November 21, 1999

Non-integrability of Cylindric Billiards and Transitive Lie Group Actions

Nándor Simányi[1]
Bolyai Institute, University of Szeged,
6720 Szeged, Aradi Vértanuk tere 1, Hungary.
E-mail: simanyi@math.u-szeged.hu

Domokos Szász[1]
Mathematical Institute of the Hungarian Academy of Sciences
H-1364, Budapest, P. O. B. 127, Hungary
E-mail: szasz@math-inst.hu

Abstract. A conjecture is formulated and discussed which provides necessary and sufficient condition for the ergodicity of cylindric billiards (this conjecture improves a previous one of the second author). 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 ($1 \le i \le k$). In this paper we can prove the necessity of our conjecture and we also formulate some notions related to transitivity. For hard ball systems, we can also show that the transitivity holds in general: for arbitrary number $N \ge 2$ of balls, arbitrary masses $m_1, \dots, m_N$ and in arbitrary dimension $\nu \ge 2$. This result implies that our conjecture is stronger than the Boltzmann-Sinai ergodic hypothesis for hard ball systems. As a by-remark, we can give a somewhat surprising characterization of the positive subspace of the second fundamental form for the evolution of a special orthogonal manifold (wavefront), namely for the parallel beam of light. Thus we obtain a new characterization of sufficiency of an orbit segment.

1. Introduction.

Semi-dispersive billiards is a class of billiards evidently without any ellipticity, actually with more or less manifest hyperbolicity; their study was initiated by

[1]Research supported by the Hungarian National Foundation for Scientific Research, grants OTKA-7275 and OTKA-16425.

Chernov and Sinai in 1987 [S-Ch(1987)]. *Cylindric billiards*, a much interesting subfamily of semi-dispersive billiards were introduced in 1992 by the second author, [Sz(1993)]. 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.

In words, cylindric billiards are toric billiards where the scatterers are cylinders. In our discussion, the bases of the cylinders will be assumed to be stricly convex, a property ensuring that the scatterers be convex, and thus the arising billiard be semi-dispersive. Because of the simplicity of our model, let us immediately start with a formal definition.

The configuration space of a cylindric billiard is $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \cdots \cup C_k)$, where $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ $(d \geq 2)$ is the unit torus. We note here that all notions and proofs work well without any significant change if we allow general flat tori $\mathbb{T}^d = \mathbb{R}^d / \mathcal{L}$, where $\mathcal{L}$ is a full lattice in $\mathbb{R}^d$, i. e. a discrete subgroup of the additive group $\mathbb{R}^d$ with maximum rank d. The cylindric scatterer C_i $(i = 1, \dots, k)$ is defined as follows:

Let $A_i \subset \mathbb{R}^d$ be a so called *lattice subspace* of the Euclidean space $\mathbb{R}^d$, which means that the discrete intersection $A_i \cap \mathbb{Z}^d$ has rank $\dim A_i$. In this case the factor $A_i/(A_i \cap \mathbb{Z}^d)$ naturally defines a subtorus of $\mathbb{T}^d$, which will be taken as the generator of the cylinder $C_i \subset \mathbb{T}^d$. Denote by $L_i = A_i^\perp$ the orthocomplement of A_i. Throughout this article we will always assume that $\dim L_i \geq 2$. Let, moreover, $D_i \subset L_i$ be a convex, compact domain with a C^2-smooth boundary ∂D_i containing the origin as an interior point. We will assume that D_i is strictly convex in the sense that the second fundamental form of its boundary ∂D_i is everywhere positive definite. Furthermore, in order to avoid unnecessary complications, we assume that the convex domain D_i does not contain any pair of points congruent modulo $\mathbb{Z}^d$. The domain D_i will be taken as the *base* of the cylinder C_i. Finally, suppose that a translation vector $t_i \in \mathbb{R}^d$ is given, playing an essential role in positioning the cylinder C_i in the ambient torus $\mathbb{T}^d$. Set

$$C_i = \{a + l + t_i \mid a \in A_i,\ l \in B_i,\ \} / \mathbb{Z}^d.$$

In order to avoid further unnecessary complications, we also assume that the interior of the configuration space $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \cdots \cup C_k)$ is connected.

The phase space $\mathbf{M}$ of our billiard will be the unit tangent bundle of $\mathbf{Q}$, i. e. $\mathbf{M} = \mathbf{Q} \times \mathbb{S}^{d-1}$. (Here, as usual, $\mathbb{S}^{d-1}$ is the $d-1$-dimensional unit sphere.)

The dynamical system $(\mathbf{M}, S^{\mathbb{R}}, \mu)$, where $S^{\mathbb{R}}$ is the dynamics defined by uniform motion inside the domain and specular reflections at its boundary (the scatterers!) and μ is the Liouville measure, is a *cylindric billiard* we want to study. (As to notions and notations in connection with semi-dispersive billiards the reader is recommended to consult the work [K-S-Sz(1990)].)

In 1994, the first general result: necessary and sufficient conditions of ergodicity, was obtained in [Sz(1994)] for the class of the so-called *orthogonal* cylindric billiards, where the generator of each cylinder is parallel to some coordinate subspace of the orthogonal system of coordinates in which the underlying torus is given. Then, in [S-Sz(1994)], sufficient conditions were found for the ergodicity, in fact, for a hyperbolic one, of a billiard with two non-orthogonal cylindric scatterers. The method of the proof already revealed the intimate role which the *transitivity on* $\mathbb{S}^{d-1}$ *of the action of a Lie-subgroup* $\mathcal{G}$ *of the special orthogonal group* $SO(d)$ would play in ensuring the hyperbolicity, and consequently the ergodicity, too, of the cylindric billiard in question. To be more precise, for any cylinder C_i $(i = 1, \dots, k)$ consider the group $\mathcal{G}_i$ consisting of all orthogonal transformations $U \in SO(d)$ that leave the points of the subspace A_i fixed. Then we consider the subgroup $\mathcal{G}$ of $SO(d)$ algebraically generated by all such groups $\mathcal{G}_i : 1 \le i \le k$, which is certainly an analytic (i. e. embedded, connected Lie-) subgroup of $SO(d)$. The Lie algebra of $\mathcal{G}$ is obviously the Lie subalgebra of $\mathfrak{so}(d)$ generated by the Lie algebras of the groups $\mathcal{G}_i$, $i = 1, \dots, k$.

Our aim in this work is to formulate precise statements about the aforementioned transitivity of the action of $\mathcal{G}$. Having collected some necessary notions in section 2, the general results are formulated in section 3. According to our Conjecture 1, transitivity of $\mathcal{G}$ is equivalent to the hyperbolic ergodicity of the flow.

Conjecture 1. *For every cylindric billiard flow* $(\mathbf{M}, \{S^t\}, \mu)$ *the transitivity of the action of* $\mathcal{G}$ *on the velocity sphere* $\mathbb{S}^{d-1}$ *is equivalent to the hyperbolic ergodicity (or, equivalently, to the hyperbolicity and Bernoulli property) of the billiard map.*

The main result of section 3 is

Theorem 3.6. *The irreducibility of the* $\mathcal{G}$*-action on* $\mathbb{R}^d$ *is equivalent to the transitivity on* $\mathbb{S}^{d-1}$.

This theorem has the following important corollary:

Corollary A. *The transitivity of the* $\mathcal{G}$*-action on* $\mathbb{S}^{d-1}$ *is a necessary condition both for the ergodicity and the complete hyperbolicity of the cylindric billiard.*

Though this corollary provides the necessity of the transitivity property, it seems to be extremely hard to establish the sufficiency of this condition. Indeed, for the "simple" subclass of cylindric billiards: hard ball systems, this is not known, in general. In section 4 of the present work, we will be able to show that for hard ball systems, transitivity of $\mathcal{G}$ holds for any number N of particles, in any dimension ν, and for arbitrary masses $m_1, \dots, m_N$ and radii r of the particles. As a contrast, we note that in our recent, quite involved paper [S-Sz(1999)], we were only able to prove hyperbolicity — and still not ergodicity — for masses $m_1, \dots, m_N$ and radii r of the particles not belonging to a countable union of exceptional analytic submanifolds of parameters. In the light of this last result, the merit of the transitivity statement of section 4 mentioned above is that it holds for arbitrary geometric parameters of the system — without any exceptional set. Motivated by the importance of the concept of transitivity, in section 3, we also introduce related notions: the tightness and the orthogonal non-splitting property of $\mathcal{G}$, the first of them being stronger than, while the second one equivalent to transitivity. Furthermore, Proposition 3.18 of Section 3 gives a new, we think beatiful and quite surprising characterization of sufficiency.

