuring at time $t(f_i)$) and, similarly, $v^+_{B_i}(t(f_i))$ is the velocity of the same particle just after the mentioned collision. We also note that, by convention, $t(f_0) = 0 > t(f_1)$ and $t(f_{h+1}) = 0 > t(f_h)$. Apparently, the time order plays an important role in this definition.

Definition 4.6. For $i \in \{0, 1, 2, \ldots, h\}$ the "contribution" $\Gamma_i(e_j)$ of an edge $e_j \in \mathcal{A}_i$ is defined as follows:

$$\Gamma_i(e_j) = \operatorname{sign}\left(t(f_i) - t(f_{i+1})\right) \frac{m_C}{m_{B_i} + m_C} \cdot \left[\left(v^+_{B_i}(t(e_j)) - v^+_C(t(e_j))\right) - \left(v^-_{B_i}(t(e_j)) - v^-_C(t(e_j))\right) \right],$$

where C is the vertex of e_j different from B_i.

Here again we adopt the convention of $t(f_0) = 0 > t(e_j)$ ($e_j \in \mathcal{A}_0$) and $t(f_{h+1}) = 0 > t(e_j)$ ($e_j \in \mathcal{A}_h$). We note that, by the definition of the set $\mathcal{A}_i$, exactly one of the two possibilities $t(f_{i+1}) < t(e_j) < t(f_i)$ and $t(f_i) < t(e_j) < t(f_{i+1})$ occurs. The subscript i of Γ is only needed because an edge $e_j \in \mathcal{A}_{i_1} \cap \mathcal{A}_{i_2}$ ($i_1 < i_2$) has two contributions at the vertices B_{i_1} and B_{i_2} which are just the endpoints of e_j.

We are now in the position of formulating the Connecting Path Formula:

Proposition 1. *Using all definitions and notations above, the following sum is an expression for* $\Delta q_\alpha(0) - \Delta q_\omega(0)$ *in terms of the advances and relative velocities of collisions:*

$$\Delta q_\alpha(0) - \Delta q_\omega(0) = \sum_{i=1}^{h} \alpha(f_i)\Gamma(f_i) + \sum_{i=0}^{h} \sum_{e_j \in \mathcal{A}_i} \alpha(e_j)\Gamma_i(e_j).$$

5. **Algebro-geometric methods.** Now we arrived at the heart of the algebro-geometric methods.

Step 6. Form of the induction. All experts had agreed that the Boltzmann–Sinai ergodic hypothesis should be proved by **induction on the number of balls** and we also follow this approach. The novelty of the approach is, however,

(6/A) to consider the masses $m_1, \ldots, m_N$ and the (for simplicity, common) radius r as *variables*, (6/B) and to execute the inductive argument by letting the mass of one, suitably selected particle converge to 0.

Step 7. Translation to algebraic language: Lifting orbits to euclidean space. The algebraic structure of the torus is not suitable for our purposes. For to apply appropriate algebraic tools, we have to lift the trajectories given on the torus to those in euclidean space.

This can be done without any difficulty. However, there is one important detail to be taken into account and to keep record of: when an orbit segment gets lifted to euclidean space, then the lift of a scatterer given on the torus is not unique, in fact, it can be indexed by vectors of $\mathbb{Z}^\nu$. For avoiding too lengthy technicalities, we are satisfied by having mentioned this circumstance, but will not introduce further notations, and from now on we will rather concentrate on the logical structure of the proof.

To this end, nevertheless, we introduce the following notations: having fixed the symbolic collision sequence $\Sigma = (\sigma_1, \ldots, \sigma_n)$ $(n \geq 1)$, consider an orbit with this specific symbolic collision sequence:

$$\left\{q_i^{k-1},\, v_i^{k-1} \mid i = 1, \ldots, N;\; k = 1, \ldots, n\right\}$$

(here, of course, $q_i^{k-1} \in \mathbb{T}^\nu$). The lifted orbit segment will be denoted then by

$$\left\{\tilde{q}_i^{k-1},\, v_i^{k-1} \mid i = 1, \ldots, N;\; k = 1, \ldots, n\right\},$$

where $\tilde{q}_i^{k-1} \in \mathbb{R}^\nu$.

