Chaotic properties of billiards in circular polygons*

Andrew Clarke† Rafael Ramírez-Ros‡

July 31, 2024

Abstract

We study billiards in domains enclosed by circular polygons. These are closed C^1 strictly convex curves formed by finitely many circular arcs. We prove the existence of a set in phase space, corresponding to generic sliding trajectories close enough to the boundary of the domain, in which the return billiard dynamics is semiconjugate to a transitive subshift on infinitely many symbols that contains the full N-shift as a topological factor for any $N \in \mathbb{N}$, so it has infinite topological entropy. We prove the existence of uncountably many asymptotic generic sliding trajectories approaching the boundary with optimal uniform linear speed, give an explicit exponentially big (in q) lower bound on the number of q-periodic trajectories as $q \to \infty$, and present an unusual property of the length spectrum. Our proofs are entirely analytical.

Keywords: Billiards, circular polygons, chaos, symbolic dynamics, periodic trajectories, length spectrum

1 Introduction

A *billiard problem* concerns the motion of a particle inside the domain bounded by a closed plane curve Γ (or the domain bounded by a hypersurface of some higher-dimensional Euclidean space). The motion in the interior of the domain is along straight lines, with elastic collisions at the boundary according to the optical law of reflection: the angles of incidence and reflection are equal. These dynamical systems were first introduced by Birkhoff [9]. See [40, 56, 17] for a general description.

In the case of dispersing billiards (i.e. when the boundary is a union of concave components), the dynamics is chaotic [54]; indeed, such billiards exhibit ergodicity, the Bernoulli property, sensitive dependence on initial conditions, and so forth. In fact, it was believed for some years that billiards without any dispersing walls could not display chaos.

Thus it came as a surprise when Bunimovich, in his famous paper, detailed a proof that the billiard in a stadium exhibits the Bernoulli property [12]. The boundary of the stadium billiard consists of two straight parallel lines connected at either end by semicircles. Stadia are C^1 and convex, but not C^2 or strictly convex. We study the class of C^1 *strictly* convex billiards bounded by finitely many circular arcs. No billiard in this class satisfies the celebrated *B-condition*—that is, the condition that all circular arcs can be completed to a disk within the billiard domain—, which is the hallmark for the *defocusing mechanism* in billiards whose focusing boundaries are all circular arcs [17, Section 8.3]. In spite of this, we observe several chaotic phenomena.

Denote by Γ a C^1, closed, strictly convex curve in $\mathbb{R}^2$. The phase space of the billiard inside the domain bounded by Γ is the 2-dimensional cylinder $\mathcal{M} = \mathbb{T} \times [0, \pi]$; the angular component of the cylinder is a parameter on the curve Γ, and the height component is the angle of incidence/reflection. We denote by $f : \mathcal{M} \to \mathcal{M}$ the billiard map (i.e. the collision map of the billiard flow with the boundary Γ of the domain; see Section 4 for a precise definition). We say that the curve Γ is a *circular polygon* if it is a union of a finite number of circular arcs, concatenated in such a way that Γ is strictly convex, and at the points where two circular arcs with different radii meet, the tangents agree, so that Γ is C^1, but not C^2. We do not consider circles as circular polygons. A *circular k-gon*

*To appear in *Communications in Mathematical Physics*

†Universitat Politècnica de Catalunya, Barcelona, Spain (andrew.michael.clarke@upc.edu)

‡Universitat Politècnica de Catalunya, Barcelona, Spain (rafael.ramirez@upc.edu)

is a circular polygon with exactly k circular arcs. Note that a circular polygon cannot have fewer than 4 circular arcs, so if Γ is a circular k-gon, then $k \geq 4$ (see Lemma 4 in Section 2 for a proof of this fact).

In this paper we are interested in the *sliding* trajectories that do not skip any arc in any of their infinite turns around Γ. These are trajectories close to the boundary $\partial\mathcal{M} = \mathbb{T} \times \{0, \pi\}$. That is, trajectories where the angle of incidence/reflection is close to 0 or π (see Definition 13).

In what follows we give heuristic statements of our main results.

Theorem A. *If Γ is a circular polygon, then there is a set $\mathcal{J} \subset \mathcal{M}$ accumulating on $\partial\mathcal{M}$ such that the return map $F : \mathcal{J} \to \mathcal{J}$ of f to $\mathcal{J}$ is topologically semiconjugate to a transitive subshift on infinitely many symbols that contains the full N-shift as a topological factor for any $N \in \mathbb{N}$, so it has infinite topological entropy.*

See Proposition 27 in Section 5 and Theorem 31 in Section 6 for a precise formulation of this theorem. Be aware that the map with infinite entropy is the return map F, not the billiard map f.

The *final sliding motions* are the possible qualitative behaviors that a sliding billiard trajectory posses as the number of impacts tends to infinity, forward or backward. Every forward counter-clockwise sliding billiard trajectory $(\varphi_n, \theta_n) = f^n(\varphi, \theta)$, where φ_n are the angles of impact and θ_n are the angles of incidence/reflection, belongs to exactly one of the following three classes:

- *Forward bounded* ($\mathcal{B}_0^+$): $\inf_{n\geq 0} \theta_n > 0$;
- *Forward oscillatory* ($\mathcal{O}_0^+$): $0 = \liminf_{n\to+\infty} \theta_n < \limsup_{n\to+\infty} \theta_n$; and
- *Forward asymptotic* ($\mathcal{A}_0^+$): $\lim_{n\to+\infty} \theta_n = 0$.

This classification also applies for backward counter-clockwise sliding trajectories when $n \leq -1$ and $n \to -\infty$, in which case we write a superindex $-$ instead of $+$ in each of the classes: $\mathcal{B}_0^-$, $\mathcal{O}_0^-$ and $\mathcal{A}_0^-$. And it also applies to (backward or forward) clockwise sliding trajectories, in which case we replace θ_n with $|\theta_n - \pi|$ in the definitions above and we write a subindex π instead of 0 in each of the classes: $\mathcal{B}_\pi^\pm$, $\mathcal{O}_\pi^\pm$ and $\mathcal{A}_\pi^\pm$.

Terminologies *bounded* and *oscillatory* are borrowed from Celestial Mechanics. See, for instance, [31]. In our billiard setting, bounded means bounded away from $\theta = \pi$ in the clockwise case and bounded away from $\theta = 0$ in the counter-clockwise case. That is, a sliding billiard trajectory is bounded when it does not approach $\partial\mathcal{M}$.

The following corollary is an immediate consequence of Theorem A, see Section 6.

Corollary B. *If Γ is a circular polygon, then $\mathcal{X}_\lambda^- \cap \mathcal{Y}_\lambda^+ \neq \emptyset$ for $\mathcal{X}, \mathcal{Y} \in \{\mathcal{B}, \mathcal{O}, \mathcal{A}\}$ and $\lambda \in \{0, \pi\}$.*

From now on, we focus on counter-clockwise sliding trajectories. Corollary B does not provide data regarding the maximal *speed* of diffusion for asymptotic trajectories. Which is the faster way in which $\theta_n \to 0$ for asymptotic sliding trajectories? The answer is provided in the following theorem.

Theorem C. *If Γ is a circular polygon, then there are uncountably many asymptotic generic sliding billiard trajectories that approach the boundary with uniform linear speed. That is, there are constants $0 < a < b$ such that if $\{\theta_n\}_{n\in\mathbb{Z}}$ is the corresponding sequence of angles of incidence/reflection of any of these uncountably many asymptotic generic sliding billiard trajectories, then*

$$a|n| \leq 1/\theta_n \leq b|n|, \qquad \forall |n| \gg 1.$$

Linear speed is optimal. That is, there is no billiard trajectory such that

$$\lim_{n\to+\infty} n\theta_n = 0.$$

See Theorem 35 in Section 7, where we also get uncountably many one-parameter families (paths) of forward asymptotic generic sliding billiard trajectories, for a more detailed version of Theorem C. The definition of *generic* billiard trajectories is a bit technical, see Definition 13 and Remark 14. The term *uniform* means that the constants $0 < a < b$ do not depend on the billiard trajectory. The term *linear* means that $1/\theta_n$ is bounded between two positive multiples of $|n|$. *Optimality* comes as no surprise since $\sum_{n\geq 0} \theta_n = +\infty$ for any billiard trajectory in any circular polygon—or, for that matter, in any strictly convex billiard table whose billiard flow is defined for all time [32]. Optimality is proved in Proposition 36 in Section 7.

There are two key insights (see Section 4) behind this theorem. First, when we iterate f along one of the circular arcs of the circular polygon Γ, the angle of reflection θ is constant, so θ can drop

only when the trajectory crosses the singularities between consecutive circular arcs. Second, the maximal drop corresponds to multiplying θ by a uniform (in θ) factor that is smaller than one. As we must iterate the map many (order θ^{-1}) times to slide fully along each circular arc, we cannot approach the boundary with a faster than linear speed.

As Theorem A gives us only a topological *semi*conjugacy to symbolic dynamics, it does not immediately provide us with the abundance of periodic orbits that the shift map possesses. However our techniques enable us to find many periodic sliding billiard trajectories. We state in the following theorem that the number of such trajectories in circular polygons grows exponentially with respect to the period. In contrast, Katok [37] showed that the numbers of isolated periodic billiard trajectories and of parallel periodic billiard trajectories grow subexponentially in any (linear) polygon.

Given any integers $1 \leq p < q$, let $\Pi(p,q)$ be the set of (p,q)-periodic billiard trajectories in Γ. That is, the set of periodic trajectories that close after p turns around Γ and q impacts in Γ, so they have rotation number p/q. Let $\Pi(q) = \cup_{1\leq p<q}\Pi(p,q)$ be the set of periodic billiard trajectories with period q. The symbol $\#$ denotes the *cardinality* of a set.

Theorem D. *If Γ is a circular k-gon and $p \in \mathbb{N}$, there are constants $c_\star(p), M_\star, h_\star > 0$ such that*

(a) $\#\Pi(p,q) \geq c_\star(p)q^{kp-1} + \mathrm{O}(q^{kp-2})$ *as $q \to \infty$ for any fixed $p \in \mathbb{N}$; and*

(b) $\#\Pi(q) \geq M_\star \mathrm{e}^{h_\star q}/q$ *as $q \to +\infty$.*

We give explicit expressions for $c_\star(p)$, $M_\star$ and $h_\star$ in Section 8. The optimal value of $c_\star(p)$ is equal to the volume of a certain $(kp-1)$-dimensional compact convex polytope with an explicitly known half-space representation, see Proposition 39. We do not give optimal values of $M_\star$ and $h_\star$. The relation between the optimal value of $h_\star$ and the topological entropy of the billiard map f is an open problem. We acknowledge that some periodic trajectories in $\Pi(p,q)$ may have period less than q when $\gcd(p,q) \neq 1$, but they are a minority, so the previous lower bounds capture the growth rate of the number of periodic trajectories with rotation number p/q and minimal period q even when p and q are not coprime.

If the circular polygon has some symmetry, we can perform the corresponding natural reduction to count the number of symmetric sliding periodic trajectories, but then the exponent $kp-1$ in the first lower bound would be smaller because there are fewer reduced arcs than original arcs. The exponent $h_\star$ would be smaller too. See [15, 27] for examples of symmetric periodic trajectories in other billiards. The first reference deals with axial symmetries. The second one deals with rotational symmetries.

Let $|\Gamma|$ be the length of Γ. If $g = \{z_0, \ldots, z_{q-1}\} \subset \Gamma$ is a q-periodic billiard trajectory, let $L(g) = |z_1 - z_0| + \cdots + |z_{q-1} - z_0|$ be its *length*. If $(g_q)_q$ is any sequence such that $g_q \in \Pi(1,q)$, then $\lim_{q\to+\infty} L(g_q) = |\Gamma|$. There are so many generic sliding $(1,q)$-periodic billiard trajectories inside circular polygons that we can find sequences $(g_q)_q$ such that the differences $L(g_q) - |\Gamma|$ have rather different asymptotic behaviors as $q \to +\infty$.

Theorem E. *If Γ is a circular polygon, then there are constants $c_- < c_+ < 0$ such that for any fixed $c \in [c_-, c_+]$ there exists a sequence $(g_q)_q$, with $g_q \in \Pi(1,q)$, such that*

$$L(g_q) = |\Gamma| + c/q^2 + \mathrm{O}(1/q^3), \quad \text{as } q \to +\infty.$$

Consequently, there exists a sequence $(h_q)_q$, with $h_q \in \Pi(1,q)$, such that

$$c_- = \liminf_{q\to+\infty} \big((L(h_q) - |\Gamma|)q^2\big) < \limsup_{q\to+\infty} \big((L(h_q) - |\Gamma|)q^2\big) = c_+, \quad \text{as } q \to +\infty.$$

Besides, $c_- \leq -\pi^2|\Gamma|/6$ and $c_+ = -\frac{1}{24}\left[\int_\Gamma \kappa^{2/3}(s)\,\mathrm{d}s\right]^3$, where $\kappa(s)$ is the curvature of Γ as a function of an arc-length parameter $s \in [0, |\Gamma|)$.