2. Prerequisites

As to the basic notions about semi-dispersive billiards we refer to the paper [K-S-Sz (1990)]. For convenience and brevity, we will throughout use the concepts and notations of Sections 2 and 3 of that paper. Here we only summarize some further notions from [K-S-Sz (1991)] , [K-S-Sz (1992)] , and [Sim(1992)]. These are either new or their exposition is simpler than that given in the original work.

An often used abbreviation is the shorthand $S^{[a,b]}x$ for the trajectory segment $\{S^t x : a \le t \le b\}$. The natural projections from $\mathbf{M}$ onto its factor spaces are denoted, as usual, by $\pi : \mathbf{M} \to \mathbf{Q}$ and $p : \mathbf{M} \to \mathcal{S}^{d-1}$ or, sometimes, we simply write $\pi(x) = Q(x) = Q$ and $p(x) = V(x) = V$ for $x = (Q, V) \in \mathbf{M}$. Any $t \in [a, b]$ with $S^t x \in \partial \mathbf{M}$ is called *a collision moment or collision time.* Denote $\partial \mathbf{Q} = \cup_{i=1}^{k} \partial \mathbf{C}_i$, where $\partial \mathbf{C}_i$ are the smooth components of the boundary. Since we want to consider typical situations, we will always assume that *at every point $q \in \partial \mathbf{Q}$ of the boundary of the configuration space $\mathbf{Q}$ the spherical angle subtended by the compact domain $\mathbf{Q}$ is strictly positive.*

As pointed out in previous works on billiards, the dynamics can only be defined for trajectories where the moments of collisions do not accumulate in any finite time interval (cf. Condition 2.1 of [K-S-Sz(1990)]). An important consequence of Theorem 5.3 of [V(1979)] is that — for semi-dispersive billiards we are considering — there are *no trajectories at all with a finite accumulation point of collision moments.*

As a result, for an arbitrary non-singular orbit segment $S^{[a,b]}x$ of the cylindric billiard flow, there is a uniquely defined sequence $a \le \tau_1 < \tau_2 < \cdots < \tau_m \le b$ ($m \ge 0$) of collision times and a uniquely defined sequence $\sigma_1, \sigma_2, \dots, \sigma_m$ ($1 \le \sigma_i \le k$) of labels of cylinders with the properties that

(i) for every $t \in [a, b], S^t x \in \partial \mathbf{M}$ if and only if $t = \tau_i$ with some $i = 1, \dots, m$, and

(ii) $\pi\left(S^{\tau_i} x\right) \in \partial \mathbf{C}_{\sigma_i}, i = 1, \dots, m$.

The sequence $\Sigma := \Sigma(S^{[a,b]}x) := (\sigma_1, \sigma_2, \dots, \sigma_m)$ is called the *symbolic collision sequence* of the trajectory segment $S^{[a,b]}x$.

As well known, billiards are dynamical systems with *singularities.* A collision at a point $x \in \partial \mathbf{M}$ such that in $\pi(x)$, at least two smooth pieces of $\partial \mathbf{Q}$ meet, is called a *multiple* collision. A collision is called *tangential* if $x \in \partial \mathbf{M}$ and $p(x) \in \mathcal{T}_{\pi(x)} \partial \mathbf{Q}$, i. e. $p(x)$ is tangential to $\partial \mathbf{Q}$ at the point of reflection. We shall use the collection $\mathcal{SR}^+$ of all singular reflections:

Definition 2.1. *The set $\mathcal{SR}^+$ is the collection of all phase points $x \in \partial \mathbf{M}$ for which the reflection, occurring at x, is singular (tangential or multiple) and, in the case of a multiple collision, x is supplied with the outgoing velocity V^+.*

It is not hard to see that $\mathcal{SR}^+$ is a compact cell-complex in $\mathbf{M}$ and $\dim(\mathcal{SR}^+) = \dim \mathbf{M} - 2 = 2d - 3$.

Neutral Subspaces and Sufficiency

Consider a *non–singular* trajectory segment $S^{[a,b]}x$. Suppose that a and b are *not moments of collision.* Before defining the neutral linear space of this trajectory segment, we note that the tangent space of the configuration space $\mathbf{Q}$ at interior points can be identified with the common linear space $\mathbb{R}^d$.

Definition 2.1. *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:*

$$\begin{aligned}\mathcal{N}_0(S^{[a,b]}x)=\big\{W\in\mathbb{R}^d:\ \exists(\delta>0)\ s.\ t.\ \forall\alpha\in(-\delta,\delta)\\ p\left(S^a\left(Q(x)+\alpha W,V(x)\right)\right)=p(S^a x)\ \&\\ p\left(S^b\left(Q(x)+\alpha W,V(x)\right)\right)=p(S^b x)\big\}.\end{aligned}$$

It is known (see (3) in Section 3 of [S-Ch (1987)]) that $\mathcal{N}_0(S^{[a,b]}x)$ is a linear subspace of $\mathbb{R}^d$ 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:

$$\mathcal{N}_t(S^{[a,b]}x)=\mathcal{N}_0\left(S^{[a-t,b-t]}(S^t x)\right). \tag{2.2}$$

It is clear that the neutral space $\mathcal{N}_t(S^{[a,b]}x)$ can be identified canonically 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, Section 2 of [K-S-Sz (1990)]). As we will see in Section 3, the neutral subspace is the orthocomplement of the positive subspace of the second fundamental form of the image of a parallel beam of light, see the proof of Proposition 3.18 later.

It is now time to bring up the basic notion of *sufficiency* of a trajectory (segment). This is the utmost important necessary condition for the proof of the Theorem on Local Ergodicity for Semi–Dispersing billiards, see Condition (ii) of Theorem 3.6 and Definition 2.12 in [K-S-Sz (1990)] .

Definition 2.3.

(1) *The non–singular 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 is said to be sufficient if and only if both branches of this trajectory segment are sufficient.*

For the notion of trajectory branches see, for example, the end of Section 2 in [Sim(1992)-I].

Definition 2.4. *The phase point $x\in\mathbf{M}$ with at most one 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 $S^{[a,b]}x$ is sufficient.*

In the case of an orbit $S^{(-\infty,\infty)}x$ with exactly one singularity, sufficiency requires that both branches of $S^{(-\infty,\infty)}x$ be sufficient.

3. Characterization of the Action of $\mathcal{G}$

As has been said in the introduction, we consider cylindric billiards with the configuration space $\mathbf{Q}=\mathbb{T}^d\setminus(C_1\cup\cdots\cup C_k)$, where $\mathbb{T}^d=\mathbb{R}^d/\mathbb{Z}^d$ ($d\geq 2$) is the unit torus, the generator space of the cylinder C_i is $A_i\subset\mathbb{R}^d$, $L_i=A_i^{\perp}$, and $\dim L_i\geq 2$.

Recall that the linear subspace $A_i \subset \mathbb{R}^d$ is a lattice subspace of $\mathbb{R}^d$, so that it defines a subtorus $T_i = A_i/\mathbb{Z}^d$ of $\mathbb{T}^d$.

For a cylinder C_i $(i = 1, \dots, k)$ consider the group $\mathcal{G}_i$ consisting of all special orthogonal transformations $U \in SO(d)$ that leave the points of the subspace A_i fixed. Since the initial studies of the close relationship between the ergodicity of cylindric billiards and the rotation groups determined by the generator spaces (cf. Lemma 4.4 and Sublemma 4.5 in [S-Sz (1994)] and the role of these lemmas in the proof of Main Lemma 4.1 there) it has become apparent that the transitivity of the action of the algebraic generate of these $\mathcal{G}_i$'s on the sphere $\mathbb{S}^{d-1}$ is vital for proving ergodicity. We investigate the subgroup $\mathcal{G}$ of $SO(d)$ algebraically generated by all such groups $\mathcal{G}_i$, which is an analytic (embedded, connected Lie-) subgroup of $SO(d)$. Observe that every orbit $\mathcal{G}V$ $(V \in \mathbb{S}^{d-1})$ of the action of $\mathcal{G}$ on the unit sphere $\mathbb{S}^{d-1}$ is certainly an embedded, smooth submanifold of $\mathbb{S}^{d-1}$, for it is naturally diffeomorphic to the homogeneous space that can be obtained by taking $\mathcal{G}$ modulo the isotropy subgroup of V.