Step 8. Calculation of orbit from the initial coordinates. Complexification and field extensions. Now the coordinates of the lifted orbit segment can be calculated step by step from the coordinates $\tilde{x} = \left\{\tilde{q}_i^0,\, v_i^0\right\} \in \mathbb{R}^{2\nu N}$ of the initial point. Two important remarks: (i) in the forthcoming algebraic arguments one can forget about the reductions prescribed by the trivial invariants of the motion, and this is why we so far have been considering the initial phase point in $2\nu N$-dimensional space; (ii) Since we were taking the mass vector $\overrightarrow{m} = (m_1, \ldots, m_N)$ and the radius r as – also complex – variables the basic variable of our initial field will be $x = (\tilde{x}, \overrightarrow{m}, r) \in \mathbb{R}^{2\nu N+N+1}$. The stepwise solution is nothing else than solving quadratic equations: the initial coordinates determine a line and we have to determine its intersection with the (lifted) scatterer and afterwards determine the outgoing velocity vector. Here again there occurs an important difficulty and two technical ones, too.

The important one: For having an algebraically closed field we should work with $\mathbb{C}^{2\nu N+N+1}$ rather than with $\mathbb{R}^{2\nu N+N+1}$. Therefore **we complexify the dynamics**. The advantage is that we can calculate the orbits step by step, though the price to pay is that we loose geometric intuition.

The technical difficulties will only be mentioned and even not formulated precisely: (i) to deal with the quadratic equations and extensions as needed, the method requires some assumptions on the coordinates at each step. These require the exclusion of certain algebraic submanifolds and for executing the procedure we are describing we should ensure at each step that these assumptions hold; (ii) Moreover, at each step, when the arising quadratic equation is reducible, then we can select any of the two, algebraically non-conjugate roots, therefore we have to make a selection and we should also keep record of the selection. Again, by keeping in mind this circumstance, we will not introduce further notations for this ambiguity.

In fact, we do not really solve the equations, but rather introduce step by step quadratic field extensions determined by the subsequent quadratic equations. Anyway, at the end of the procedure, from the CPFs characterizing the set of non-hyperbolicity conditions for the given trajectory segment, one obtains a finite set $\{P_j(x) \mid j \in J\}$ of polynomials over $\mathbb{C}^{2\nu N+N+1}$. These polynomials only depend on the fixed symbolic collision sequence (and as mentioned above: on the selection of the set of $\mathbb{Z}^\nu$ lattice vectors when lifting and on the sequence of decisions about which root to select).

Remark. The procedure of solving these quadratic equations or equivalently of the subsequent field extensions is completely analogous to the field extensions arising in the classical theory of geometric constructions by a compass and a ruler.

Step 9. Polynomial equations for non-hyperbolicity. As the result of the previous procedure we have arrived at the so-called *hyperbolicity polynomials* of the orbit segment.

Pseudo-Definition 5.1. The set of polynomials $\{P_j(x) \mid j \in J\}$ is called the hyperbolicity polynomials of the orbit segment.

The reader can already guess that the following statement is true. Since we do not formulate it in a technical way, we will call it (and other similar statements to come) a *pseudo-Theorem*.

Pseudo-Theorem 5.2 (dichotomy corollary)**.** *Given a symbolic collision sequence (and the other parameters, decisions described above), either at least one hyperbolicity polynomial does not vanish identically, and then the phase points with the given collision structure are sufficient (except perhaps a proper algebraic subvariety), or every phase point with the given collision structure is non-sufficient.*

(for the precise statement cf. Corollary 4.7 of [25]).