Let us put these results into perspective by comparing them with the observed behavior in sufficiently smooth (say C^6) and strictly convex billiards, which for the purpose of this discussion we refer to as *Birkhoff billiards*. Lazutkin's theorem (together with a refinement due to Douady) implies that Birkhoff billiards possess a family of caustics[1] accumulating on the boundary [22, 41]. These caustics divide the phase space into invariant regions, and therefore guarantee a certain *regularity* of the dynamics near the boundary, in the sense that the conclusion of Theorem A never holds for

[1] A closed curve γ contained in the interior of the region bounded by Γ is called a *caustic* if it has the following property: if one segment of a billiard trajectory is tangent to γ, then so is every segment of that trajectory.

Birkhoff billiards. Not only does the conclusion of Theorem C not hold for Birkhoff billiards, but in such systems there are no trajectories approaching the boundary asymptotically as the orbits remain in invariant regions bounded by the caustics. As for Theorem D, a well-known result of Birkhoff [9] implies that Birkhoff billiards have $\#\Pi(p,q) \geq 2$ for each coprime pair p, q such that $1 \leq p < q$. This lower bound turns out to be sharp, in the sense that for any such pair p, q, there exist Birkhoff billiards with exactly two geometrically distinct periodic orbits of rotation number p/q [50]; a simple example is that the billiard in a non-circular ellipse has two periodic orbits of rotation number $1/2$, corresponding to the two axes of symmetry. It follows that the conclusion of Theorem D does not hold in general for Birkhoff billiards. Finally, as for Theorem E, a well-known result of Marvizi-Melrose [44] implies that if $(g_q)_q$, with $g_q \in \Pi(1,q)$, is any sequence of periodic billiard trajectories in a Birkhoff billiard Γ, then

$$L(g_q) = |\Gamma| + c_+/q^2 + \mathrm{O}(1/q^4), \quad \text{as } q \to +\infty,$$

where $c_+ = -\frac{1}{24}\left[\int_\Gamma \kappa^{2/3}(s)\,\mathrm{d}s\right]^3$. Hence, $(1,q)$-periodic billiard trajectories in circular polygons are *asymptotically shorter* than the ones in Birkhoff billiards.

An interesting question in general that has been considered to a significant extent in the literature is: what happens to the caustics of Lazutkin's theorem, and thus the conclusions of Theorems A and C, if we loosen the definition of a Birkhoff billiard? Without altering the basic definition of the billiard map f, there are three ways that we can generalise Birkhoff billiards: (i) by relaxing the strict convexity hypothesis, (ii) by relaxing the smoothness hypothesis, or (iii) by increasing the dimension of the ambient Euclidean space.

(i) Mather proved that if the boundary is convex and C^r for $r \geq 2$, but has at least one point of zero curvature, then there are no caustics and there exist trajectories which come arbitrarily close to being positively tangent to the boundary and also come arbitrarily close to being negatively tangent to the boundary [45]. Although this result is about finite segments of billiard trajectories, there are also infinite trajectories tending to the boundary both forward and backward in time in such billiards: $\mathcal{A}_0^{\pm} \cap \mathcal{A}_\pi^{\mp} \neq \emptyset$, see [46].

(ii) Despite six continuous derivatives being the stated smoothness requirement for Lazutkin's theorem [22, 41], there is some uncertainty regarding what happens for C^5 boundaries, and in fact it is generally believed that 4 continuous derivatives should suffice. Halpern constructed billiard tables that are strictly convex and C^1 but not C^2 such that the billiard particle experiences an infinite number of collisions in finite time [32]; that is to say, the billiard flow is incomplete. This construction does not apply to our case, as our billiard boundaries have only a finite number of singularities (points where the boundary is only C^1 and not C^2), whereas Halpern's billiards have infinitely many.

The case of boundaries that are strictly convex and C^1 but not C^2 and have only a finite number (one, for example) of singularities was first considered by Hubacher [35], who proved that such billiards have no caustics in a neighborhood of the boundary. This result opens the door for our analysis.

(iii) It has been known since the works of Berger and Gruber that in the case of strictly convex and sufficiently smooth billiards in higher dimension (i.e. the billiard boundary is a codimension 1 submanifold of $\mathbb{R}^d$ where $d \geq 3$), only ellipsoids have caustics [7, 29]. However Gruber also observed that in this case, even in the absence of caustics, the Liouville measure of the set of trajectories approaching the boundary asymptotically is zero [28]. The question of existence of such trajectories was thus left open.

It was proved in [19] (combined with results of [18]) that generic strictly convex analytic billiards in $\mathbb{R}^3$ (and 'many' such billiards in $\mathbb{R}^d$ for $d \geq 4$) have trajectories approaching the boundary asymptotically. It is believed that the meagre set of analytic strictly convex billiard boundaries in $\mathbb{R}^d$, $d \geq 3$, for which these trajectories do not exist consists entirely of ellipsoids, but the perturbative methods of [19] do not immediately extend to such a result.

Billiards in circular polygons have been studied numerically in the literature [6, 23, 33, 34, 43]. In the paper [4] the authors use numerical simulations and semi-rigorous arguments to study billiards in a 2-parameter family of circular polygons. They conjecture that, for certain values of the parameters, the billiard is ergodic. In addition they provide heuristic arguments in favor of this conjecture.

A related problem is the lemon-shaped billiard, which is known to display chaos [11, 16, 36]. These billiards are strictly convex but not C^1, so the billiard map is well-defined only on a proper subset of the phase space.

The *elliptic flowers* recently introduced by Bunimovich [13] are closed C^0 curves formed by finitely many pieces of ellipses. *Elliptic polygons* are elliptic flowers that are C^1 and strictly convex, so they are a natural generalisation of circular polygons. One can obtain a 1-parameter family of elliptic polygons with the string construction from any convex (linear) polygon. The *string construction* consists of wrapping an inelastic string around the polygon and tracing a curve around it by keeping the string taut. Billiards in elliptic polygons can be studied with the techniques presented here for circular polygons. We believe that all results previously stated in this introduction, with the possible exception of the inequality $c_- \leq -\pi^2|\Gamma|/6$ given in Theorem E, hold for generic elliptic polygons. However, there are elliptic polygons that are globally C^2, and not just C^1, the *hexagonal string billiard* first studied by Fetter [24] being the most celebrated example. We do not know how to deal with C^2 elliptic polygons because jumps in the curvature of the boundary are a key ingredient in our approach to get chaotic billiard dynamics. Fetter suggested that the hexagonal string billiard could be integrable, in which case it would be a counterexample to the Birkhoff conjecture[2]. However, such integrability was numerically put in doubt in [8].

In what follows we describe the main ideas of our proofs. Let Γ be a circular k-gon. It is well-known that the angle of incidence/reflection is a constant of motion for billiards in circles. Therefore, for the billiard in Γ, the angle of incidence/reflection can change only when we pass from one circular arc to another, and not when the billiard has consecutive impacts on the same circular arc. The main tool that we use to prove our theorems is what we call the *fundamental lemma* (Lemma 18 below), which describes how trajectories move up and down after passing from one circular arc to the next. The phase space $\mathcal{M}$ of the billiard map f is a cylinder, with coordinates (φ, θ) where $\varphi \in \mathbb{T}$ is a parameter on the boundary Γ, and where $\theta \in [0, \pi]$ is the angle of incidence/reflection. We consider two vertical segments $\mathcal{L}_j$ and $\mathcal{L}_{j+1}$ in $\mathcal{M}$ corresponding to consecutive singularities of Γ, and sufficiently small values of θ. The index j that labels the singularities is defined modulo k. The triangular region $\mathcal{D}_j$ bounded by $\mathcal{L}_j$ and $f(\mathcal{L}_j)$ is a fundamental domain of the billiard map f; that is, a set with the property that sliding trajectories have exactly one point in $\mathcal{D}_j$ on each turn around Γ. Consider now the sequence of backward iterates $\left\{f^{-n}(\mathcal{L}_{j+1})\right\}$ of $\mathcal{L}_{j+1}$. This sequence of slanted segments divides the fundamental domain $\mathcal{D}_j$ into infinitely many quadrilaterals, which we call *fundamental quadrilaterals*. The fundamental lemma describes which fundamental quadrilaterals in $\mathcal{D}_{j+1}$ we can visit if we start in a given fundamental quadrilateral in $\mathcal{D}_j$.

In order to prove Theorem A, we apply the fundamental lemma iteratively to describe how trajectories visit different fundamental quadrilaterals consecutively in each of the k fundamental domains $\mathcal{D}_j$ in $\mathcal{M}$. A particular coding of possible sequences of k fundamental quadrilaterals that trajectories can visit gives us our symbols. We then use a method due to Papini and Zanolin [48, 49] (extended to higher dimensions by Pireddu and Zanolin [51, 52, 53]) to prove that the billiard dynamics is semiconjugate to a shift map on the sequence space of this set of symbols; this method is called *stretching along the paths*. Observe that we could equally have used the method of correctly aligned windows [3, 26, 58], or the crossing number method [39]; note however that the latter would not have provided us with the large amount of periodic orbits that the other two methods do. We note that, although Theorem A provides us with a topological semiconjugacy to symbolic dynamics, we expect that this could be improved to a full conjugacy by using other methods.

Once the proof of Theorem A is completed, Theorems C, D, and E are proved by combining additional arguments with the symbolic dynamics we have constructed. With respect to the symbolic dynamics, we choose a coding of the fundamental quadrilaterals visited by a trajectory that corresponds to θ tending to 0 in the fastest way possible. We then prove that the corresponding billiard trajectories satisfy the conclusion of Theorem C. As for Theorem D, the method of stretching along the paths guarantees the existence of a periodic billiard trajectory for every periodic sequence of symbols. Consequently, the proof of Theorem D amounts to counting the number of sequences of symbols that are periodic with period p (because each symbol describes one full turn around the table; see Section 5 for details) such that the corresponding periodic sliding billiard trajectories after p turns around the table have rotation number p/q. It turns out that this reduces to counting the number of integer points whose coordinates sum q in a certain kp-dimensional convex polytope. We do

[2]It is well-known that billiards in ellipses are integrable. The so-called *Birkhoff conjecture* says that elliptical billiards are in fact the only integrable billiards. This conjecture, in its full generality, remains open.

this by proving that the given convex polytope contains a hypercube with sides of a certain length, and finally by counting the number of integer points whose coordinates sum q in that hypercube.

The structure of this paper is as follows. In Section 2 we describe the salient features of circular polygons. We summarise the *stretching along the paths* method in Section 3. In Section 4 we define the billiard map and dicuss the billiard dynamics in circular polygons, before giving the definition of fundamental quadrilaterals, as well as the statement and proof of the fundamental lemma. Symbols are described in Section 5. Chaotic dynamics, and thus the proofs of Theorem A and Corollary B, is established in Section 6, whereas Section 7 contains the proof of Theorem C. In Section 8, we count the periodic orbits, thus proving Theorem D. Finally, Theorem E is proved in Section 9. Some technical proofs are relegated to appendices.

2 Circular polygons

In this section we define our relevant curves, construct their suitable parametrisations, and introduce notations that will be extensively used in the rest of the paper.

A *piecewise-circular curve* (or *PC curve* for short) is given by a finite sequence of circular arcs in the Euclidean plane $\mathbb{R}^2$, with the endpoint of one arc coinciding with the beginning point of the next. PC curves have been studied by several authors. See [5, 14] and the references therein. Lunes and lemons (two arcs), yin-yang curves, arbelos and PC cardioids (three arcs), salinons, Moss's eggs and pseudo-ellipses (four arcs), and Reuleaux polygons (arbitrary number of arcs) are celebrated examples of simple closed PC curves [21, 1]. A simple closed PC curve is a PC curve not crossing itself such that the endpoint of its last arc coincides with the beginning point of its first arc.

All simple closed PC curves are Jordan curves, so we could study the billiard dynamics in any domain enclosed by a simple closed PC curve. However, such domains are too general for our purposes. We will only deal with strictly convex domains without corners or cusps. Strict convexity is useful, because then any ordered pair of points on the boundary defines a unique billiard trajectory. Absence of cusps and corners implies that the corresponding billiard map is a global homeomorphism in the phase space $\mathcal{M}$, see Section 4.