Being aware of our general, three-step strategy of proving ergodicity for semi-dispersive billiards, it is clear that without the mentioned transitivity there would not be a great chance to prove the hoped ergodicity for such billiards. Later on, in Theorem 3.6, we will see that without the transitivity of the $\mathcal{G}$-action actually there must exist some very simple first integrals of the flow, thus hindering the ergodicity. At this point, however, it is already easy to see that without that transitivity the flow $(\mathbf{M}, \{S^t\}, \mu)$ simply can not be hyperbolic. Indeed, hyperbolicity just means that for $\mu-$almost every point $x \in M$ the relevant Lyapunov exponents of the system are nonzero. On the other hand, if for an $x \in M$, the relevant exponents do not vanish, then the point is necessarily sufficient. By the following lemma, however, the existence of just one sufficient point already implies the transitivity of the action of $\mathcal{G}$.

Denote by δ the maximum dimension of the orbits of the action of $\mathcal{G}$ on the unit sphere $\mathbb{S}^{d-1}$.

Lemma 3.1. *The action of $\mathcal{G}$ on $\mathbb{S}^{d-1}$ is transitive if and only if $\delta = d-1$, i. e. there is an open orbit.*

Proof. Transitivity obviously implies $\delta = d-1$.

Suppose that the orbit $\mathcal{G}V$ of some $V \in \mathbb{S}^{d-1}$ is an open subset of the unit sphere $\mathbb{S}^{d-1}$. Then there exists an $\epsilon_0 > 0$ such that $B(V; \epsilon_0) \subset \mathcal{G}V$. Since $\mathcal{G} \subset SO(d)$, we have that $B(gV; \epsilon_0) \subset \mathcal{G}V$ for all $g \in \mathcal{G}$. Using the connectedness of $\mathbb{S}^{d-1}$, this is only possible if $\mathcal{G}V = \mathbb{S}^{d-1}$. $\square$

First we put forward a simple observation regarding the (possibly existing) dense orbits of the $\mathcal{G}$-action.

Lemma 3.2. *Suppose that the orbit $\mathcal{G}V \subset \mathbb{S}^{d-1}$ is dense for some $V \in \mathbb{S}^{d-1}$. Then every orbit of the $\mathcal{G}$-action is also dense.*

Proof. Denote by $\overline{\mathcal{G}} \subset SO(d)$ the closure of the group $\mathcal{G}$ in the orthogonal group $SO(d)$. It is obvious that the closure $\overline{\mathcal{G}V}$ of the orbit $\mathcal{G}V$ is precisely the orbit $\overline{\mathcal{G}}V$ of the $\overline{\mathcal{G}}$-action on $\mathbb{S}^{d-1}$. Thus, $\overline{\mathcal{G}V} = \mathbb{S}^{d-1}$ means that the $\overline{\mathcal{G}}$-action is transitive, hence $\overline{\mathcal{G}}V' = \overline{\mathcal{G}V'} = \mathbb{S}^{d-1}$ for every $V' \in \mathbb{S}^{d-1}$. $\square$

Let us observe that the existence of a single dense orbit implies the irreducibility of the $\mathcal{G}$-action on the space $\mathbb{R}^d$. Thus, irreducibility of the $\mathcal{G}$-action is a necessary condition for transitivity.

Now we present another necessary condition for transitivity:

Definition of Orthogonal Non-Splitting Property, ONSP. *We say that the system of subspaces* $L_1, \dots, L_k$ *has the Orthogonal Non-Splitting Property (ONSP) iff there is no orthogonal splitting* $\mathbb{R}^d = B_1 \oplus B_2$ *with* $\dim B_i > 0$ *and with the property that for every* $i = 1, \dots, k$ *either* $L_i \subset B_1$ *or* $L_i \subset B_2$.

Lemma 3.3. *The irreducibility of the* $\mathcal{G}$*-action on* $\mathbb{R}^d$ *implies the ONSP.*

Proof. If ONSP fails to hold, then the subspaces B_i in the splitting described above (with the property that for every $i = 1, \dots, k$ either $L_i \subset B_1$ or $L_i \subset B_2$) are $\mathcal{G}$-invariant.

We note that in this case, in addition to the total kinetic energy $\frac{1}{2}\|V\|^2$, the cylindric billiard clearly has further first integrals (invariant quantities), namely, the kinetic energies corresponding to the invariant norm square $\|P_{B_i}(V)\|^2$ $(i = 1,\, 2)$ of the orthogonal projection of the velocity V on the subspace B_i. $\square$

Lemma 3.4. *The irreducibility of the action of* $\mathcal{G}$ *on* $\mathbb{R}^d$ *is equivalent to the ONSP.*

Proof. We have seen that the irreducibility implies the ONSP. Suppose now the reducibility, i. e. the existence of an orthogonal decomposition $\mathbb{R}^d = B_1 \oplus B_2$, $\dim B_i > 0$, and B_i is $\mathcal{G}$-invariant. (Recall that the orthocomplement of any invariant subspace is also invariant, hence we necessarily have such a decomposition.) We show that ONSP is violated by the same splitting $\mathbb{R}^d = B_1 \oplus B_2$.

Consider and fix an arbitrary index i, $1 \le i \le k$.

Sublemma. *If the linear subspace* $B \subset \mathbb{R}^d$ *is* $\mathcal{G}_i$*-invariant, then* B *has the orthogonal direct sum splitting* $B = B^* \oplus B^{**}$*, where* $B^{**} \subset A_i$ *and* $B^* = 0$ *or* $B^* = L_i$*. (Recall that* $\mathcal{G}_i \subset \mathcal{G}$ *consists of all orthogonal transformations* $U \in SO(d)$ *that leave all vectors of* A_i *fixed.)*

Proof. If B is a subspace of $A_i = L_i^{\perp}$, then we are done: $B^{**} = B$, $B^* = 0$. Assume now that the orthogonal projection $P_{L_i}(B)$ is nonzero. Select a vector $b \in B$ with $P_{L_i}(b) \ne 0$. Since the tangent space $\mathcal{T}_b\,(\mathcal{G}_i b)$ is equal to the orthocomplement $L_i \ominus P_{L_i}(b) = H_i$, by the invariance of B we have that $H_i \subset B$. Since $\mathcal{G}_i$ acts transitively on the unit sphere of L_i, it follows that $L_i \subset B$. Take $B^* = L_i$, $B^{**} = B \ominus L_i$. $\square$

Consider now the splittings $B_1 = B_1^* \oplus B_1^{**}$, $B_2 = B_2^* \oplus B_2^{**}$. Then we have that either $B_1^* = L_i$ or $B_2^* = L_i$, for $B_1 \oplus B_2 \not\subset A_i$. This shows that either $L_i \subset B_1$ or $L_i \subset B_2$, $i = 1, \dots, k$. $\square$

Remark 3.5. Each of the above properties obviously implies that $\bigcap_{i=1}^k A_i = 0$.

It is now a bit surprising fact that the irreducibility of the $\mathcal{G}$-action (or, equivalently, the ONSP) in turn implies the transitivity of the action on $\mathbb{S}^{d-1}$, thus giving us an easily checkable necessary and sufficient condition for transitivity.