6. **Hyperbolicity of typical hard ball systems.** Denote by $R_0 := R_0(N, \nu, L)$ the interval of those values of $r > 0$, for which the interior $\mathrm{Int}\mathbf{M} = \mathrm{Int}\mathbf{Q} \times \mathcal{E}$ of the phase space of the standard billiard ball flow is connected.

Our aim now is to sketch the proof of

Theorem 6.1 ([25])**.** *For $N \geq 2$, $\nu \geq 2$ and $r \in R_0$, none of the relevant Lyapunov exponents of the standard billiard ball system $(\mathbf{M}, \{S^{\mathbb{R}}\}, \mu)_{\vec{m},r}$ vanishes — apart from a countable union of proper analytic submanifolds of the outer geometric parameters $(\vec{m}; r) \in \mathbb{R}_+^{N+1}$.*

We note that this proof does not use the local ergodicity theorem so we do not have to worry about the Chernov–Sinai Ansatz either.

Step 10. Inductive lemma about sufficiency. For preparing the induction we formulate a simple statement which is, heuristically at least, not much surprising and is utmost important.

Consider a symbolic collision sequence: $\Sigma = (\sigma_1, \dots, \sigma_n)$ $(n \geq 1)$ and assume that the N-th particle is "infinitely light" compared to the others, i.e. suppose that $m_N = 0$ but $m_1 \cdot \dots \cdot m_{N-1} \neq 0$.

Pseudo-Lemma 6.2. *Suppose that $m_N = 0$ and $m_1 \cdot \dots \cdot m_{N-1} \neq 0$. 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 $j = p+1, \dots, q-1$. Then an orbit segment with the fixed collision sequence is sufficient (i.e. $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$) if*

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

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

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 $i_p(i_q)$ 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 $N \notin \sigma_i$.)

(for the precise statement cf. Lemma 4.9 of [25]).

Step 11. Combinatorial lemma for richness. The next, purely combinatorial lemma is the last ingredient of the inductive proof of the key lemma.

Lemma 6.3. *Define the sequence of positive numbers* $C(N)$ *recursively by taking* $C(2) = 1$ *and* $C(N) = \frac{N}{2} \cdot \max\{C(N-1); 3\}$ *for* $N \geq 3$*. Let* $N \geq 3$*, and suppose that the symbolic collision sequence* $\Sigma = (\sigma_1, \ldots, \sigma_n)$ *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) $\sigma_p = \sigma_q \Longrightarrow \exists j\ p < j < q\ \&\ \sigma_p \cap \sigma_j \neq \emptyset$,

and

(iv) Σ' *is* $C(N-1)$*-rich on the vertex set* $\{1, \ldots, N-1\}$*.*

Here we denote by Σ' *the symbolic sequence that can be obtained from* Σ *by discarding all edges containing* N*.*

Step 12. Pseudo-induction. Now we can give our pseudo-arguments for proving the following

Pseudo-Theorem 6.4. *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* $\Sigma = (\sigma_1, \ldots, \sigma_n)$ *is* $C(N)$*-rich, then* $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$ *(i.e. the trajectory is sufficient) for almost every phase point* $x \in \mathbb{C}^{2\nu N + N + 1}$*. The real version of this result is also valid.*

(cf. Key Lemma 4.1 in [25]).

Pseudo-Proof of Pseudo-Theorem 6.4. First we treat the complex version. The key ingredients of the proof are now collected, and we can sketch the proof. Assume that the statement is true for all $C(N-1)$-rich symbolic collision sequences of $(N-1)$ balls. Select $C(N)$ as indicated by Lemma 6.3. Assume on the contrary to the statement that there exists a $C(N)$-rich symbolic collision sequence $\Sigma = (\sigma_1, \ldots, \sigma_n)$, such that all its hyperbolicity polynomials (cf. Pseudo-Definition 5.1) identically vanish. Then select the ball N according to Lemma 6.3. By substituting $m_N = 0$ we obtain the hyperbolicity polynomials $\{P_j(x)_{m_N=0} \mid j \in J\}$ of the system of balls $1, 2, \ldots, (N-1)$ that also vanish. But, because of Lemma 6.3, the reduced symbolic collision sequence of the system of balls $1, 2, \ldots, (N-1)$ is $C(N-1)$-rich and we arrive at a contradiction.