Therefore, we will only consider *circular polygons*, defined as follows.

Definition 1. A *circular k-gon* is a simple closed strictly convex curve in $\mathbb{R}^2$ formed by the concatenation of $k > 1$ circular arcs, in such a way that the curve is C^1, but not C^2, at the intersection points of any two consecutive circular arcs. The *nodes* of a circular polygon are the intersection points of each pair of consecutive circular arcs.

Reuleaux polygons, lemons, lunes, yin-yang curves, arbelos, salinons, PC cardioids are not circular polygons, but pseudo-ellipses and Moss's eggs (described later on; see also [21, Section 1.1]) are. We explicitly ask that consecutive arcs always have different radii, so the curvature has jump discontinuities at all nodes. We do not consider circumferences as circular polygons since circular billiards are completely integrable.

Let Γ be a circular k-gon with arcs $\Gamma_1, \dots, \Gamma_k$, listed in the order in which they are concatenated, moving in a counter-clockwise direction. Each arc Γ_j is completely determined by its *center* O_j, its *radius* $r_j > 0$ and its *angular range* $[a_j, b_j] \subset \mathbb{T}$. Then $\delta_j = b_j - a_j$ is the *central angle* of Γ_j. Using the standard identification $\mathbb{R}^2 \simeq \mathbb{C}$, let

$$A_j = O_j + r_j \mathrm{e}^{\mathrm{i}a_j}, \qquad B_j = O_j + r_j \mathrm{e}^{\mathrm{i}b_j}$$

be the two nodes of arc Γ_j. We denote by

$$\Gamma_\star = \{A_1, \dots, A_k\} = \{B_1, \dots, B_k\} \tag{1}$$

the *set of nodes* of Γ.

Notation 2. The index j that labels the arcs of any circular k-gon is defined modulo k. Hence, $\Gamma_j = \Gamma_{j \mod k}$, $r_j = r_{j \mod k}$, $a_j = a_{j \mod k}$ and so forth. In particular, $\Gamma_{k+1} = \Gamma_1$.

Definition 3. The *polar parametrisation* of Γ is the counter-clockwise parametrisation

$$z : \mathbb{T} \to \Gamma \subset \mathbb{R}^2 \simeq \mathbb{C}, \qquad z(\varphi) = O_j + r_j \mathrm{e}^{\mathrm{i}\varphi}, \qquad \forall \varphi \in [a_j, b_j].$$

The points $a_1, \dots, a_k$ are the *singularities* of Γ.

This parametrisation is well-defined because, by definition, $B_j = A_{j+1}$ (the endpoint of any arc coincides with the beginning point of the next), and $b_j = a_{j+1}$ (two consecutive arcs have the same oriented tangent line at their intersecting node). From now on, the reader should keep in mind that singularities $a_1, \dots, a_k$ are always ordered in such a way that

$$a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_{k-1} = a_k < b_k = a_1 + 2\pi. \tag{2}$$

As far as we know, all the billiards in circular polygons that have been studied in the past correspond to cases with *exactly* four arcs [4, 6, 23, 33, 34, 43]. It turns out that this is the simplest case, in the context of the next lemma.

Lemma 4. *Let Γ be a circular k-gon with radii $r_j > 0$, singularities $a_j \in \mathbb{T}$ (or $b_j = a_{j+1}$) and central angles $\delta_j = b_j - a_j \in (0, 2\pi)$. Set $w_j = \mathrm{e}^{\mathrm{i}b_j} - \mathrm{e}^{\mathrm{i}a_j} \in \mathbb{C}$. Then Γ has at least four arcs: $k \geq 4$, and*

$$\sum_{j=1}^{k} \delta_j = 2\pi, \qquad \sum_{j=1}^{k} r_j w_j = 0. \tag{3}$$

Proof. Clearly, $\sum_{j=1}^{k} \delta_j = \sum_{j=1}^{k} (b_j - a_j) = b_k - a_1 = 2\pi$. It is known that a bounded measurable function $\rho \in L(\mathbb{T})$ is the radius of curvature of a closed curve if and only if

$$\int_{a_1}^{b_k} \rho(\varphi) \mathrm{e}^{\mathrm{i}\varphi} \, \mathrm{d}\varphi = \int_0^{2\pi} \rho(\varphi) \mathrm{e}^{\mathrm{i}\varphi} \, \mathrm{d}\varphi = 0. \tag{4}$$

Since the radius of curvature of Γ is the piecewise constant function $\rho_{|(a_j, b_j)} \equiv r_j$, the general condition (4) becomes $-\mathrm{i} \sum_{j=1}^{k} r_j w_j = 0$. Note that $\sum_{j=1}^{k} w_j = 0$, $w_j \neq 0$ for all j, and $\dim_{\mathbb{R}}[w_1, \dots, w_k] = 2$ when $k \geq 3$. If Γ has just two arcs: $k = 2$, then

$$r_1 w_1 + r_2 w_2 = 0, \qquad w_1 + w_2 = 0, \qquad w_1, w_2 \neq 0.$$

This implies that $r_1 = r_2$ and contradicts our assumption about radii of consecutive arcs. If Γ has just three arcs: $k = 3$, then

$$r_1 w_1 + r_2 w_2 + r_3 w_3 = 0, \qquad w_1 + w_2 + w_3 = 0, \qquad \dim_{\mathbb{R}}[w_1, w_2, w_3] = 2.$$

This implies that $r_1 = r_2 = r_3$ and we reach the same contradiction. □

Necessary conditions (2) and (3) are sufficient ones too. To be precise, if the radii $r_j > 0$, the angular ranges $[a_j, b_j] \subset \mathbb{T}$ and the central angles $\delta_j = b_j - a_j \in (0, 2\pi)$ satisfy (2) and (3), then there exists a 2-parameter family of circular k-gons sharing all those elements. Moreover, all circular k-gons in this family are the same modulo translations. Let us prove this claim. Once we put the center O_1 at an arbitrary location, all other centers are recursively determined by imposing that $A_{j+1} = B_j$, which implies (since $b_j = a_{j+1}$) that

$$O_{j+1} = O_j + (r_j - r_{j+1}) \mathrm{e}^{\mathrm{i}b_j}, \qquad j = 1, \dots, k-1.$$

The obtained PC curve $\Gamma = z(\mathbb{T})$, where $z(\varphi)$ is the polar parametrisation introduced in Definition 3, is closed by (4) and it is C^1 and strictly convex by construction. Hence, Γ is a circular k-gon. This means that any circular k-gon is completely determined once we know its first center O_1, its first singularity a_1, its radii $r_1, \dots, r_k$, and its central angles $\delta_1, \dots, \delta_k$.

The above discussion shows that circular k-gons form, modulo translations and rotations, a $(2k-3)$-parameter family. More precisely, if we set $O_1 = (0, 0)$ and $a_1 = 0$ by means of a translation and a rotation, then parameters $r_1, \dots, r_k, \delta_1, \dots, \delta_k$ are restricted by (3), which has codimension three. If, in addition, we normalise somehow (in the literature one can find many different choices) our circular k-gons with a scaling, we get that they form a $(2k-4)$-parameter family modulo similarities. The reader can find a complete geometric description, modulo similarities, of the four-parameter family of (convex and nonconvex, symmetric and nonsymmetric) closed C^1 PC curves with four arcs in [33], whose goal was to exhibit numerically the richness of the billiard dynamics in those C^1 PC curves.

For brevity, we only give a few simple examples of symmetric and non-symmetric circular polygons with four and six arcs. We skip many details.

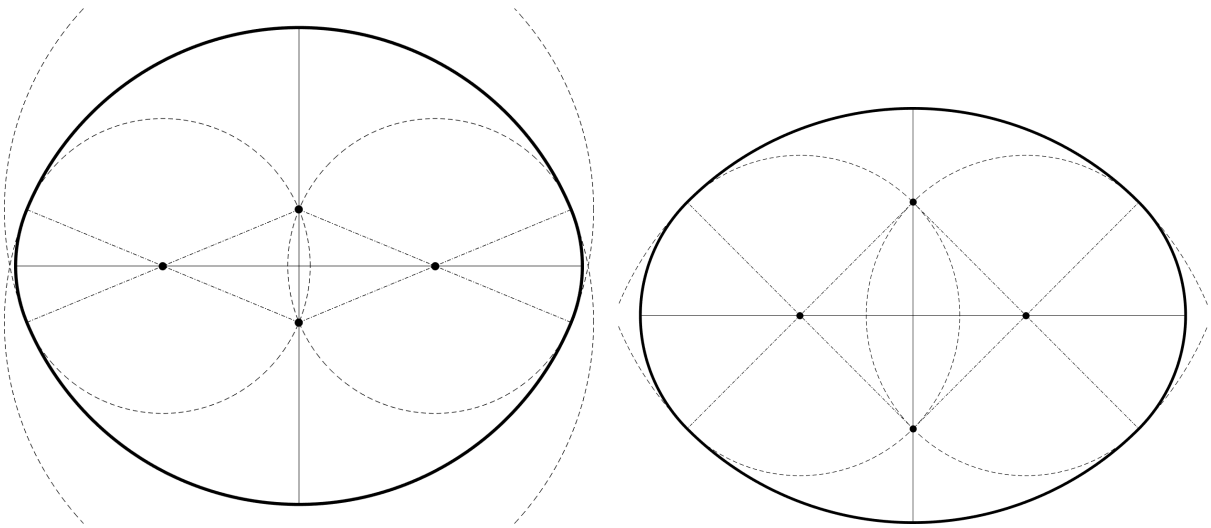

Figure 1: Left: Pseudo-ellipse $E_{\pi/4,1,2}$. Right: Squared pseudo-ellipse $E_{\pi/2,1,2}$. Pseudo-ellipses are represented with thick lines, their pairs of symmetry lines with thin continuous lines, their centers O_j with solid dots, the circumferences of radii r_j centered at O_j with dashed thin lines, their angular ranges $[a_j, b_j]$ with dash-dotted thin lines, and their nodes are the intersections of the thick and dash-dotted thin lines.

Pseudo-ellipses are the simplest examples. We may define them as the circular 4-gons with a $\mathbb{Z}_2 \times \mathbb{Z}_2$-symmetry. They form, modulo translations and rotations, a three-parameter family. The radii and central angles of any pseudo ellipse have the form

$$r_1 = r_3 = r, \qquad r_2 = r_4 = R, \qquad \delta_1 = \delta_3 = \alpha, \qquad \delta_2 = \delta_4 = \pi - \alpha,$$

for some free parameters $\alpha \in (0, \pi)$, and $r, R > 0$. We will assume that $0 < r < R$ for convenience. We will denote by $E_{\alpha,r,R}$ the corresponding pseudo-ellipse. Given any pseudo-ellipse $E_{\alpha,r,R}$, its centers form a *rhombus* (4 equal sides) and its nodes form a *rectangle* (4 equal angles). If $\alpha = \pi/2$, then $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \pi/2$ and we say that $E_{\pi/2,r,R}$ is a *squared pseudo-ellipse*. The term *squared* comes from the fact that the centers of such pseudo-ellipses form a square. See Figure 1. The nodes of a squared pseudo-ellipse still form a rectangle, not a square. On the contrary, the celebrated Benettin-Strelcyn ovals, whose billiard dynamics was numerically studied in [6, 34, 43], are pseudo-ellipses whose nodes form a square, but whose centers only form a rhombus. Later on, the extent of chaos in billiards associated to general pseudo-ellipses was numerically studied in [23].

Another celebrated example of a circular 4-gon is *Moss's egg* [21, Section 1.1], whose radii and central angles have the form

$$r_1 = r, \quad r_2 = 2r = r_4, \quad r_3 = (2 - \sqrt{2})r, \quad \delta_1 = \pi, \quad \delta_2 = \pi/4 = \delta_4, \quad \delta_3 = \pi/2,$$

for some free parameter $r > 0$, called the radius of the egg. All Moss's eggs are congruent modulo similarities. They have a $\mathbb{Z}_2$-symmetry, so their nodes form an *isosceles trapezoid* (2 pairs of consecutive equal angles) and its centers form a *kite* (2 pairs of adjacent equal-length sides). In fact, this kite is somewhat degenerate since it is, in fact, a triangle. See Figure 2. Billiards in a 2-parameter family of circular 4-gons with $\mathbb{Z}_2$-symmetry, but not containing Moss's egg, were considered in [4]. The heuristic analysis of sliding trajectories contained in Section 4.5 of that paper is closely related to our study.