Theorem 3.6. *The irreducibility of the* $\mathcal{G}$*-action on* $\mathbb{R}^d$ *implies the transitivity on* $\mathbb{S}^{d-1}$*, i. e. these properties are actually equivalent.*

Corollary A. *The transitivity of the* $\mathcal{G}$*-action on* $\mathbb{S}^{d-1}$ *is a necessary condition both for the ergodicity and the complete hyperbolicity of the cylindric billiard.*

Proof. Without transitivity there must exist a non-trivial, orthogonal $\mathcal{G}$-nvariant splitting $\mathbb{R}^d = B_1 \oplus B_2$, and the partial kinetic energies $\frac{1}{2}\|P_{B_i}(V)\|^2$ are first integrals of the motion. Similarly, the translation invariance of the entire system in the direction of (say) B_1 shows that the flow is not completely hyperbolic. $\square$

Corollary B. The Orbit Structure. *The natural representation of the group $\mathcal{G}$ in $\mathbb{R}^d$ is always completely reducible, i. e. the space $\mathbb{R}^d$ splits into the orthogonal direct sum $\mathbb{R}^d = \bigoplus_{j=1}^r B_j$, where each subspace B_j is $\mathcal{G}$-invariant, and the restriction of the representation to B_j is irreducible. Note that the above splitting is also unique, as long as the subspaces L_i span the whole space $\mathbb{R}^d$. Two subspaces L_i and L_j belong to the same component B_s $(s = 1, \dots, r)$ if and only if these subspaces can be connected by a finite chain of spaces $L_{i(1)}, L_{i(2)}, \dots, L_{i(p)}$ so that the consecutive elements in this chain are not orthogonal to each other. It follows from Theorem 3.6 that every orbit of the $\mathcal{G}$-action on $\mathbb{S}^{d-1}$ has the following form:*

$$\left\{V \in \mathbb{S}^{d-1} \middle| \ \|P_{B_j}(V)\|^2 = 2E_j, \quad j = 1, \dots, r\right\}.$$

Here the quantity E_j is the flow invariant partial kinetic energy in the direction of the subspace B_j, $j = 1, \dots, r$. The orbit is uniquely determined by the vector $(E_1, E_2, \dots, E_r)$ fulfilling $E_j \geq 0$ and $\sum_{j=1}^r E_j = 1/2$. Thus every orbit is compact, for it is the product of spheres. The typical orbit is the topological product of $(\nu_j - 1)$-dimensional spheres, $j = 1, \dots, r$, where $\nu_j = \dim B_j$.

Proof of Theorem 3.6. The proof will be split into several lemmas, some of them being just very simple observations.

Lemma 3.7. *Let us introduce the following equivalence relation $\sim_i$ $(i = 1, \dots, k)$ between vectors of $\mathbb{R}^d$:*

$$\begin{aligned} &V_1 \sim_i V_2 \ \textit{if and only if} \\ &V_1 - P_i V_1 = V_2 - P_i V_2 \ \textit{and} \ \|P_i V_1\| = \|P_i V_2\|. \end{aligned} \tag{3.8}$$

(Recall that P_i denotes the orthogonal projection onto L_i.) Suppose that some subgroups (not necessarily Lie subgroups) $\mathcal{H}_i \subset \mathcal{G}_i = SO(L_i)$ are given with the property that $\mathcal{H}_i$ acts transitively on the unit sphere of L_i, $i = 1, \dots, k$.

Then the algebraic generate $\mathcal{H} = \langle \mathcal{H}_1, \dots, \mathcal{H}_k \rangle$ of the groups $\mathcal{H}_i$ acts transitively on $\mathbb{S}^{d-1}$ if and only if the transitive hull of the equivalence relations $\sim_i$ on $\mathbb{S}^{d-1}$ $(i = 1, \dots, k)$ is the trivial equivalence relation on $\mathbb{S}^{d-1}$ making every pair of velocities equivalent.

Proof. The assertion of the lemma immediately follows from the definitions. $\square$

Based on the previous lemma, it will be convenient to say that the system of subspaces $\{L_1, \dots, L_k\}$ of $\mathbb{R}^d$ $(\dim L_i \geq 2,\ i = 1, \dots, k)$ is *transitive* if and only if the transitive hull of the relations $\sim_i$ $(i = 1, \dots, k)$ associated with the subspaces $L_1, \dots, L_k$ is the trivial relation $\mathbb{S} \times \mathbb{S}$ on the unit sphere $\mathbb{S}$ of the linear span $L_1 + L_2 + \cdots + L_k$ of these subspaces.

Lemma 3.9. *If the systems of linear subspaces $\{L_1, L_2, \dots, L_p\}$ and*

$$\{L_1 + L_2 + \cdots + L_p, L_{p+1}, \dots, L_k\}$$

are both transitive $(1 \leq p \leq k)$, then the system $\{L_1, \dots, L_k\}$ is transitive, as well.

Proof. The lemma is an immediate consequence of Lemma 3.7. □

The next proposition deals with the transitivity of two subspaces $\{L_1, L_2\}$ of $\mathbb{R}^d$.

Proposition 3.10. *The system of two subspaces $\{L_1, L_2\}$ of $\mathbb{R}^d$ ($\dim L_i \geq 2$, $i = 1, 2$) is transitive if and only if L_1 and L_2 are not orthogonal to each other.*

Proof. Non-orthogonality is, of course, necessary for transitivity. Assume now that $L_1 \not\perp L_2$, and show that the system $\{L_1, L_2\}$ is transitive. In order to simplify the notations, we assume that $L_1 + L_2 = \mathbb{R}^d$.

Case I. $L_1 \cap L_2 \neq 0$. Select a vector $V \in \mathbb{S}^{d-1}$ such that $V \notin (L_1 \cap L_2)^\perp$. It is easy to see that the tangent space $\mathcal{T}_V \mathcal{G}_i V$ of the orbit $\mathcal{G}_i V$ at V ($i = 1, 2$) is the linear space $V^\perp \cap L_i$ where, as usual, $\mathcal{G}_i$ denotes the group of all special orthogonal transformations of $\mathbb{R}^d$ that keep the vectors of $L_i^\perp = A_i$ fixed. Elementary computation with the dimensions shows that

$$\begin{aligned}&\dim\left[(V^\perp \cap L_1) + (V^\perp \cap L_2)\right] = \dim L_1 - 1 + \dim L_2 - 1 - (\dim L_1 \cap L_2 - 1)\\&= \dim(L_1 + L_2) - 1 = d - 1.\end{aligned}$$

Due to Lemma 3.1, the system $\{L_1, L_2\}$ is indeed transitive in Case I.

Case II. $L_1 \cap L_2 = 0$. In this case the above calculus of dimensions provides

$$\dim\left[(V^\perp \cap L_1) + (V^\perp \cap L_2)\right] = d - 2,$$

so we have by one less dimension than the needed $d - 1$. Interestingly enough, the missing one dimension of Case II and the trick to obtain transitivity of the action by using Lie brackets already appeared in [K-S-Sz(1991)]. Actually, we apply now the commutator method used in the proof of Sublemma 4.5 of [S-Sz(1994)]. Indeed, we select two-dimensional subspaces $\tilde{L}_i \subset L_i$ ($i = 1, 2$) so that $\tilde{L}_1$ and $\tilde{L}_2$ are still not orthogonal to each other, then we consider the four-dimensional linear span $H = \tilde{L}_1 + \tilde{L}_2$. In the spirit of Lemma 3.9, in order to prove the transitivity of the system $\{L_1, L_2\}$ it is enough to show that

(A) the system $\{\tilde{L}_1, \tilde{L}_2\}$ is transitive,

and

(B) the system $\{L_1, L_2, H\}$ is transitive.

By the result of Case I, both systems $\{L_1 + H, L_2\}$ and $\{L_1, H\}$ are transitive and, therefore, according to Lemma 3.9, the statement of (B) is indeed true.

The only outstanding task in the proof of Proposition 3.10 is to show (A).

In order to simplify the notations, we assume that $\tilde{L}_1 + \tilde{L}_2 = \mathbb{R}^4$. We have arrived at the set-up of Sublemma 4.5 of [S-Sz(1994)], except that now we do not have the bit stronger assumption $\tilde{L}_1 \cap (\tilde{L}_2)^\perp = 0$ but, rather, we only have that $\tilde{L}_1 \not\perp \tilde{L}_2$. This means that the cosine

$$b = \min\left\{\|P_2(x)\| \,\Big|\, x \in \tilde{L}_1, \quad ||x|| = 1\right\} \tag{3.11}$$

of the maximum angle between a unit vector of $\tilde{L}_1$ and the plane $\tilde{L}_2$ (see (4.6) in [S-Sz(1994)]) may be zero, whereas the cosine

$$a = \max\left\{\|P_2(x)\| \,\Big|\, x \in \tilde{L}_1, \quad ||x|| = 1\right\} \tag{3.12}$$

of the minimum such angle (see (4.5) in [S-Sz(1994)]) must be strictly between zero and one, thanks to the conditions $\tilde{L}_1 \cap \tilde{L}_2 = 0$ and $\tilde{L}_1 \not\perp \tilde{L}_2$.