Finally, since a nonzero polynomial $P_i(\vec{x})$ $(1 \leq i \leq s)$ takes nonzero values almost everywhere on the real space $\mathbb{R}^{2\nu N + N + 1}$, we immediately obtain the validity of the real version of the statement. $\square\ \square$

7. **Ergodicity of typical hard ball systems.** The statement of Theorem 6.1 got improved in 2004 by N. Simányi, who could establish the Boltzmann–Sinai ergodic hypothesis, in general, for typical parameters.[1]

Theorem 7.1 ([22])**.** *For $N \geq 2$, $\nu \geq 2$ and $r \in R_0$, the standard billiard ball system $(\mathbf{M}, \{S^{\mathbb{R}}\}, \mu)_{\vec{m},r}$ is ergodic and completely hyperbolic — apart from a countable union of proper analytic submanifolds of the outer geometric parameters $(\vec{m}; r) \in \mathbb{R}_+^{N+1}$.*

The idea of the proof uses the fact that for a family of polynomials not to have a a non-trivial common divisor is equivalent to the property that the common solution set of the polynomials has topological codimension at least two. Thus our goal is to show that the hyperbolicity polynomials do not have any non-constant common divisor. The basic observation the proof of the theorem is based upon is elementary.

Lemma 7.2 ([22])**.** *The hyperbolicity polynomials $\{P_j(x) \mid j \in J\}$ are homogeneous in the masses $m_1, \dots, m_N$, and consequently, any common divisor of these polynomials is also homogeneous in the masses.*

The combinatorial Lemma 6.3 now gets replaced by a bit stronger lemma.

Lemma 7.3. *Define the sequence of positive numbers $C(N)$ recursively by taking $C(2) = 1$ and $C(N) = \frac{N}{2} \cdot (2C(N-1)+1)$ 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) $\sigma_p = \sigma_q \Longrightarrow \exists j \; p < j < q \;\&\; \sigma_p \cap \sigma_j \neq \emptyset$,

and

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

Here we denote by Σ' the symbolic sequence that can be obtained from Σ by discarding all edges containing N.

Again, we dismiss talking about — not easy — technical details, nontheless, we recall Simányi's key lemma revealing the main new element of his approach.

Pseudo-Theorem 7.4 ([22])**.** *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, then $\dim_{\mathbb{C}} \mathcal{N}(\omega) = \nu + 1$ (i.e. the trajectory is sufficient) apart from an algebraic variety of codimension at least two, in other words the hyperbolicity polynomials $\{P_j(x) \mid j \in J\}$ do not have a non-constant common divisor.*

Pseudo-proof. If, on the contrary, for some $C(N)$-rich collision sequence the claim does not hold, then there exists a non-constant polynomial $Q(x)$ such that $\forall j Q(x) \mid P_j(x)$. Then choose again the ball N by the combinatorial Lemma 7.3 and put $m_N = 0$. For a polynomial $R(x)$ denote $\tilde{R}(x_{(N)}) = R(\tilde{q}_i, v_i;\ i = 1, \dots, N;\ m_1, \dots, m_{N-1}, m_N = 0)$. Of course, $\forall j \tilde{Q} \mid \tilde{P}_j$. We claim that $\tilde{Q} \neq$ const. If it were, then we would have $Q = c + m_N Q^*$. $c \neq 0$ is impossible since then Q were not a homogeneous polynomial contradicting to Lemma 7.2. However, if $c = 0$, then the polynomials

[1] For his work [22] Nándor Simányi was awarded the 2004 Prize of Annales Henri Poincaré.