Next, we describe a way to construct some circular 6-gons. Fix a triangle $\triangle ABC$ with vertexes A, B and C ordered in the *clockwise* direction. Let α, β and γ be its internal angles. Let a, b and c be the lengths of its sides, following the standard convention. That is, a refers to the side opposed to vertex A and so forth. Then we look for circular 6-gons with centers $O_1 = O_4 = A$, $O_2 = O_5 = B$, $O_3 = O_6 = C$ and central angles $\delta_1 = \delta_4 = \alpha$, $\delta_2 = \delta_5 = \beta$ and $\delta_3 = \delta_6 = \gamma$. In this setting, all radii are determined by the choice of the first one. Namely, we can take

$$r_1 = r, \quad r_2 = r + c, \quad r_3 = r + c - a, \quad r_4 = r + c - a + b, \quad r_5 = r + b - a, \quad r_6 = r + b,$$

for any $r > \max\{0, a - c, a - b\}$. Therefore, we obtain a one-parameter family of parallel circular 6-gons, parameterised by the first radius $r_1 = r$. See Figure 2 for a non-symmetric example with $A = (3, -1)$, $B = (-1, -1)$, $C = (0, 1)$ and $r = 1$.

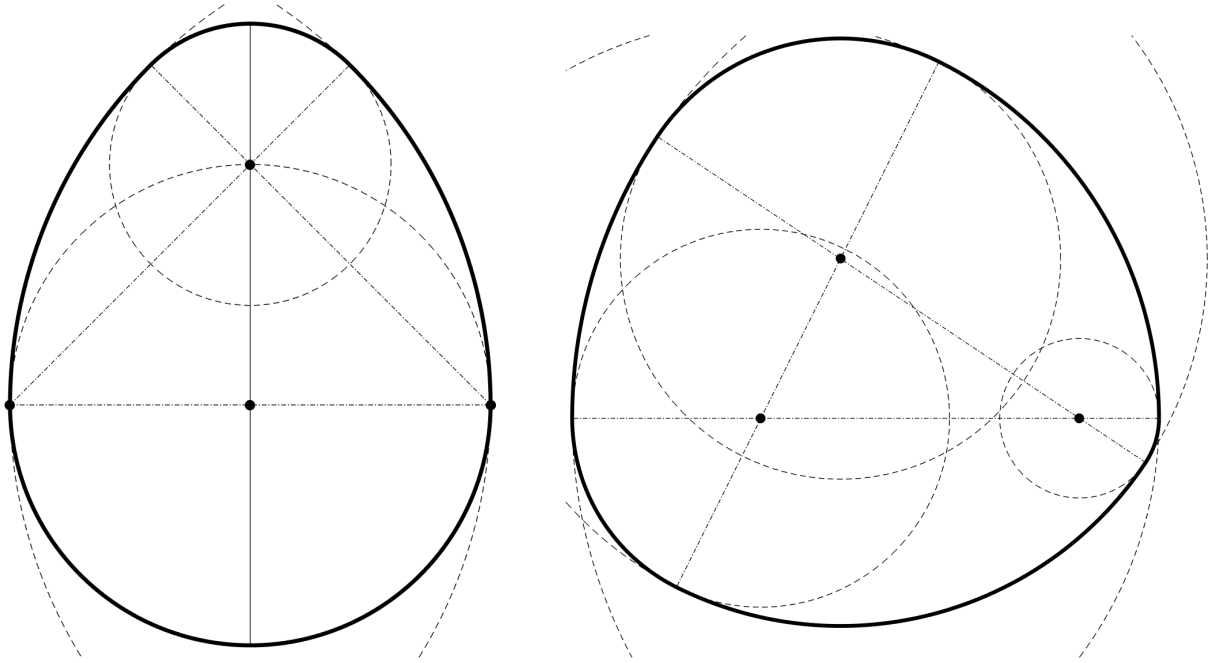

Figure 2: Left: Moss's egg. Right: A nonsymmetric circular 6-gon with centers $O_1 = O_4 = (3, -1)$, $O_2 = O_5 = (-1, -1)$ and $O_3 = O_6 = (0, 1)$, which form a triangle. Circular polygons are represented with thick lines, the symmetry line of Moss's egg with a thin continuous line, their centers O_j with solid dots, the circumferences of radii r_j centered at O_j with dashed thin lines, their angular ranges $[a_j, b_j]$ with dash-dotted thin lines, and their nodes are the intersections of the thick and dash-dotted thin lines.

One can draw circular polygons with many arcs by applying similar constructions, but that challenge is beyond the scope of this paper. The interested reader can look for inspiration in the nice construction of elliptic flowers due to Bunimovich [13].

To end this section, we emphasise that all our theorems are general. They can be applied to any circular polygon. Thus, we do not need to deal with concrete circular polygons.

3 The 'stretching along the paths' method

In this short section, we present the main ideas of the *stretching along the paths* method developed by Papini and Zanolin [48, 49], and extended by Pireddu and Zanolin [52, 53, 51]. The reader acquainted with the method can take note of the notation introduced in Definition 7 and skip the rest of this section.

This method is a technical tool to establish the existence of *topological chaos*; that is, chaotic dynamics in continuous maps. We present a simplified version of the method because we work in the two-dimensional annulus $\mathcal{M} = \mathbb{T} \times [0, \pi]$ and our maps are *homeomorphisms* on $\mathcal{M}$. We also change some terminology because our maps stretch along *vertical* paths, instead of *horizontal* paths. The reader interested in more general statements about higher dimensions, finding fixed and periodic points in smaller compact sets, the study of crossing numbers, non-invertible maps, and maps not defined in the whole space $\mathcal{M}$, is referred to the original references.

Let $\mathcal{M} = \mathbb{T} \times [0, \pi]$. By a *continuum* we mean a compact connected subset of $\mathcal{M}$. *Paths* and *arcs* are the continuous and the homeomorphic images of the unit interval $[0, 1]$, respectively. Most definitions below are expressed in terms of paths, but we could also use arcs or continua, see [48, Table 3.1]. *Cells* are the homeomorphic image of the unit square $[0, 1]^2$ so they are simply connected and compact. The Jordan-Shoenflies theorem implies that any simply connected compact subset of $\mathcal{M}$ bounded by a Jordan curve is a cell. Figure 3 provides a visual guide for the following definitions.

Definition 5. An *oriented cell* $\widetilde{\mathcal{Q}}$ is a cell $\mathcal{Q} \subset \mathcal{M}$ where we have chosen four different points $\widetilde{\mathcal{Q}}_{\mathrm{bl}}$ (base-left), $\widetilde{\mathcal{Q}}_{\mathrm{br}}$ (base-right), $\widetilde{\mathcal{Q}}_{\mathrm{tr}}$ (top-right) and $\widetilde{\mathcal{Q}}_{\mathrm{tl}}$ (top-left) over the boundary $\partial\mathcal{Q}$ in a counter-clockwise order. The *base side* of $\widetilde{\mathcal{Q}}$ is the arc $\widetilde{\mathcal{Q}}_{\mathrm{b}} \subset \partial\mathcal{Q}$ that goes from $\widetilde{\mathcal{Q}}_{\mathrm{bl}}$ to $\widetilde{\mathcal{Q}}_{\mathrm{br}}$ in the counter-clockwise orientation. Similarly, $\widetilde{\mathcal{Q}}_{\mathrm{l}}$, $\widetilde{\mathcal{Q}}_{\mathrm{r}}$ and $\widetilde{\mathcal{Q}}_{\mathrm{t}}$ are the *left*, *right* and *top sides* of $\widetilde{\mathcal{Q}}$. Finally,

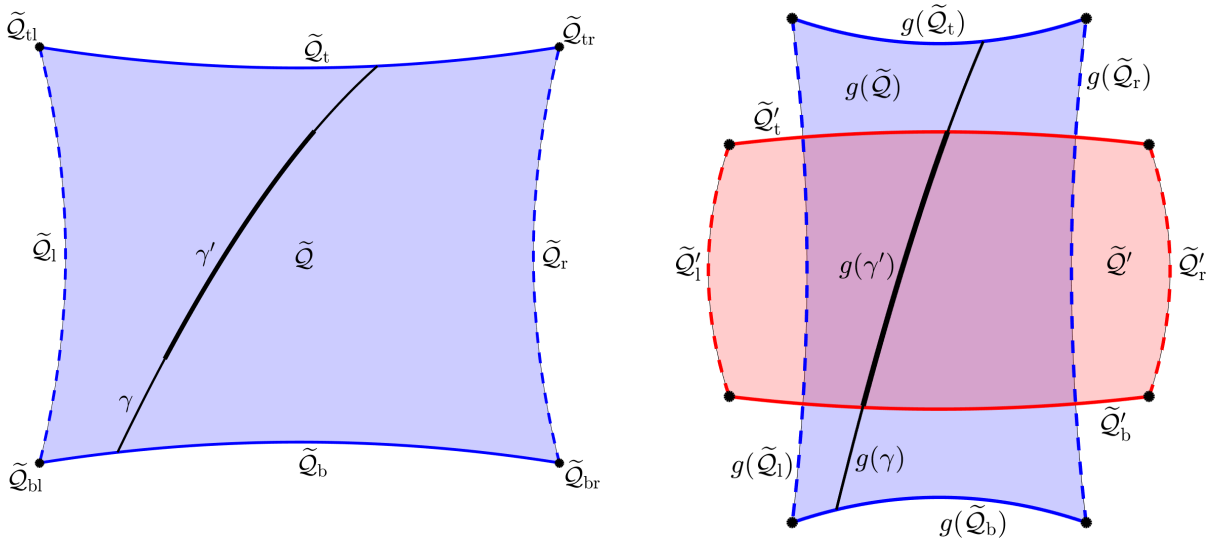

Figure 3: Left: An oriented cell $\widetilde{\mathcal{Q}}$ (in blue) with a vertical path γ (in black). Right: A homeomorphism g stretches $\widetilde{\mathcal{Q}}$ to a second oriented cell $\widetilde{\mathcal{Q}}'$ (in red) along vertical paths. The horizontal and vertical sides of both cells are represented with continuous and dashed lines, respectively. The vertical path γ and its subpath γ' (and their corresponding images in the picture on the right) are represented with thin and thick lines, respectively.

$\widetilde{\mathcal{Q}}_{\rm h} = \widetilde{\mathcal{Q}}_{\rm b} \cup \widetilde{\mathcal{Q}}_{\rm t}$ and $\widetilde{\mathcal{Q}}_{\rm v} = \widetilde{\mathcal{Q}}_{\rm l} \cup \widetilde{\mathcal{Q}}_{\rm r}$ are the *horizontal* and *vertical sides* of $\widetilde{\mathcal{Q}}$.

All our cells will have line segments as vertical sides, some being even quadrilaterals.

Definition 6. Let $\widetilde{\mathcal{Q}}$ be an oriented cell. A path $\gamma : [a, b] \to Q$ is *vertical* (respectively, *horizontal*) in $\widetilde{\mathcal{Q}}$ when it connects the two horizontal (respectively, vertical) sides of $\widetilde{\mathcal{Q}}$ and $\gamma(t) \notin \widetilde{\mathcal{Q}}_{\rm h}$ (respectively, $\gamma(t) \notin \widetilde{\mathcal{Q}}_{\rm v}$) for all $t \in (a, b)$. We say that an oriented cell $\widetilde{\mathcal{K}}$ is a *horizontal slab* in $\widetilde{\mathcal{Q}}$ and write

$$\widetilde{\mathcal{K}} \subset_{\rm h} \widetilde{\mathcal{Q}}$$

when $\mathcal{K} \subset \mathcal{Q}$ and, either $\widetilde{\mathcal{K}}_{\rm l} \subset \widetilde{\mathcal{Q}}_{\rm l}$ and $\widetilde{\mathcal{K}}_{\rm r} \subset \widetilde{\mathcal{Q}}_{\rm r}$, or $\widetilde{\mathcal{K}}_{\rm l} \subset \widetilde{\mathcal{Q}}_{\rm r}$ and $\widetilde{\mathcal{K}}_{\rm r} \subset \widetilde{\mathcal{Q}}_{\rm l}$. If, in addition, $\mathcal{K} \cap \widetilde{\mathcal{Q}}_{\rm h} = \emptyset$, then we say that $\widetilde{\mathcal{K}}$ is a *strict horizontal slab* in $\widetilde{\mathcal{Q}}$ and write

$$\widetilde{\mathcal{K}} \subsetneq_{\rm h} \widetilde{\mathcal{Q}}.$$

Vertical slabs can be defined analogously. Note that $\widetilde{\mathcal{K}} \subsetneq_{\rm h} \widetilde{\mathcal{Q}}$ is a much stronger condition than $\widetilde{\mathcal{K}} \subset_{\rm h} \widetilde{\mathcal{Q}}$ and $\mathcal{K} \subsetneq \mathcal{Q}$.