However, the whole machinery of computing the commutator $[X_1, X_2]$ of the infinitesimal generators X_1, X_2 of the one-parameter rotation groups $\mathcal{G}_1 = SO\left(\tilde{L}_1\right)$, $\mathcal{G}_2 = SO\left(\tilde{L}_2\right)$ (presented in the proof of Sublemma 4.5 of [S-Sz(1994)]) still works, and we obtain the matrix expansions for X_1, X_2, and the commutator $[X_1, X_2]$

$$
\begin{aligned}
&X_1 = \begin{bmatrix} 0 & -1 & 0 & -b \\ 1 & 0 & a & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -b & 0 & -1 \\ a & 0 & 1 & 0 \end{bmatrix}, \\
&[X_1, X_2] = \begin{bmatrix} -ab & 0 & -b & 0 \\ 0 & -ab & 0 & -a \\ b & 0 & ab & 0 \\ 0 & a & 0 & ab \end{bmatrix}
\end{aligned}
\tag{3.13}
$$

written in an appropriate basis $\{e_1, f_1, e_2, f_2\}$ of $\mathbb{R}^4$. Here $e_1 \in \tilde{L}_1$ is a unit vector (out of the two possible, antipodal unit vectors) for which the maximum value a of $\|P_2(x)\|$ is attained, $e_2 = P_2(e_1)/a \in \tilde{L}_2$, whereas $f_1 \in \tilde{L}_1 \ominus e_1$ is a unit vector (out of the two possible, antipodal ones) for which the minimum value b of $\|P_2(x)\|$ is attained, and, finally, $f_2 \in \tilde{L}_2 \ominus e_2$ is a unit vector making the angle $\arccos(b)$ with f_1. Note that $\langle e_1,\, e_2\rangle = a$, $\langle f_1,\, f_2\rangle = b$, and $\langle e_i,\, f_j\rangle = 0$, $i,\, j = 1,\, 2$. In order to give the reader a little glimpse of the (not at all complicated) computation of the matrix expansions of the infinitesimal generators X_1 and X_2 in the given basis, we just write down here the computation of the, say, fourth column of the matrix of X_1. Recall that the infinitesimal generator X_1 is a 90 degree rotation in the plane $\tilde{L}_1$ transporting e_1 to f_1 (with the appropriate orientation) and acting as the zero operator in the orthocomplement of $\tilde{L}_1$ in $\mathbb{R}^4 = \tilde{L}_1 + \tilde{L}_2$. Thus, $X_1(f_2) = X_1\left(bf_1 + (f_2 - bf_1)\right) = -be_1$, for the vector bf_1 is just the orthogonal projection of f_2 onto the plane $\tilde{L}_1$ and the vector $f_2 - bf_1$ is therefore orthogonal to the plane $\tilde{L}_1$.

Observe that the first columns of the three matrices above are linearly independent, unless $b = 0$. (Recall that $0 \le b \le a < 1$ and $0 < a$.) In the case $b = 0$, however, the sums of the second and fourth columns of the above matrices are still linearly independent or, in other words, the vectors $X_1(f_1 + f_2)$, $X_2(f_1 + f_2)$, and $[X_1, X_2](f_1 + f_2)$ are linearly independent. Anyhow, we find a unit vector in $\mathbb{R}^4$ with an open orbit of the $\mathcal{G}$-action in $\mathbb{S}^3$. By Lemma 3.1, this implies the transitivity of the $\mathcal{G}$-action on the unit sphere $\mathbb{S}^3$. This completes the proof of Proposition 3.10. $\square$

Finishing the proof of Theorem 3.6. By using induction on the number k of the subspaces $L_1, L_2, \dots, L_k$ ($\dim L_i \ge 2$), we prove that the system $\{L_1, \dots, L_k\}$ is transitive, as long as it has the ONSP (see the definition preceding Lemma 3.3) in the linear span $L_1 + L_2 + \cdots + L_k$.

Indeed, the statement is obviously true for $k = 1$. Let $k > 1$, and assume that the theorem has been proven for all smaller values of k. By the ONSP of the system $\{L_1, \dots, L_k\}$, one can find two subspaces among $L_1, \dots, L_k$, say L_1 and L_2, such that $L_1 \not\perp L_2$. According to Proposition 3.10, the system $\{L_1, L_2\}$ is transitive.

The ONSP of the collection $\{L_1, \dots, L_k\}$ immediately implies the ONSP for the system $\{L_1 + L_2, L_3, \dots, L_k\}$. By the induction hypothesis, the latter system is transitive. Finally, Lemma 3.9 and the mentioned transitivity of $\{L_1, L_2\}$ yield the transitivity of $\{L_1, \dots, L_k\}$. Theorem 3.6 is now proved. $\square$

It is worth mentioning the following, easily checkable sufficient condition for transitivity:

Definition 3.14. Tightness. *We say that the system of subspaces $L_1, \dots, L_k$ is tight if and only if there exists a system of $d-1$ linearly independent vectors $e_1 \in L_{i(1)}$, $e_2 \in L_{i(2)}$, $\dots$, $e_{d-1} \in L_{i(d-1)}$ such that the linear span $\mathrm{span}\{e_1, \dots, e_{d-1}\}$ of these vectors does not contain any of the subspaces $L_{i(1)}, \dots, L_{i(d-1)}$.*

As a matter of fact, as we will see in the proof of the forthcoming lemma, under the condition of tightness, transitivity can be obtained directly by merely taking the linear span of the Lie algebras of $\mathcal{G}_i$ instead of the entire generated Lie algebra.

Lemma 3.15. *The tightness of the system $L_1, \dots, L_k$ implies the transitivity of the action of $\mathcal{G}$ on $\mathbb{S}^{d-1}$.*

Proof. Select a system of linearly independent vectors $e_1, \dots, e_{d-1}$ featuring the definition of tightness. Let $V \in \mathbb{S}^{d-1}$ be orthogonal to all vectors $e_1, \dots, e_{d-1}$. Consider one of these vectors, say, e_1. Let us denote by $P_{i(1)}$ the orthogonal projection onto the subspace $L_{i(1)}$. The projection $P_{i(1)}V$ is nonzero, for $L_{i(1)} \not\subset \mathrm{span}\{e_1, \dots, e_{d-1}\}$. Since the tangent space $\mathcal{T}_V\left(\mathcal{G}_{i(1)}V\right)$ of the orbit $\mathcal{G}_{i(1)}V \subset \mathcal{G}V$ at V is the orthocomplement $L_{i(1)} \ominus \mathrm{span}\left\{P_{i(1)}V\right\}$ of the nonzero vector $P_{i(1)}V$ in $L_{i(1)}$, we have that $e_1 \in \mathcal{T}_V\left(\mathcal{G}_{i(1)}V\right) \subset \mathcal{T}_V\left(\mathcal{G}V\right)$. By the same argument, $e_j \in \mathcal{T}_V\left(\mathcal{G}V\right)$ for $j = 1, \dots, d-1$ and, therefore, $\dim\left(\mathcal{G}V\right) = d-1$. $\square$

Basic examples related to the notion of tightness are given in Remarks 3.24 and 3.26. The first of them shows that transitivity does not imply tightness, whereas the second one provides an example of a tight family.

Characterization of the Positive Subspace of the Second Fundamental Form

We define the "parallel beam of light" B (formerly called an orthogonal manifold) around the phase point $x_0 = (Q_0, V_0) \in \mathbf{M}$ as follows:

$$B = \{x = (Q, V_0) \in \mathbf{M} : \; Q - Q_0 \perp V_0 \text{ and } ||Q - Q_0|| < \varepsilon_0\} \tag{3.16}$$

with a fixed and sufficiently small number $\varepsilon_0 > 0$. There are two interpretations of the manifold B:

(a) a $(d-1)$-dimensional submanifold of the phase space $\mathbf{M}$, or

(b) a codimension-one flat submanifold $\pi(B)$ of the configuration space $\mathbf{Q}$ supplied with a field of unit normal vectors, where $\pi : \mathbf{M} \to \mathbf{Q}$ is the natural projection.