P_j have a common divisor m_N, i. e. $m_N = 0$ alone implies non-sufficiency, which statement can be easily falsified by a constructive dynamical argument. $\square$

8. **Cylindrical billiards.** *Cylindric billiards*, a much interesting subfamily of semi-dispersing billiards were introduced in 1992 in my paper [27]. Cylindric billiards are interesting for,

on one hand, they contain *hard ball systems*, a fundamental model from the aspects of statistical physics and a much beautiful one, we believe, from the point of view of mathematics; and

on the other hand, this is apparently the widest subclass of semi-dispersive billiards where the search for *transparent necessary and sufficient conditions of ergodicity* is promising.

Indeed, in [27], it was conjectured that such a condition only depends on the generator subspaces of the scatterers-cylinders. Also, such a condition: *there exists at least one sufficient trajectory* got formulated in [27]. A constructive form of such a condition was conjectured in [26]: *This condition requires that the action of a Lie-subgroup $\mathcal{G}$ of the orthogonal group $SO(d)$ (d being the dimension of the billiard in question) be transitive on the unit sphere S^{d-1}. If $C_1, \dots, C_k$ are the cylindric scatterers of the billiard, then $\mathcal{G}$ is generated by the embedded Lie-subgroups $\mathcal{G}_i$ of $SO(d)$, where $\mathcal{G}_i$ consists of all transformations $g \in SO(d)$ of $\mathbb{R}^d$ that leave the points of the generator subspace of C_i fixed ($1 \le i \le k$).* It is worth mentioning that $\mathcal{G}_i$ is nothing else than the full rotation group around the axis subspace of the cylinder C_i. [2]

In their paper, the authors showed that

(1) this condition is necessary for ergodicity,
(2) and cylindric billiards isomorphic to hard ball systems satisfy this condition; consequently, the conjecture is stronger than the Boltzmann–Sinai ergodic hypothesis.

9. **Results for every hard ball system.** In [21], Simányi proved hyperbolicity for cylindric billiards satisfying some additional conditions beyond the transitivity one. In particular, his theorem implies the following improvement of Theorem 6.1:

Theorem 9.1. *Every hard ball system is completely hyperbolic.*

Recall that the proof of hyperbolicity does not use the local ergodicity theorem, and is consequently independent of establishing the Chernov–Sinai Ansatz. Now, — for me unexpectedly — at present the final verification of the Boltzmann–Sinai ergodic hypothesis hangs on the Chernov–Sinai Ansatz. Concretely, Simányi, [23] has recently shown

Theorem 9.2 ([23])**.** *For $N \ge 2$, $\nu \ge 2$, the standard billiard ball system $(\mathbf{M}, \{S^{\mathbb{R}}\}, \mu)_{\vec{m},r}$ is hyperbolic and ergodic for every value of the outer geometric parameters $(\vec{m}; r) \in \mathbb{R}_+^N \times (0, R_0)$ if we assume that the Chernov-Sinai Ansatz (see Condition 2 above) holds true for the standard billiard ball system and all its subsystems.*

One can only hope that soon there will be a non-hypothetical, final proof of the Boltzmann–Sinai ergodic hypothesis.

[2] The authors called the conjecture based on this condition the Erdőtarcsa Conjecture, in honor of the village, where a mansion belonging to the Hungarian Academy of Sciences is situated; the authors found this condition and the results of [26] during their work there.

Acknowledgement. The author expresses his sincere thanks to the referees for their valuable remarks.

REFERENCES

[1] P. Batchourine, *On the structure of singularity submanifolds of dispersing billiards*, preprint, http://www.arxiv.org/ps/math.DS/0505620.

[2] P. Bachurin, P. Bálint and I.P. Tóth, *Local ergodicity for systems with growth properties including multi-dimensional dispersing billiards*, preprint, to appear in Israel Journal of Mathematics.

[3] P. Batchourine and Ch. Fefferman, *The volume near the zeros of a smooth function*, Revista Mat. Iberoamericana, to appear, 2005.