Definition 7. Let $g : \mathcal{M} \to \mathcal{M}$ be a homeomorphism. Let $\widetilde{\mathcal{Q}}$ and $\widetilde{\mathcal{Q}}'$ be oriented cells in $\mathcal{M}$. We say that *g stretches $\widetilde{\mathcal{Q}}$ to $\widetilde{\mathcal{Q}}'$ along vertical paths* and write

$$g : \widetilde{\mathcal{Q}} \rightsquigarrow \widetilde{\mathcal{Q}}'$$

when every path $\gamma : [a, b] \to \mathcal{Q}$ that is vertical in $\widetilde{\mathcal{Q}}$ contains a *subpath* $\gamma' = \gamma_{|[s,t]}$ for some $a \le s < t \le b$ such that the image path $g \circ \gamma' : [s, t] \to \mathcal{Q}'$ is vertical in $\widetilde{\mathcal{Q}}'$.

This stretching condition does not imply that $g(\mathcal{Q}) \subset \mathcal{Q}'$. In fact, we see $\mathcal{Q}'$ as a 'target set' that we want to 'visit', and not as a codomain. If $\gamma : [a, b] \to \mathcal{M}$ is a path, we also use the notation γ to mean the set $\gamma([a, b]) \subset \mathcal{M}$. This allows us to state the stretching condition more succinctly. Namely, we ask that every path γ vertical in $\widetilde{\mathcal{Q}}$ contains a subpath $\gamma' \subset \gamma$ such that the image path $g(\gamma')$ is vertical in $\widetilde{\mathcal{Q}}'$.

Definition 8. Let $f : \mathcal{M} \to \mathcal{M}$ be a homeomorphism. Let $(\mathcal{Q}_i; n_i)_{i \in I}$ be a two-sided sequence: $I \in \mathbb{Z}$, one-sided sequence: $I \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, p-periodic sequence: $I = \mathbb{Z}/p\mathbb{Z}$, or finite sequence $I = \{0, 1, \dots, k\}$ with $\mathcal{Q}_i \subset \mathcal{M}$ and $n_i \in \mathbb{N}$. Let $x \in \mathcal{Q}_0$. We say that the point x *f-realises* the sequence $(\mathcal{Q}_i; n_i)_{i \in I}$ when

$$f^{-(n_{-1}+\cdots+n_{-i})}(x) \in \mathcal{Q}_{-i}, \qquad f^{n_0+\cdots+n_{i-1}}(x) \in \mathcal{Q}_i, \qquad \forall i \ge 1.$$

Clearly, condition $f^{-(n_{-1}+\cdots+n_{-i})}(x) \in \mathcal{Q}_{-i}$ does not apply in the case of one-sided or finite sequences. A subset of $\mathcal{Q}_0$ *f-realises* the sequence $(\mathcal{Q}_i; n_i)_{i \in I}$ when all its points do so.

The subsets $\mathcal{Q}_i$ in this definition do not have to be cells, but that is the case considered in the following powerful 3-in-1 theorem about the existence of points and paths of the phase space $\mathcal{M}$ that f-realise certain two-sided, one-sided, and periodic sequences of oriented cells.

Theorem 9 (Papini & Zanolin [49]). *Let $f : \mathcal{M} \to \mathcal{M}$ be a homeomorphism. Let $(\widetilde{\mathcal{Q}}_i; n_i)_{i\in I}$ be a two-sided, one-sided or p-periodic sequence where $\widetilde{\mathcal{Q}}_i$ are oriented cells with $\mathcal{Q}_i \subset \mathcal{M}$ and $n_i \in \mathbb{N}$. If*

$$f^{n_i} : \widetilde{\mathcal{Q}}_i \rightsquigarrow \widetilde{\mathcal{Q}}_{i+1}, \qquad \forall i,$$

then the following statements hold.

- **(T)** *If $I = \mathbb{Z}$, there is a point $x \in \mathcal{Q}_0$ that f-realises the two-sided sequence $(\mathcal{Q}_i; n_i)_{i\in\mathbb{Z}}$.*
- **(O)** *If $I = \mathbb{N}_0$, there is a path γ horizontal in $\widetilde{\mathcal{Q}}_0$ that f-realises the sequence $(\mathcal{Q}_i; n_i)_{i\geq 0}$.*
- **(P)** *If $(\widetilde{\mathcal{Q}}_{i+p}; n_{i+p}) = (\widetilde{\mathcal{Q}}_i; n_i)$ for all $i \in \mathbb{Z}$ and $n = n_0 + \cdots + n_{p-1}$, there is a point $x \in \mathcal{Q}_0$ such that $f^n(x) = x$ and x f-realises the p-periodic sequence $(\mathcal{Q}_i; n_i)_{i\in\mathbb{Z}/p\mathbb{Z}}$.*

Remark 10. We believe that the following finite version (**F**) also holds: "If $I = \{0, \ldots, k\}$, there is a horizontal slab $\widetilde{\mathcal{K}} \subset_{\mathrm{h}} \widetilde{\mathcal{Q}}_0$ such that $\mathcal{K}$ f-realises the finite sequence $(\mathcal{Q}_i; n_i)_{i=0,\ldots,k}$.", but we have not found such statement in the literature. Therefore, we will not use it.

We refer to Theorem 2.2 in [49] for a more general statement which deals with sequences of maps that are not some power iterates of a single map. Version (**T**) of Theorem 9 is the key tool to obtain orbits that follow prescribed itineraries, so that we can establish the existence of topological chaos and we can construct a suitable symbolic dynamics in Section 6. We will use version (**O**) of Theorem 9 to prove the existence of 'paths' of generic sliding billiard trajectories that approach the boundary asymptotically with optimal uniform speed in Section 7. Finally, we will establish several lower bounds on the number of periodic billiard trajectories from version (**P**) of Theorem 9 in Section 8.

4 The fundamental lemma for circular polygons

In this section, we define the billiard map, and describe in Lemma 11 the sliding dynamics in circular polygons. That is, the dynamics when the angle of reflection θ is small. See Definition 13 for the precise formulation. We then introduce *fundamental quadrilaterals*, which will later serve as symbol sets for symbolic dynamics. In Lemma 17 we compute the extreme points of the fundamental quadrilaterals, as well as those of their iterates after crossing the singularities between two consecutive circular arcs. With these estimates on hand, we finally state and prove the *fundamental lemma* (Lemma 18, as well as an important consequence, Corollary 19), which describes how orbits of the fundamental quadrilaterals visit other fundamental quadrilaterals.

We begin with the definition of the billiard map. Recall that the phase space of the billiard map is $\mathcal{M} = \mathbb{T} \times [0, \pi]$, and let $(\varphi, \theta) \in \operatorname{Int} \mathcal{M}$. Write $z = z(\varphi)$, where z is the polar parametrisation of the circular polygon Γ introduced in Definition 3, and $v = R_\theta z'(\varphi)$, where R_θ is the standard 2×2 counter-clockwise rotation matrix by an angle θ. The straight line $L = L(\varphi, \theta)$ passing through z in the direction v has exactly two points of intersection with Γ since $\theta \in (0, \pi)$. One of these is z; denote by $\bar{z}$ the other. Then there is a unique $\bar{\varphi} \in \mathbb{T}$ such that $\bar{z} = z(\bar{\varphi})$. Denote by $\bar{\theta}$ the angle between L and $z'(\bar{\varphi})$ in the counter-clockwise direction. The *billiard map* $f : \operatorname{Int} \mathcal{M} \to \operatorname{Int} \mathcal{M}$ is defined by $f(\varphi, \theta) = (\bar{\varphi}, \bar{\theta})$, see Figure 4.

Note that f is continuous since Γ is C^1 and strictly convex. The billiard map can be extended continuously to $\partial\mathcal{M}$ by setting $f(\varphi, 0) = (\varphi, 0)$ and $f(\varphi, \pi) = (\varphi, \pi)$ for each $\varphi \in \mathbb{T}$. The billiard map $f : \mathcal{M} \to \mathcal{M}$ is a homeomorphism; indeed, the map $f^{-1} = I \circ f \circ I$ is a continuous inverse where the involution $I : \mathcal{M} \to \mathcal{M}$ is defined by $I(\varphi, \theta) = (\varphi, \pi - \theta)$.

A key geometric property of the billiard dynamics in the case of impacts in consecutive arcs was presented in [35]. Later on, a more detailed description was given in [4]. Both results follow from trigonometric arguments. The following lemma summarises these properties, thus giving a clear picture of the billiard dynamics near $\partial\mathcal{M}$.

Lemma 11. *The billiard map $f : \mathcal{M} \to \mathcal{M}$ satisfies the following properties.*

- *(a) If $a_j \leq \varphi \leq \varphi + 2\theta \leq b_j$, then $(\bar{\varphi}, \bar{\theta}) = f(\varphi, \theta) = (\varphi + 2\theta, \theta)$.*

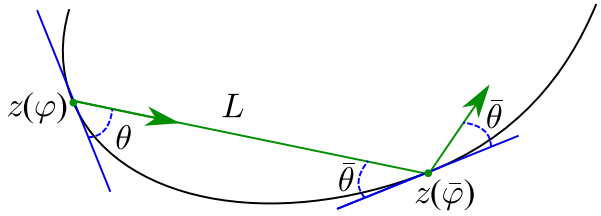

Figure 4: Definition of the billiard map $f(\varphi,\theta)=(\bar\varphi,\bar\theta)$.

(b) Let $g(\theta;\mu) = \operatorname{acos}\big((1-\mu^2)+\mu^2\cos\theta\big)$ *and* $\mu_j = \sqrt{r_j/r_{j+1}} \neq 1$. *If* $0<\theta\le\delta_j$ *and* $\bar\theta = g(\theta;\mu_j)\le\delta_{j+1}$, *then*

$$f(b_j-\theta,\theta)=(a_{j+1}+\bar\theta,\bar\theta) \qquad \textit{and} \qquad \begin{cases}\bar\theta<\mu_j\theta, & \textit{when } \mu_j<1,\\ \bar\theta>\mu_j\theta, & \textit{when } \mu_j>1.\end{cases} \tag{5}$$

(c) Given any $\epsilon>0$ *there exists* $\psi=\psi(\epsilon)>0$ *such that*

$$\left.\begin{array}{l} f(\varphi,\theta)=(\bar\varphi,\bar\theta) \textit{ with } 0<\theta\le\psi\\ \textit{and } a_j\le\varphi\le a_{j+1}\le\bar\varphi\le a_{j+2}\end{array}\right\} \Rightarrow \begin{cases}\bar\theta>(\mu_j-\epsilon)\theta, & \textit{when } \mu_j<1,\\ \bar\theta<(\mu_j+\epsilon)\theta, & \textit{when } \mu_j>1.\end{cases}$$

Proof. (a) If $a_j\le\varphi\le\varphi+2\theta\le b_j$, then $z(\varphi),z(\varphi+2\theta)\in\Gamma_j$, so f behaves as a circular billiard map, in which case is well-known that $f(\varphi,\theta)=(\varphi+2\theta,\theta)$.

(b) Set $\varphi=b_j-\theta$ and $(\bar\varphi,\bar\theta)=f(\varphi,\theta)$. Condition $0<\theta\le\delta_j$ implies that $z(\varphi)\in\Gamma_j$. Identity $\varphi+\theta=b_j$ implies that lines $L=L(\varphi,\theta)$ and N_j are perpendicular, where N_j is the normal to Γ at $z(b_j)$. If, in addition, $z(\bar\varphi)\in\Gamma_{j+1}$, then Hubacher proved (5) in [35, page 486]. Finally, we note that $\bar\theta\le\delta_{j+1}$ implies that $z(\bar\varphi)\in\Gamma_{j+1}$.