We shall use these two interpretations alternately: when dealing with the second fundamental form we use (b), however, when defining the image $S^t(B)$ of B under the flow, we use the first interpretation. Consider a non-singular trajectory segment $S^{[0,T]}x_0$ and denote by $\mathcal{W}_+ = \mathcal{W}_+\left(S^{[0,T]}x_0\right)$ the positive subspace of the positive semi-definite, symmetric, second fundamental form W of $S^T(B)$ at the point $S^T x_0 = (Q'_0, V'_0)$. Since W acts in the orthocomplement $(V'_0)^{\perp}$ of V'_0 in $\mathbb{R}^d$, $\mathcal{W}_+$ is a subspace of $(V'_0)^{\perp}$.

With the first collision with the cylinder C_{σ_1} we associate a $(\nu_1 - 1)$-dimensional real projective space $\mathcal{P}_1 \cong \mathbb{P}^{\nu_1 - 1}(\mathbb{R})$ of all orthogonal reflections of the space $\mathbb{R}^d$ across all possible hyperplanes H that contain the generator space A_{σ_1} of the cylinder C_{σ_1}. Plainly, such reflections (or hyperplanes H) can be characterized uniquely by the orthocomplement line $H^{\perp} \subset L_{\sigma_1}$, so the collection $\mathcal{P}_1$ of all such reflections is naturally diffeomorphic to the real projective space $\mathbb{P}^{\nu_1 - 1}(\mathbb{R})$, where $\nu_1 = \dim L_{\sigma_1}$.

Similarly, other real projective spaces $\mathcal{P}_i \cong \mathbb{P}^{\nu_i - 1}(\mathbb{R})$ are attached to the symbolic collisions σ_i, $i = 1, \dots, m$. By using these definitions, we obtain a mapping

$$\Phi_\Sigma = \Phi : \quad \mathbb{S}^{d-1} \times \prod_{i=1}^{m} \mathcal{P}_i \to \mathbb{S}^{d-1} \tag{3.17}$$

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

$$(V; h_1, h_2, \dots, h_m) \in \mathbb{S}^{d-1} \times \prod_{i=1}^{m} \mathcal{P}_i$$

the image $V h_1 h_2 \cdot \ldots \cdot h_m = V'$ of $V \in \mathbb{S}^{d-1}$ under the composite action $h_1 h_2 \cdot \ldots \cdot h_m$. (Here, by convention, h_1 is applied first.) Plainly $V'_0 = \Phi(V_0; g_1, g_2, \dots, g_m)$, where g_i denotes the orthogonal reflection that causes the abrupt change of the velocity at the i-th collision σ_i of the given orbit segment $S^{[0,T]} x_0$. The space $\mathbb{S}^{d-1} \times \prod_{i=1}^{m} \mathcal{P}_i$ will often be called the phase space of the virtual dynamics, or the phase space of the velocity process.

Now we can consider the partial derivative $\dfrac{\partial \Phi}{\partial \tilde{\mathcal{P}}}(V; h_1, h_2, \dots, h_m)$ of Φ with respect to the factor $\tilde{\mathcal{P}} = \prod_{i=1}^{m} \mathcal{P}_i$: It is a linear mapping from the tangent space $\mathcal{T}_{\tilde{h}}\left(\tilde{\mathcal{P}}\right)$ into the tangent space $\mathcal{T}_{V'} \mathbb{S}^{d-1}$, where $V' = \Phi(V; h_1, h_2, \dots, h_m)$. The next result characterizes the positive subspace $\mathcal{W}_+$ as the range of the above mentioned partial derivative:

Proposition 3.18. *Using the definitions and notations from above,*

$$\mathcal{W}_+ \left(S^{[0,T]} x_0 \right) = \mathit{Ran} \left(\frac{\partial \Phi}{\partial \tilde{\mathcal{P}}} (V_0; g_1, g_2, \dots, g_m) \right).$$

Proof. The left-hand-side is obviously a subspace of the one on the right. Conversely, suppose a vector

$$Y = (y_1, y_2, \dots, y_N) \in (V'_0)^{\perp}$$

is orthogonal to $\mathcal{W}_+$. We will show that Y is orthogonal to

$$\mathrm{Ran} \left(\frac{\partial \Phi}{\partial \tilde{\mathcal{P}}} (V_0; g_1, g_2, \dots, g_m) \right),$$

as well. Because of

$$\operatorname{Ran}\left(\frac{\partial \Phi}{\partial \tilde{\mathcal{P}}}(V_0; g_1, g_2, \dots, g_m)\right) = \mathcal{L}\left\{\operatorname{Ran}\left(\frac{\partial \Phi}{\partial \mathcal{P}_i}(V_0; g_1, g_2, \dots, g_m)\right) : \; i = 1, \dots, m\right\}$$

it is sufficient to show that for every integer i $(1 \le i \le m)$

$$Y \perp \operatorname{Ran}\left(\frac{\partial \Phi}{\partial \mathcal{P}_i}(V_0; g_1, g_2, \dots, g_m)\right) := R_i.$$

For the sake of simplifying the notations we suppose that $\sigma_i = 1$. Then

$$R_i = \left\{ z g_{i+1} \dots g_m : \; z \in L_1, \; z \perp V_i^+ \right\},$$

where V_i^+ is the velocity right after the i-th collision with the cylinder C_1. By the assumed neutrality of Y with respect to $S^{[0,T]}x_0$ (which says that the vector $Y g_m^{-1} \dots g_{i+1}^{-1}$ belongs to the linear span of $A_1 = A_{\sigma_i}$ and V_i^+), we have that $Y g_m^{-1} \dots g_{i+1}^{-1} \perp z$ for every vector $z \in L_1$, $z \perp V_i^+$. Thus, by the orthogonality of the mapping $g_{i+1} \dots g_m$, we have $Y \perp R_i$.

Hence Proposition 3.18 follows. $\square$

Remark 3.19. The big advantage of the above lemma is that it gives us a new characterization of sufficiency in terms of the pure velocity process without the configuration history.

Proposition 3.18 pregnantly shows an intimate relationship between the sufficiency of a trajectory segment $S^{[0,T]}x_0$ (i. e. when the left-hand-side in the statement of that lemma has the maximum dimension $d-1$) and the transitivity of the action of $\mathcal{G}$ on the velocity sphere $\mathbb{S}^{d-1}$. This circumstance and the results of the papers [S-Sz(1994)], [S-Sz(1995)], and [Sim(1999)] (Especially Lemma 4.4 and Sublemma 4.5 in [S-Sz (1994)], and the role of those lemmas in the proof of Main Lemma 4.1 in [S-Sz(1994)]) strongly suggest the following conjectures:

Conjecture 1. *For every cylindric billiard flow* $(\mathbf{M}, \{S^t\}, \mu)$ *the transitivity of the action of* $\mathcal{G}$ *on the velocity sphere* $\mathbb{S}^{d-1}$ *is equivalent to the hyperbolic ergodicity (or, equivalently, to the hyperbolicity and Bernoulli property; cf. [Ch-H(1996)] and [O-W(1998)]) of the billiard map.*

Note. The hyperbolicity of the cylindric billiard flow alone automatically implies the transitivity of the $\mathcal{G}$-action, see Corollary A following Theorem 3.6. Thus, transitivity is equivalent to the hyperbolicity, as well.

Corollary 3.20. Dichotomy. If the above conjecture holds true, then for every cylindric billiard flow $(\mathbf{M}, \{S^t\}, \mu)$ exactly one of the following two possibilities will occur:

(I) The system $\{L_1, \dots, L_k\}$ of the orthocomplements of the generators has the ONSP in $\mathbb{R}^d$, and the billiard map is hyperbolic and enjoys the Bernoulli property;

(II) The system $\{L_1, \dots, L_k\}$ has an orthogonal splitting $\mathbb{R}^d = B_1 \oplus B_2$ (see the definition right before Lemma 3.3), and the partial kinetic energies $\|P_{B_1}(V)\|^2$ and $\|P_{B_2}(V)\|^2$ are trivial first integrals of the motion.

The above dichotomy shows that non-ergodicity can only be caused by the presence of some very simple invariant quantities, namely the kinetic energies of a "subsystems".

Corollary 3.21. Another consequence of Conjecture 1 is that "the more cylinders the better", i. e. (hyperbolic) ergodicity can not be spoilt by the removal of additional cylinders from the configuration space, that is, by adjoining more cylindric scatterers to the billiard.

Corollary 3.22. If Conjecture 1 is valid, then the phenomena of ergodicity and partial hyperbolicity (i. e. the case when there exist zero and nonzero relevant Lyapunov exponents) can not coexist in cylindric billiards.