[4] L.A. Bunimovich and J. Rehacek, *Nowhere Dispersing 3D Billiards with Non-vanishing Lyapunov Exponents*, Commun. Math. Phys., **189** (1997), 729–758.

[5] P. Bálint and I.P. Tóth, *Exponential decay of correlations in multi-dimensional dispersing billiards*, manuscipt, 2007.

[6] P. Bálint, N. Chernov, D. Szász and I.P. Tóth, *Multi-dimensional Semi-Dispersing Billiards: Singularities and the Fundamental Theorem*, Annales Henri Poincaré, **3** (2002), 451–482.

[7] A. Bunimovich and Ya.G. Sinai, *Fundamental Theorem in the Theory of Dispersing Billiards*, Mat. Sbornik, **90** (1973), 415–431.

[8] B.V. Chirikov, *A universal instability of many-dimensional oscillator systems*, Phys. Rep., **52** (1979), 263–379.

[9] N. Chernov and D. Dolgopyat, *Hyperbolic billiards and statistical physics*, Proceedings ICM-2006 **II.** (2006), 1679–1704.

[10] N. Chernov and N. Simányi, *Flow-invariant hypersurfaces in semi-dispersing billiards*, Annales Henri Poincaré, to appear, 2007.

[11] N. Chernov and L.-Y. Young, *Decay of Correlations for Lorentz Gases and Hard Balls*, Encyclopaedia of Mathematical Sciences, 101:51-8889-120, 2000.

[12] A. Katok and J.-M. Strelcyn, *Invariant manifolds, Entropy and Billiards; Smooth maps with singularities*, Lecture Notes in Mathematics, **1222**, 1986.

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

[14] C. Liverani, *Interacting particles*, Encyclopaedia of Mathematical Sciences, **101** (2000), 179–216.

[15] Ya.G. Pesin, *Characteristic exponents and smooth ergodic theory*, Russian Math. Surveys, **32** (1977), 55–114.

[16] Ya.G. Pesin, *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, Ergodic Th. & Dynam. Sys., **12** (1992), 123–151.

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

[18] Ya.G. Sinai, *Dynamical Systems with Elastic Reflections*, Usp. Mat. Nauk, **25** (1970), 141–192.

[19] N. Simányi, *The K-property of N billiard balls*, Invent. Math., I. **108** 521–548, II. **110** (1992), 151–172.

[20] N. Simányi, *Hard Ball Systems and Semi-Dispersive Billiards*, Encyclopaedia of Mathematical Sciences, **101** 51–88.

[21] N. Simányi, *The complete hyperbolicity of cylindric billiards*, Ergodic Th. & Dynam. Sys., **22** (2002), 281–302.

[22] N. Simányi, *Proof of the ergodic hypothesis for typical hard ball systems*, Annales H. Poincaré, **5** (2004), 203–233.

[23] N. Simányi, *Conditional proof of the Boltzmann–Sinai ergodic hypothesis*, (Assuming the hyperbolicity of typical singular orbits.), arxiv:math.DS0605358, 2007.

[24] Ya.G. Sinai and N.I. Chernov, *Ergodic Properties of Some Systems of 2-D Discs and 3-D Spheres*, Usp. Mat. Nauk, **42** (1987), 15.

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

[26] N. Simányi and D. Szász, *Non-integrability of cylindric billiards and transitive Lie-group actions*, Ergodic Th. & Dynam. Sys., **20** (2000), 593–610.

[27] D. Szász, *Ergodicity of classical billiard balls*, Physica A., **194** (1993), 86–92.

[28] D. Szász, *Boltzmann's Ergodic Hypothesis, a Conjecture for Centuries*, Studia Sci. Math. Hung., **31** (1996), 266–322. (Reprinted in [29], 423–446).

[29] D. Szász, Editor, *Hard Ball Systems and the Lorentz Gas*, Encyclopaedia of Mathematical Sciences, **101** 2000.

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

Received XX; revised XX.

E-mail address: szasz@math.bme.hu