(c) Bálint *et al.* [4] proved the following generalisation of Hubacher computation. Set $f(\varphi,\theta)=(\bar\varphi,\bar\theta)$. If $a_j\le\varphi\le a_{j+1}\le\bar\varphi\le a_{j+2}$, so that $z(\varphi)\in\Gamma_j$ and $z(\bar\varphi)\in\Gamma_{j+1}$, then there exist angles $\varphi^+\in[0,2\theta]$ and $\varphi^-\in[0,2\bar\theta]$ such that

$$\varphi=b_j-\varphi^+, \qquad \bar\varphi=a_{j+1}+\varphi^-, \qquad \varphi^++\varphi^-=\theta+\bar\theta,$$

and

$$\bar\theta=\operatorname{acos}\left((1-\mu_j^2)\cos(\theta-\varphi^+)+\mu_j^2\cos\theta\right).$$

Hubacher's computation corresponds to the case $\varphi^+=\theta$ and $\varphi^-=\bar\theta$. We introduce the auxiliary coordinate $s=1-\varphi^+/\theta\in[-1,1]$ and the positive function

$$\Omega_j:[-1,1]\to\mathbb{R}_+, \qquad \Omega_j(s)=\sqrt{\mu_j^2+(1-\mu_j^2)s^2}. \tag{6}$$

Function $\Omega_j(s)$ is even, $\Omega_j(0)=\mu_j$ and $\Omega_j(\pm1)=1$. If $\mu_j<1$, then $\Omega_j(s)$ increases for $s>0$ and decreases for $s<0$, so $\mu_j\le\Omega_j(s)\le1$ for all $s\in[-1,1]$. If $\mu_j>1$, then $\Omega_j(s)$ decreases for $s>0$ and increases for $s<0$, so $1\le\Omega_j(s)\le\mu_j$ for all $s\in[-1,1]$.
A straightforward computation with Taylor expansions shows that

$$1-\bar\theta^2/2+\mathrm{O}(\bar\theta^4)=\cos\bar\theta=(1-\mu_j^2)\cos(s\theta)+\mu_j^2\cos\theta=1-\Omega_j^2(s)\theta^2/2+\mathrm{O}(\theta^4)$$

as $\theta\to0^+$, where the error term $\mathrm{O}(\theta^4)$ is uniform in $s\in[-1,1]$ and $j=1,\dots,k$. Therefore,

$$\bar\theta=\left[\Omega_j(s)+\mathrm{O}(\theta^2)\right]\theta$$

as $\theta\to0^+$, where the error term $\mathrm{O}(\theta^2)$ is uniform in $s\in[-1,1]$ and $j=1,\dots,k$. □

If $a_j \leq \varphi < \varphi + 2\theta = b_j$, then $z(\varphi) \in \Gamma_j$ and $z(\bar{\varphi}) = b_j = a_{j+1} \in \Gamma_j \cap \Gamma_{j+1}$, but part (a) of Lemma 11 still applies, because the tangents to Γ_j and Γ_{j+1} agree at $z(\bar{\varphi})$ by the definition of circular polygon. This fact will be used in Proposition 46 to construct some special periodic *nodal* billiard trajectories in *rational* circular polygons, which are introduced in Definition 45. Similar nodal periodic billiard trajectories were constructed in [10] to answer a question about length spectrum and rigidity.

Lemma 11 and the above observation describe two rather different ways in which the angle θ can vary as a sliding billiard trajectory jumps from one arc to the next. On the one hand, if the trajectory impacts at the corresponding node, there is no change: $\bar{\theta} = \theta$. On the other hand, if the billiard trajectory is perpendicular at the normal line at the corresponding node, we have the largest possible change: $\bar{\theta} < \mu_j \theta$ for $\mu_j < 1$ or $\bar{\theta} > \mu_j \theta$ for $\mu_j > 1$. The great contrast between these two situations is the crucial fact behind the non-existence of caustics near the boundary obtained by Hubacher [35]. It is also the main ingredient to obtain all chaotic properties stated in the introduction.

Next, we introduce the main geometric subsets of the phase space $\mathcal{M} = \mathbb{T} \times [0, \pi]$. All of them are denoted with calligraphic letters.

Definition 12. The *j-singularity segment* is the vertical segment $\mathcal{L}_j = \{a_j\} \times [0, \pi] \subset \mathcal{M}$. Given any $s > 0$, the *$(j, \pm s)$-singularity segments* are the slanted segments

$$\mathcal{L}_j^{-s} = \big\{(\varphi, \theta) \in \mathcal{M} : a_{j-1} \leq \varphi = a_j - 2\theta s\big\}, \quad \mathcal{L}_j^{s} = \big\{(\varphi, \theta) \in \mathcal{M} : \varphi = a_j + 2\theta s \leq a_{j+1}\big\}.$$

The *j-fundamental domain* is the triangular domain

$$\mathcal{D}_j = \{(\varphi, \theta) \in \mathcal{M} : a_j \leq \varphi \leq a_j + 2\theta \leq a_{j+1}\}.$$

Finally, $\mathcal{L} = \bigcup_{j=1}^{k} \left(\mathcal{L}_j \cup \mathcal{L}_j^{1/2} \cup \mathcal{L}_j^{1}\right)$ is the *extended singularity set*.

Note that $\mathcal{L}_j^n \subset f^n(\mathcal{L}_j)$ for all $n \in \mathbb{Z}$, so $\mathcal{L}_j^s$ is a generalisation of the forward and backward iterates under the billiard map of the j-singularity segments when $s \notin \mathbb{Z}$. We will only need the segments $\mathcal{L}_j^s$ for values $s = n$ and $s = n + 1/2$ with $n \in \mathbb{Z}$. The left (respectively, right) side of the triangle $\mathcal{D}_j$ is contained in the vertical segment $\mathcal{L}_j$ (respectively, coincides with the slanted segment $\mathcal{L}_j^1$).

We have used the term 'sliding' in a clumsy way until now. Let us clarify its precise meaning. Let $\Pi_\varphi : \mathcal{M} \to \mathbb{T}$ and $\Pi_\theta : \mathcal{M} \to [0, \pi]$ be the projections $\Pi_\varphi(\varphi, \theta) = \varphi$ and $\Pi_\theta(\varphi, \theta) = \theta$. Let $J : \mathcal{M} \setminus \mathcal{L} \to \mathbb{Z}/k\mathbb{Z}$ be the piece-wise constant map defined by $a_j < \Pi_\varphi(x) < b_j \Rightarrow J(x) = j$. This map is well-defined since $\Pi_\varphi(x) \notin \{a_1, \ldots, a_k\} = \{b_1, \ldots, b_k\}$ when $x \notin \mathcal{L}$.

Definition 13. A billiard orbit is *(counter-clockwise) sliding* when any consecutive impact points are either in the same arc or in consecutive arcs in the counter-clockwise direction. An orbit is *generic* when it avoids the extended singularity set. We denote by $\mathcal{S}_0$ the set of all initial conditions that give rise to generic counter-clockwise sliding orbits. That is,

$$\mathcal{S}_0 = \big\{x \in \mathcal{M} : J(f^{n+1}(x)) - J(f^n(x)) \in \{0, 1\} \text{ and } f^n(x) \notin \mathcal{L} \text{ for all } n \in \mathbb{Z}\big\}.$$

The *(counter-clockwise) generic sliding set* $\mathcal{S}_0$ is f-invariant. The term *glancing* —see, for instance, [45]— is also used in the literature, but sliding is the most widespread term. A consequence of part (a) of Lemma 11 is that any generic sliding billiard orbit has exactly one point in $\operatorname{Int} \mathcal{D}_j$ on each turn around Γ. This fact establishes the *fundamental* character of $\mathcal{D}_j$. Following the notation used in the introduction, $\mathcal{S}_\pi$ is the clockwise generic sliding set, but we are not going to deal with it.

Remark 14. If $x \in \mathcal{M}$ is a point such that $x_i = (\varphi_i, \theta_i) = f^i(x) \in \mathcal{L}$ for some $i \in \mathbb{Z}$, then its billiard trajectory $\big(z_n = z(\Pi_\varphi(f^n(x)))\big)_{n \in \mathbb{Z}}$ has some impact point $z_m \in \Gamma_\star$, where $\Gamma_\star$ is the set of nodes (1), or has two consecutive impact points $z_m \in \Gamma_j$ and $z_{m+1} \in \Gamma_{j+1}$ such that the segment from z_m to z_{m+1} is perpendicular to the normal to Γ at the node $\Gamma_j \cap \Gamma_{j+1}$.

Lemma 15. *Let $s, t \geq 0$ and j such that $s + t \geq \delta_j / 2\pi$. Then $\mathcal{L}_j^s \cap \mathcal{L}_{j+1}^{-t} \neq \emptyset$ and*

$$\Pi_\varphi\left(\mathcal{L}_j^s \cap \mathcal{L}_{j+1}^{-t}\right) = a_j + \frac{s\delta_j}{s+t}, \qquad \Pi_\theta\left(\mathcal{L}_j^s \cap \mathcal{L}_{j+1}^{-t}\right) = \frac{\delta_j}{2s + 2t}.$$

Proof. By definition, $(\varphi, \theta) \in \mathcal{L}_j^s \cap \mathcal{L}_{j+1}^{-t} \Leftrightarrow a_j \leq a_j + 2\theta s = \varphi = a_{j+1} - 2\theta t \leq a_{j+1}$. Identity $a_j + 2\theta s = a_{j+1} - 2\theta t$ implies that $2\theta = \delta_j/(s+t)$. Then inequality $s + t \geq \delta_j/2\pi$ implies that $\theta \leq \pi$. Finally, $a_j + 2\theta s \leq a_{j+1}$ and $a_{j+1} - 2\theta t \geq a_j$ because $s/(s+t), t/(s+t) \leq 1$. □

This lemma implies that segments $\mathcal{L}_{j+1}^{-n+1}$ and $\mathcal{L}_{j+1}^{-n}$ intersect segments $\mathcal{L}_j$ and $\mathcal{L}_j^1$ for any integer $n \geq 2 > 1 + \delta_j/2\pi$, so the following definition makes sense. See Figure 5.

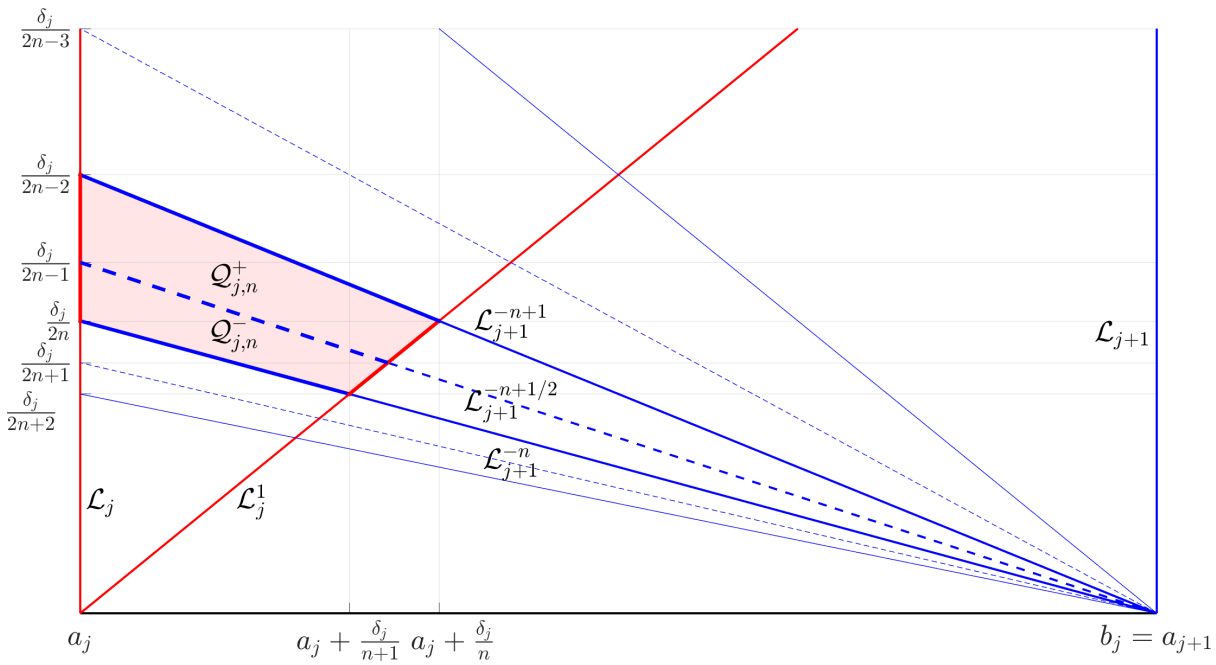

Figure 5: The fundamental quadrilateral $\mathcal{Q}_{j,n} = \mathcal{Q}^-_{j,n} \cup \mathcal{Q}^+_{j,n}$. Its horizontal (base and top) and vertical (left and right) sides are displayed in blue and red, respectively. Accordingly, singularity segments $\mathcal{L}^s_j$ and $\mathcal{L}^{-t}_{j+1}$ are displayed in red and blue, respectively. Moreover, they are displayed with continuous and dashed lines when $s, t \in \mathbb{N}$ and $t \in \mathbb{N} + \frac{1}{2}$, respectively. This is a quantitative representation computed for $\delta_j = \pi/2$ and $n = 3$. The images $f^n(\mathcal{Q}^\pm_{j,n})$ are displayed in Figure 6.