It is interesting to note that for billiards with two cylindric scatterers the recent generalization of the result of [S-Sz(1994)] by Péter Bálint [B(1999)] actually verifies Conjecture 1 for that case.

For completeness, we recite here the conjecture appeared in [Sz(1996)] which easily turns out to be a weakened version of the previous conjecture:

Conjecture 2. *For every cylindric billiard flow* $(\mathbf{M}, \{S^t\}, \mu)$ *the existence of a single sufficient trajectory is equivalent to the hyperbolic ergodicity (or, equivalently, to the hyperbolicity and Bernoulli property) of the billiard map.*

Remark 3.23. By virtue of Proposition 3.18 it is obvious that the existence of a sufficient trajectory implies the transitivity of the $\mathcal{G}$-action. Thus Conjecture 1 is formally stronger than Conjecture 2.

Remark 3.24. Transitivity does not imply tightness. Indeed, in the model of [S-Sz(1994)] we had $d = 4$, and $\mathbb{R}^4$ was the linear direct sum of the two-dimensional subspaces L_1 and L_2. Thus, for any system $e_1 \in L_{i(1)}$, $e_2 \in L_{i(2)}$, $e_3 \in L_{i(3)}$ of linearly independent vectors either L_1 or L_2 is a subspace of $\operatorname{span}\{e_1, e_2, e_3\}$, and the system $\{L_1, L_2\}$ is not tight, although, as has been shown in [S-Sz(1994)], the $\mathcal{G}$-action is still transitive.

Remark 3.25. Orthogonal Cylindric Billiards. In his paper [Sz(1994)] the second author studied the so-called orthogonal cylindric billiards. i. e. the cylindric billiards for which each space $L_i \subset \mathbb{R}^d$ is spanned by the coordinate axes belonging to the set $K^i \subset \{1, 2, \dots, d\}$. The *sufficient and necessary condition of ergodicity* found there was the following one: There is no splitting $\{1, 2, \dots, d\} = B_1 \cup B_2$, $B_j \neq \emptyset$, $B_1 \cap B_2 = \emptyset$, such that every set K^i is the subset of B_1 or B_2. It is clear that this condition is equivalent to our ONSP.

An interesting, special family of the above orthogonal cylindric billiards is the one for which $\bigcap_{i=1}^k K^i \neq \emptyset$, $\bigcup_{i=1}^k K^i = \{1, 2, \dots, d\}$. For simplicity assume that $d \in \bigcap_{i=1}^k K^i$. Then we can select the standard coordinate unit vectors $e_1, \dots, e_{d-1}$ (in the direction of the first, second, ..., $(d-1)$-th coordinate axes), and this system obviously fulfills all requirements in the definition of tightness. By the result of [Sz(1994)], the corresponding cylindric billiard map is hyperbolic and ergodic.

Remark 3.26. Consider the case when $\dim A_j = 1$ for $j = 1, \dots, k$. Assume that not all lines A_j are parallel, for example, A_1 is not parallel to A_2. Choose a linear basis $e_1, \dots, e_d$ in $\mathbb{R}^d$ so that $\{e_1, e_3, \dots, e_d\}$ is a basis of L_1 and $\{e_2, e_3, \dots, e_d\}$ is a basis of L_2. (Thus, $\{e_3, \dots, e_d\}$ is automatically a basis of $L_1 \cap L_2$.) Then the system $\{e_1, e_2, \dots, e_{d-1}\}$ obviously fulfills all requirements in the definition of tightness. Thus, we obtained that any cylindric billiard with one-dimensional generators is necessarily tight, unless all generators are parallel. (In which case the $\mathcal{G}$-action is clearly not transitive.) We note that for the required transitivity it is actually enough to have at least two generator spaces, say, A_1 and A_2 such that $A_1 \cap A_2 = \{0\}$ and $A_1 + A_2 \neq \mathbb{R}^d$, see Case I in the proof of Proposition 3.10.

For such models the proof of (hyperbolicity and) ergodicity can go ahead straightforwardly along the lines of the proof of ergodicity developed in the papers [K-S-Sz(1989)] and [S-Sz(1995)] for two cylinders. Thus, for the case of one-dimensional generator spaces, the methods of [K-S-Sz(1989)] and [S-Sz(1995)] actually prove Conjecture 1.

4. Hard Sphere Systems

Consider the system of N (≥ 2) hard spheres, labelled by $1, 2, \dots, N$, with positive masses $m_1, \dots, m_N$ $(\sum_{i=1}^N m_i = 1)$ moving uniformly and colliding elastically in the unit ν-torus $(\nu \geq 2)$ $\mathbb{T}^\nu = \mathbb{R}^\nu / \mathbb{Z}^\nu$. For simplicity we assume that these spheres have the common radius $r > 0$, so that even the interior of the configuration space is connected. We make the standard reductions $I = \sum_{i=1}^N m_i v_i = 0$, $2E = \sum_{i=1}^N m_i \|v_i\|^2 = 1$, where $v_i = \dot{q}_i$ is the velocity of the i-th sphere, while $q_i \in \mathbb{T}^\nu$ is the position of its center. Corresponding to the reduction $I = 0$, we need to factorize the configuration space with respect to uniform translations as follows: the configuration $(q_1, q_2, \dots, q_N)$ is considered to be equivalent to another configuration $(q_1', q_2', \dots, q_N')$ iff there is an element $a \in \mathbb{T}^\nu$ such that $q_i - q_i' = a$ for every index i. After this factorization the configuration space $\mathbf{Q}$ is still a torus (of dimension $d = \nu(N-1)$) from which we remove the convex sets

$$C_{i,j} = \left\{ (q_1, \dots, q_N) \,\middle|\, \|q_i - q_j\|_e < 2r \right\},$$

$1 \leq i < j \leq N$. Here $\| \, . \, \|_e$ denotes the original Euclidean metric of $\mathbb{T}^\nu = \mathbb{R}^\nu / \mathbb{Z}^\nu$.

The relevant Riemannian metric in $\mathbf{Q}$, however, is given by (the double of) the kinetic energy:

$$\|ds\|^2 = \sum_{i=1}^N m_i \, \|dq_i\|_e^2 \,. \tag{4.1}$$

With the above Riemannian metric $\| \, . \, \|$ the configuration space $\mathbf{Q}$ is still a flat torus and — as it is easy to see — the removed scatterers $C_{i,j}$ are cylinders with a subtorus $T_{i,j}$ as the generator, and a ν-dimensional, solid disk $D_{i,j}$ as the base. Let us understand in more detail the tangent space $A_{i,j}$ of the generator subtorus $T_{i,j}$ and its orthocomplement $L_{i,j} = A_{i,j}^\perp$ with respect to the inner product given by the kinetic energy (4.1).

First of all, the common tangent space of $\mathbf{Q}$ at all of its points is the d-dimensional Euclidean space

$$(4.2)\qquad \mathcal{Z} = \left\{ z = (dq_k)_{k=1}^N \,\middle|\, \sum_{k=1}^N m_k dq_k = 0 \right\}$$

with the Euclidean structure (4.1). Secondly, the tangent space $A_{i,j} \subset \mathcal{Z}$ of $T_{i,j}$ is easily seen to be

$$(4.3)\qquad A_{i,j} = \left\{ z = (dq_k)_{k=1}^N \,\middle|\, \sum_{k=1}^N m_k dq_k = 0 \text{ and } dq_i = dq_j \right\}.$$

The orthocomplement $L_{i,j} = A_{i,j}^{\perp}$ of $A_{i,j}$ is

$$(4.4)\qquad L_{i,j} = \left\{ z = (dq_k)_{k=1}^N \,\middle|\, dq_k = 0 \text{ for } k \neq i,j \text{ and } m_i dq_i + m_j dq_j = 0 \right\}.$$

For the norm square $\|z\|^2 = m_i \|dq_i\|_e^2 + m_j \|dq_j\|_e^2$ of an element $z \in L_{i,j}$ in (4.4) we immediately have

$$(4.5)\qquad \|z\|^2 = \frac{m_i(m_i + m_j)}{m_j} \|dq_i\|_e^2 = \frac{m_j(m_i + m_j)}{m_i} \|dq_j\|_e^2.$$

Therefore, the base $D_{i,j}$ of the removed body $C_{i,j}$ is

$$(4.6)\qquad \begin{aligned} D_{i,j} &= \left\{ z = (dq_k)_{k=1}^N \in L_{i,j} \,\middle|\, \|dq_i - dq_j\|_e < 2r \right\} \\ &= \left\{ z \in L_{i,j} \,\middle|\, \|z\| < 2r \cdot \sqrt{\frac{m_i m_j}{m_i + m_j}} \right\}. \end{aligned}$$

The $(d-1)$-dimensional sphere $\mathbb{S}^{d-1}$ (on which the action of $\mathcal{G}$ is defined) is, of course, the unit sphere

$$(4.7)\qquad \mathbb{S}^{d-1} = \left\{ z = (dq_k)_{k=1}^N \in \mathcal{Z} \,\middle|\, \sum_{k=1}^N m_k \|dq_k\|_e^2 = 1 \right\}$$

of the space $\mathcal{Z}$.