Definition 16. Let n be an integer such that $n \geq 2$. The (j,n)-*fundamental quadrilateral* is the oriented cell $\widetilde{\mathcal{Q}}_{j,n} \subset \mathcal{M}$ bounded by $\mathcal{L}_j$ (left side), $\mathcal{L}^{-n}_{j+1}$ (base side), $\mathcal{L}^1_j$ (right side) and $\mathcal{L}^{-n+1}_{j+1}$ (top side). We split $\mathcal{Q}_{j,n}$ in two by means of the segment $\mathcal{L}^{-n+1/2}_j$, which gives rise to two smaller oriented cells: $\widetilde{\mathcal{Q}}^-_{j,n}$ (the lower one) and $\widetilde{\mathcal{Q}}^+_{j,n}$ (the upper one), whose left and right sides are still contained in $\mathcal{L}_j$ and $\mathcal{L}^1_j$, respectively. We say that $\tilde{\mathcal{Q}}^\pm_{j,n}$ is the $(\pm, j, n)$-*fundamental quadrilateral*.

In order to find sufficient conditions for $f^n : \widetilde{\mathcal{Q}}^\varsigma_{j,n} \rightsquigarrow \widetilde{\mathcal{Q}}^{\varsigma'}_{j+1,n'}$, we need the extreme values of $\Pi_\theta(f^n(x))$ when x moves on the horizontal sides of $\widetilde{\mathcal{Q}}^\varsigma_{j,n}$ and the extreme values of $\Pi_\theta(x)$ when $x \in \mathcal{Q}_{j+1,n'} = \mathcal{Q}^-_{j+1,n'} \cup \mathcal{Q}^+_{j+1,n'}$. These extreme values are defined and estimated in the lemma below. Those estimates are used in the proof of Lemma 18.

Lemma 17. *Fix any j. With the above notations, if $\chi_j \geq 2$ is a large enough integer, then the following properties hold for all $n \geq \chi_j$.*

(a) $\nu_{j,n} := \min_{x \in \mathcal{Q}_{j,n}} \Pi_\theta(x) = \delta_j/(2n+2)$ *and* $\omega_{j,n} := \max_{x \in \mathcal{Q}_{j,n}} \Pi_\theta(x) = \delta_j/(2n-2)$.

(b) If $\nu^s_{j,n} := \min_{x \in \mathcal{Q}_{j,n} \cap \mathcal{L}^{-n+s}_{j+1}} \Pi_\theta(f^n(x))$ *and* $\omega^s_{j,n} := \max_{x \in \mathcal{Q}_{j,n} \cap \mathcal{L}^{-n+s}_{j+1}} \Pi_\theta(f^n(x))$, *then*

i) $\nu^0_{j,n} = \delta_j/(2n+2)$, $\nu^1_{j,n} = \delta_j/2n = \omega^0_{j,n}$ *and* $\omega^1_{j,n} = \delta_j/(2n-2)$;

ii) $\omega^{1/2}_{j,n} < \mu_j \delta_j/(2n-1)$ *when* $\mu_j < 1$; *and*

iii) $\nu^{1/2}_{j,n} > \mu_j \delta_j/(2n+1)$ *when* $\mu_j > 1$.

Proof. The fundamental domain $\mathcal{Q}_{j,n}$ is only well-defined for $n \geq 2$. The reader must keep in mind Lemmas 11 and 15. See Figure 5 for a visual guide.

(a) The minimum and maximum values are attained at the intersections $\mathcal{L}^1_j \cap \mathcal{L}^{-n}_{j+1}$ and $\mathcal{L}_j \cap \mathcal{L}^{-n+1}_{j+1}$, respectively.

(b) i) If $x \in \mathcal{Q}_{j,n} \cap \mathcal{L}^{-n}_{j+1}$ or $x \in \mathcal{Q}_{j,n} \cap \mathcal{L}^{-n+1}_{j+1}$, then $\Pi_\theta(f^n(x)) = \Pi_\theta(x)$ by part (a) of Lemma 11. Therefore, the four extreme values $\nu^0_{j,n}$, $\nu^1_{j,n}$, $\omega^0_{j,n}$ and $\omega^1_{j,n}$ are attained at the four intersections $\mathcal{L}^1_j \cap \mathcal{L}^{-n}_{j+1}$, $\mathcal{L}^1_j \cap \mathcal{L}^{-n+1}_{j+1}$, $\mathcal{L}_j \cap \mathcal{L}^{-n}_{j+1}$ and $\mathcal{L}_j \cap \mathcal{L}^{-n+1}_{j+1}$, respectively.

ii) First, the value $\max_{x\in\mathcal{Q}_{j,n}\cap\mathcal{L}_{j+1}^{-n+1/2}}\Pi_\theta(x)$ is attained at $\mathcal{L}_j\cap\mathcal{L}_{j+1}^{-n+1/2}$. Second, if $x\in\mathcal{Q}_{j,n}\cap\mathcal{L}_{j+1}^{-n+1/2}$ and $\mu_j<1$, then $\Pi_\theta(f^n(x))<\mu_j\Pi_\theta(x)$ by part (b) of Lemma 11. We need hypotheses $0<\theta\le\delta_j$ and $\bar\theta=g(\theta;\mu_j)\le\delta_{j+1}$ to apply Lemma 11. In order to guarantee them, it suffices to take $n\ge\chi_j$ with

$$\chi_j\ge 1+\lceil\mu_j\delta_j/2\delta_{j+1}\rceil, \tag{7}$$

since then $\chi_j\ge 2$ and $2\chi_j-2\ge\mu_j\delta_j/\delta_{j+1}$, so $\theta\le\omega_{j,n}\le\delta_j/(2\chi_j-2)\le\delta_j/2<\delta_j$ and $\bar\theta<\mu_j\theta\le\mu_j\delta_j/(2\chi_j-2)\le\delta_{j+1}$. Here $\lceil\cdot\rceil$ denotes the *ceiling* function.

iii) First, the value $\min_{x\in\mathcal{Q}_{j,n}\cap\mathcal{L}_{j+1}^{-n+1/2}}\Pi_\theta(x)$ is attained at $\mathcal{L}_j^1\cap\mathcal{L}_{j+1}^{-n+1/2}$. Second, if $x\in\mathcal{Q}_{j,n}\cap\mathcal{L}_{j+1}^{-n+1/2}$ and $\mu_j>1$, then $\Pi_\theta(f^n(x))>\mu_j\Pi_\theta(x)$ by part (b) of Lemma 11. We still need hypotheses $0<\theta\le\delta_j$ and $\bar\theta=g(\theta;\mu_j)\le\delta_{j+1}$ in Lemma 11. In order to guarantee them, it suffices to take $n\ge\chi_j$ for some large enough integer χ_j, since $\lim_{n\to+\infty}\omega_{j,n}=0$ and $\lim_{\theta\to0^+}g(\theta;\mu_j)=0$. □

The following lemma (which we refer to as the *fundamental lemma*) is the key step in constructing generic sliding billiard trajectories that approach the boundary in optimal time, and in constructing symbolic dynamics. It describes which fundamental quadrilaterals in $\mathcal{D}_{j+1}$ we can 'nicely' visit if we start in a given fundamental quadrilateral in $\mathcal{D}_j$. See Figure 6 for a visual guide.

Lemma 18 (Fundamental Lemma)**.** *With the above notations, let*

$$\Upsilon_j=\left\{(n,n')\in\mathbb{N}^2:\alpha_j^-n+\beta_j^-\le n'\le\alpha_j^+n-\beta_j^+,\ n\ge\chi_j,\ n'\ge\chi_{j+1}\right\}, \tag{8}$$

where $\alpha_j^-=\delta_{j+1}/\delta_j\max\{1,\mu_j\}$, $\alpha_j^+=\delta_{j+1}/\delta_j\min\{1,\mu_j\}$ *and* $\beta_j^\pm=\alpha_j^\pm+1$ *for all* j*. Then* $f^n:\widetilde{\mathcal{Q}}_{j,n}^\varsigma\rightsquigarrow\widetilde{\mathcal{Q}}_{j+1,n'}^{\varsigma'}$ *for all* j*,* $(n,n')\in\Upsilon_j$ *and* $\varsigma,\varsigma'\in\{-,+\}$*.*

Proof. We fix an index j such that $\mu_j<1$, so $\alpha_j^-=\delta_{j+1}/\delta_j$ and $\alpha_j^+=\delta_{j+1}/\mu_j\delta_j$. Let $(n,n')\in\Upsilon_j$. We want to show that $f^n:\widetilde{\mathcal{Q}}_{j,n}^\varsigma\rightsquigarrow\widetilde{\mathcal{Q}}_{j+1,n'}^{\varsigma'}$ for any $\varsigma,\varsigma'\in\{-,+\}$.

Let $\gamma^\varsigma:[a,b]\to\mathcal{Q}_{j,n}^\varsigma$ be a vertical path in $\widetilde{Q}_{j,n}^\varsigma$. We assume, without loss of generality, that $\gamma^\varsigma(a)$ and $\gamma^\varsigma(b)$ belong to the base side and top side of $\tilde{\mathcal{Q}}_{j,n}^\varsigma$. Note that $\overline{\mathcal{D}_{j+1}\setminus\mathcal{Q}_{j+1,n'}}$ has a connected component above and other one below $\mathcal{Q}_{j+1,n'}$. By continuity and using that $\mathcal{Q}_{j+1,n'}^{\varsigma'}\subset\mathcal{Q}_{j+1,n'}$ and $f^n(\gamma^\varsigma)\subset\mathcal{D}_{j+1}$, we know that if $f^n(\gamma^\varsigma(a))$ and $f^n(\gamma^\varsigma(b))$ are in different connected components of $\overline{\mathcal{D}_{j+1}\setminus\mathcal{Q}_{j+1,n'}}$, then there is a subpath $\eta^{\varsigma,\varsigma'}\subset\gamma^\varsigma$ such that the image path $f^n(\eta^{\varsigma,\varsigma'})$ is vertical in $\widetilde{\mathcal{Q}}_{j+1,n'}^{\varsigma'}$. Thus, we only have to check that endpoints $f^n(\gamma^\varsigma(a))$ and $f^n(\gamma^\varsigma(b))$ are in different connected components of $\overline{\mathcal{D}_{j+1}\setminus\mathcal{Q}_{j+1,n'}}$ for any path γ^ς vertical in $\widetilde{Q}_{j,n}^\varsigma$.

First, we consider the case $\varsigma=-$, so $\gamma^-(a)\in\mathcal{Q}_{j,n}^-\cap\mathcal{L}_{j+1}^{-n}$ and $\gamma^-(b)\in\mathcal{Q}_{j,n}^-\cap\mathcal{L}_{j+1}^{-n+1/2}$. We deduce from Lemma 17 that

$$\begin{aligned}\Pi_\theta(f^n(\gamma^-(a)))&\ge\min_{x\in\mathcal{Q}_{j,n}^-\cap\mathcal{L}_{j+1}^{-n}}\Pi_\theta(f^n(x))=\min_{x\in\mathcal{Q}_{j,n}\cap\mathcal{L}_{j+1}^{-n}}\Pi_\theta(f^n(x))=\delta_j/(2n+2),\\ \Pi_\theta(f^n(\gamma^-(b)))&\le\max_{x\in\mathcal{Q}_{j,n}^-\cap\mathcal{L}_{j+1}^{-n+1/2}}\Pi_\theta(f^n(x))<\mu_j\delta_j/(2n-1).\end{aligned}$$

The above results and identities $\nu_{j+1,n'}=\delta_{j+1}/(2n'+2)$ and $\omega_{j+1,n'}=\delta_{j+1}/(2n'-2)$ imply that if inequalities

$$\mu_j\delta_j/(2n-1)\le\delta_{j+1}/(2n'+2),\qquad\delta_{j+1}/(2n'-2)\le\delta_j/(2n+2) \tag{9}$$

hold, then $f^n(\gamma^-(a))$ and $f^n(\gamma^-(b))$ are above and below $\mathcal{Q}_{j+1,n'}$, respectively, so they are in different connected components of $\overline{\mathcal{D}_{j+1}\setminus\mathcal{Q}_{j+1,n'}}$ for any path γ^- vertical in $\widetilde{Q}_{j,n}^-$.