Remark 4.8. In Section 2 we assumed that the cylindric billiard has the following property:

At every point $q \in \partial\mathbf{Q}$ of the boundary of the configuration space $\mathbf{Q}$ the spherical angle subtended by the compact domain $\mathbf{Q}$ is strictly positive.

The reason for postulating this axiom was to ensure that in any trajectory the collision times can not accumulate in finite time. Unfortunately, this property does not hold for *every* hard sphere system, it only holds for typical values of the radius r. (Apart from countably many exceptional values.) However, even if this property fails to hold, the set of phase points

$$A_\infty^{\pm} = \{x \in \mathbf{M} |\ S^{[0,\,b)}x \text{ contains infinitely many collisions with some } \pm b > 0\}$$

has Liouville measure zero, which is a direct consequence of the invariance of the Liouville measure. Then we can discard the zero measure union of the above sets, and this minor modification of the phase space does not influence in any way the hyperbolicity status of the hard sphere system.

The result of this section is

Proposition 4.9. *In the case of the hard sphere system $(N; \nu; r; m_1, m_2, \ldots, m_N)$ the action of $\mathcal{G}$ on $\mathbb{S}^{d-1}$ is transitive.*

Proof. By using (4.4), we will simply check the meaning of the orthogonality of two subspaces $L_{i,j}$ and $L_{k,l}$ ($1 \le i < j \le N$, $1 \le k < l \le N$). One immediately sees that $L_{i,j} \perp L_{k,l}$ if and only if the pairs $\{i, j\}$ and $\{k, l\}$ have no common element. By the connectedness of the full collision graph, the system of subspaces $\{L_{i,j} |\ 1 \le i < j \le N\}$ has the ONSP. Therefore, according to Lemma 3.4 and Theorem 3.6, the $\mathcal{G}$-action is transitive on $\mathbb{S}^{d-1}$. This finishes the proof of the proposition. $\square$

Remark 4.10. Let us call a transformation $g \in \mathcal{G}_{i,j}$ (with some pair (i, j), $1 \le i < j \le N$) elementary.

By the compactness of the sphere $\mathbb{S}^{d-1}$, there exists an upper bound

$$C = C(N; \nu; m_1, \ldots, m_N) < \infty$$

such that for every pair of velocities V_1, $V_2 \in \mathbb{S}^{d-1}$ the velocity V_2 can be obtained from V_1 by applying at most C elementary transformations or, equivalently, at most C elementary reflections. We note that every product gV, $V \in \mathbb{S}^{d-1}$, g elementary, can also be obtained as $g'V$, where g' is the product of two orthogonal reflections across hyperplanes.

Example 4.11. The upper bound $C(N; \nu; m_1, \ldots, m_N)$ is far from uniform; it depends heavily on the masses $m_1, m_2, \ldots, m_N$! Actually, the bound C may be forced to tend to infinity, as some ratios between masses blow up. A simple example showing this phenomenon is the following one:

Let $m_1 = m_2 = M >> 1$ (very big), $m_3 = m_4 = \cdots = m_N = 1$, $v_1 = v_2 = 0$, $N \ge 4$. By the theorem and the previous remark this velocity configuration can be transformed into $(V, -V, 0, 0, \ldots, 0)$ (where $\|V\|^2 = 1/(2M)$) by applying at most C elementary transformations. We can, however, easily estimate (from above) the maximum amount of kinetic energy that can be conveyed from the subsystem $\{3, 4, \ldots, N\}$ to the system $\{1, 2\}$. Assume, for instance, that the first and third spheres collide. It is easy to see that the maximum amount of such an energy exchange occurs when

(a) the collision is a "head-on" collision, i. e. the two spheres move on a straight line so that (by identifying this line with $\mathbb{R}$ by also suitably orienting it) the pre-collision velocities of the colliding spheres are $v_3^- > v_1^- > 0$;

(b) the velocities v_1^-, v_3^- have the maximum possible values.

The simple reason for (b) to hold is that by increasing either v_1^- or v_3^- the amount of energy gained by the first particle during this collision also increases. Straightforward upper bounds for these positive velocities are $v_1^- \le 1/\sqrt{M}$, $v_3^- \le 1$. An elementary calculation shows that in the extreme case $v_1^- = 1/\sqrt{M}$, $v_3^- = 1$ the after-collision velocity v_3^+ is equal to

$$\frac{2\sqrt{M} + 1 - M}{M + 1}$$

and, therefore, the maximum order of magnitude of the amount of kinetic energy conveyed from the subsystem $\{3, 4, \ldots, N\}$ to the subsystem $\{1, 2\}$ (by a single collision) is not greater than $2/\sqrt{M}$. Thus, in order for the subsystem $\{3, \dots, N\}$ to lose its energy $1/2$, there must be at least $\sqrt{M}/4$ collisions between the two subsystems of heavy and light particles, i. e. $C(N; \nu; M, M, 1, 1, \dots, 1) \geq \sqrt{M}/4$, $M >> 1$.

Remark 4.12. Arbitrary collision graphs. Suppose that not all cylinders $C_{i,j}$ are removed from the original, $\nu(N-1)$-dimensional configuration torus, but, instead, only the cylinders whose pair of labels $\{i, j\}$ belongs to the set of edges $E(G)$ of a prescribed graph G of allowed collisions on the set of vertices $\{1, 2, \dots, N\}$. Then the proof of Proposition 4.9 yields that the $\mathcal{G}$-action corresponding to the arising cylindric billiard is transitive if and only if the graph G is connected on the whole vertex set $\{1, 2, \dots, N\}$. Hyperbolic ergodicity, that is, actually Conjecture 1 was proved to be true in [S-Sz(1995)] for such classes of generalized hard sphere systems in the case $\nu \geq 4$, provided that the connected graph G is either the simple path of length $N-1$ or the simple loop of length N. For such graphs of allowed collisions, in the case of $\nu = 3$, the complete hyperbolicity of the flow is proved there, too.

Finally, by taking into account (4.3), observe the interesting fact that, in the realm of such generalized hard sphere systems, the condition $\bigcap_{(i,j)\in E(G)} A_{i,j} = 0$ is also equivalent to the connectedness of the graph G. Thus, at least for these generalized hard sphere systems, the transitivity and the otherwise much weaker property $\bigcap_{i,j} A_{i,j} = 0$ are equivalent.

References

[B(1999)] P. Bálint, *Chaotic and Ergodic Properties of Cylindric Billiards*, Manuscript.

[Ch-H(1996)] N. I. Chernov, C. Haskell, *Nonuniformly Hyperbolic K-systems are Bernoulli*, Ergodic Thoery and dynamical Systems **16**, 19-44 (1996).

[G(1981)] G. Galperin, *On Systems of Locally Interacting and Repelling Particles Moving in Space*, Trudy MMO **43**, 142-196 (1981).

[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).

[O-W(1998)] D. Ornstein, B. Weiss, *On the Bernoulli Nature of Systems with Some Hyperbolic Structure*, Ergodic Theory and Dynamical Systems **18**, 441-456.

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

[Sim(1999)] N. Simányi, *Ergodicity of Hard Spheres in a Box*, Ergodic theory and dynamical systems **19**, 741-766.

[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(1994)] N. Simányi, D. Szász, *The K-property of 4-D Billiards with Non-Orthogonal Cylindric Scatterers*, J. Stat. Phys. **76, Nos. 1/2**, 587-604 (1994).

[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**, 35-96.

[Sz(1993)] D. Szász, *Ergodicity of Classical Billiard Balls*, Physica A **194**, 86-92 (1993).

[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).