Second, we consider the case $\varsigma=+$, so $\gamma^+(a)\in\mathcal{Q}_{j,n}^+\cap\mathcal{L}_{j+1}^{-n+1/2}$ and $\gamma^+(b)\in\mathcal{Q}_{j,n}^+\cap\mathcal{L}_{j+1}^{-n+1}$. Similar arguments show that if inequalities

$$\mu_j\delta_j/(2n-1)\le\delta_{j+1}/(2n'+2),\qquad\delta_{j+1}/(2n'-2)\le\delta_j/2n \tag{10}$$

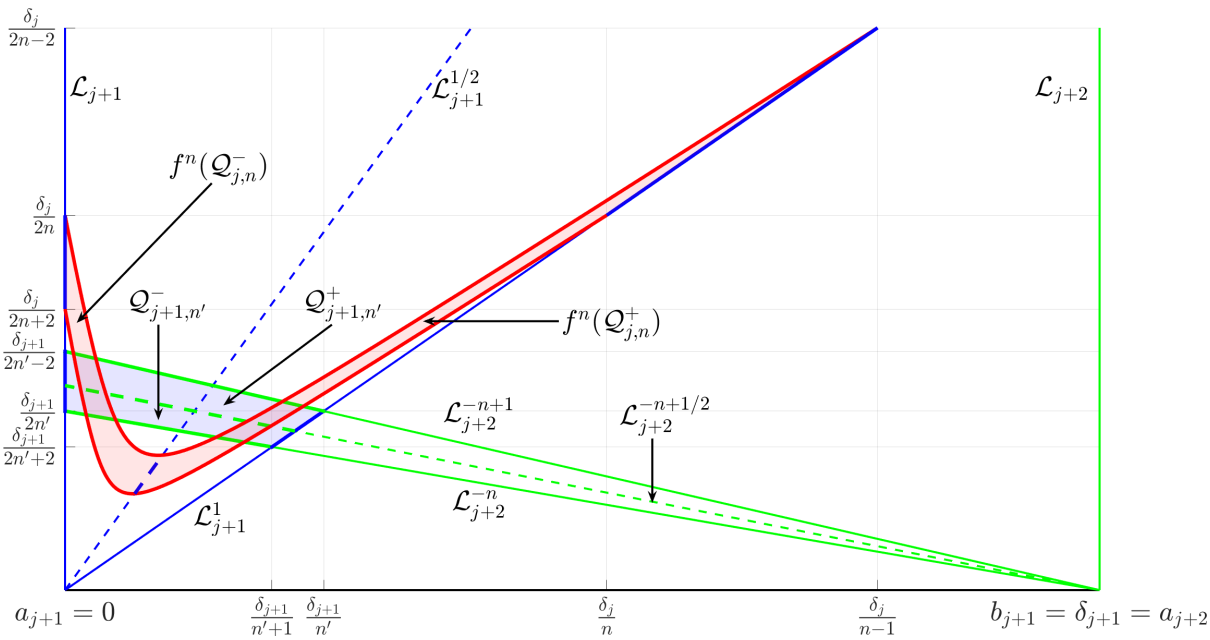

Figure 6: Overlapping of the image under f^n of the fundamental quadrilateral $\mathcal{Q}_{j,n} = \mathcal{Q}_{j,n}^- \cup \mathcal{Q}_{j,n}^+$ displayed in Figure 5 with the 'target' fundamental quadrilateral $\mathcal{Q}_{j+1,n'} = \mathcal{Q}_{j+1,n'}^- \cup \mathcal{Q}_{j+1,n'}^+$ in the case $\mu_j < 1$. Singularity segments $\mathcal{L}_{j+1}^s$ and $\mathcal{L}_{j+2}^{-t}$ are displayed in blue and green, respectively. Besides, they are displayed with continuous and dashed lines when $s, t \in \mathbb{N}$ and $t \in \mathbb{N} + \frac{1}{2}$, respectively. The upper and lower red thick continuous curves are the images of the left and right sides of $\mathcal{Q}_{j,n}$, respectively. We have assumed that $a_{j+1} = 0$ to avoid the overlapping of labels in the horizontal axis. This is a quantitative representation computed for $\delta_j = \pi/2$, $\delta_{j+1} = 1$, $\mu_j = \sqrt{r_j/r_{j+1}} = 0.3$, $n = 3$ and $n' = 4$, showing that $f^n : \widetilde{\mathcal{Q}}_{j,n}^{\varsigma} \rightsquigarrow \widetilde{\mathcal{Q}}_{j+1,n'}^{\varsigma'}$ for all $\varsigma, \varsigma' \in \{-,+\}$. The parabolic-like shape of $f^n(\mathcal{Q}_{j,n})$ was expected, see (6).

hold, then $f^n(\gamma^+(a))$ and $f^n(\gamma^+(b))$ are below and above $\mathcal{Q}_{j+1,n'}$, respectively, so they are in different connected components of $\overline{\mathcal{D}_{j+1} \setminus \mathcal{Q}_{j+1,n'}}$ for any path γ^+ vertical in $\widetilde{Q}_{j,n}^+$.

Finally, after a straightforward algebraic manipulation, we check that inequalities (9) and (10) hold for any $(n, n') \in \Upsilon_j$. This ends the proof for the case $\mu_j < 1$.

The case $\mu_j > 1$ follows from similar arguments. We skip the details. □

No inequality in (8) is strict. However, we need some strict inequalities for a technical reason. Let us explain it. We will use the objects defined above and the fundamental lemma to construct our symbolic dynamics in Section 6. However, if we were to try to construct our symbolic dynamics directly with the symbol sets being the fundamental quadrilaterals $Q_{j,n}^{\varsigma}$, we would run into problems at the boundaries, where neighboring quadrilaterals intersect. To be precise, $\mathcal{Q}_{j,n}^+$ and $\mathcal{Q}_{j,n}^-$ have a common side contained in $\mathcal{L}_j^{-n+1/2}$, whereas $\mathcal{Q}_{j,n}^-$ and $\mathcal{Q}_{j,n+1}^+$ have a common side contained in $\mathcal{L}_j^{-n}$. The following corollary to Lemma 18 solves this problem by establishing the existence of pairwise disjoint strict horizontal slabs $\widetilde{\mathcal{K}}_{j,n}^{\varsigma} \subsetneq_{\mathrm{h}} \widetilde{\mathcal{Q}}_{j,n}^{\varsigma}$ with *exactly* the same stretching properties as the original fundamental quadrilaterals $\widetilde{\mathcal{Q}}_{j,n}^{\varsigma}$. It requires some strict inequalities.

Corollary 19 (Fundamental Corollary)**.** *With the above notations, let*

$$\Xi_j = \left\{(n, n') \in \mathbb{N}^2 : \alpha_j^- n + \beta_j^- < n' < \alpha_j^+ n - \beta_j^+,\ n \geq \chi_j,\ n' \geq \chi_{j+1}\right\},$$

where $\alpha_j^- = \delta_{j+1}/\delta_j \max\{1, \mu_j\}$, $\alpha_j^+ = \delta_{j+1}/\delta_j \min\{1, \mu_j\}$ *and* $\beta_j^{\pm} = \alpha_j^{\pm} + 1$ *for all* j. *There are pairwise disjoint strict horizontal slabs*

$$\widetilde{\mathcal{K}}_{j,n}^{\varsigma} \subsetneq_{\mathrm{h}} \widetilde{\mathcal{Q}}_{j,n}^{\varsigma}, \qquad \forall j,\ n \geq \chi_j,\ \varsigma \in \{-,+\},$$

such that:

(a) $f^n : \widetilde{\mathcal{K}}_{j,n}^{\varsigma} \rightsquigarrow \widetilde{\mathcal{K}}_{j+1,n'}^{\varsigma'}$*; and*

(b) $f^n(\mathcal{K}^{\varsigma}_{j,n}) \cap \mathcal{K}^{\varsigma'}_{j+1,n'} \cap \mathcal{L} = \emptyset$,

for all j, $(n,n') \in \Xi_j$ *and* $\varsigma, \varsigma' \in \{-,+\}$. *(See Definition 12 for the meaning of* $\mathcal{L}$*.)*

Proof. Inequalities (9) and (10) become *strict* for any $(n,n') \in \Xi_j$. We consider the oriented cells $\widetilde{\mathcal{R}}_{j+1,n}$, where

$$\mathcal{R}_{j+1,n} = \bigcup_{(n,n')\in\Xi_j} \mathcal{Q}_{j+1,n'} \subset \mathcal{D}_{j+1},$$

and orientations are chosen in such a way that the left and right sides of these big oriented cells are still contained in $\mathcal{L}_{j+1}$ and $\mathcal{L}^1_{j+1}$, respectively. Note that $\mathcal{D}_{j+1} \setminus \mathcal{R}_{j+1,n}$ has a connected component above and other one below $\mathcal{R}_{j+1,n}$.

Fix an index j such that $\mu_j < 1$. Let $n \geq \chi_j$. We refer to Figure 6 for a visual guide. The reader should imagine that the blue quadrilateral shown in that figure is our whole cell $\mathcal{R}_{j+1,n}$.

Strict versions of inequalities (9) and (10) imply that the images by f^n of both the base side of $\widetilde{\mathcal{Q}}^-_{j,n}$ and the top side of $\widetilde{\mathcal{Q}}^+_{j,n}$ are *strictly* above $\mathcal{R}_{j+1,n}$; whereas the image by f^n of the top side of $\widetilde{\mathcal{Q}}^-_{j,n}$, which coincides with the base side of $\widetilde{\mathcal{Q}}^+_{j,n}$, is *strictly* below $\mathcal{R}_{j+1,n}$. Thus, there are strict horizontal slabs $\widetilde{\mathcal{K}}^{\varsigma}_{j,n} \subsetneq_{\mathrm{h}} \widetilde{\mathcal{Q}}^{\varsigma}_{j,n}$, with $\varsigma \in \{-,+\}$, such that the images by f^n of both the base side of $\widetilde{\mathcal{K}}^-_{j,n}$ and the top side of $\widetilde{\mathcal{K}}^+_{j,n}$ are strictly above $\mathcal{R}_{j+1,n}$; whereas the image by f^n of both the top side of $\widetilde{\mathcal{K}}^-_{j,n}$ and the base side of $\widetilde{\mathcal{K}}^+_{j,n}$ are strictly below $\mathcal{R}_{j+1,n}$. Consequently, $f^n(\mathcal{K}^{\varsigma}_{j,n}) \cap \mathcal{R}_{j+1,n} \cap \mathcal{L} = \emptyset$ and $f^n : \widetilde{\mathcal{K}}^{\varsigma}_{j,n} \rightsquigarrow \widetilde{\mathcal{R}}_{j+1,n}$, which implies that $f^n(\mathcal{K}^{\varsigma}_{j,n}) \cap \mathcal{K}^{\varsigma'}_{j+1,n'} \cap \mathcal{L} = \emptyset$ and $f^n : \widetilde{\mathcal{K}}^{\varsigma}_{j,n} \rightsquigarrow \widetilde{\mathcal{K}}^{\varsigma'}_{j+1,n'}$ for all $(n,n') \in \Xi_j$ and $\varsigma, \varsigma' \in \{-,+\}$, since $\widetilde{\mathcal{K}}^{\varsigma'}_{j+1,n'} \subset_{\mathrm{h}} \widetilde{\mathcal{R}}_{j+1,n}$ for all $(n,n') \in \Xi_j$ and $\varsigma' \in \{-,+\}$. This ends the proof of the stretching and intersecting properties when $\mu_j < 1$. The case $\mu_j > 1$ follows from similar arguments. We omit the details.

Finally, these strict horizontal slabs are necessarily pairwise disjoint because the original fundamental quadrilaterals $\widetilde{\mathcal{Q}}^{\varsigma}_{j,n}$ only share some of their horizontal sides. □

5 Symbols, shift spaces and shift maps

In this section, we define an alphabet $\boldsymbol{Q} \subset \mathbb{Z}^k$ with infinitely many symbols, then we consider two shift spaces $\mathfrak{Q}^+ \subset \boldsymbol{Q}^{\mathbb{N}_0}$ and $\mathfrak{Q} \subset \boldsymbol{Q}^{\mathbb{Z}}$ of admissible one-sided and two-sided sequences, and finally we study some properties of the shift map $\sigma : \mathfrak{Q} \to \mathfrak{Q}$. We present these objects in a separate section, minimizing their relation with circular polygons, since we believe that they will be useful in future works about other problems.

For brevity, we will use the term shift instead of subshift, but $\mathfrak{Q}^+ \subsetneq \boldsymbol{Q}^{\mathbb{N}_0}$ and $\mathfrak{Q} \subsetneq \boldsymbol{Q}^{\mathbb{Z}}$. The sets $\boldsymbol{Q}$, $\mathfrak{Q}^+$ and $\mathfrak{Q}$ are defined in terms of some positive factors $\alpha^{\pm}_j$, some positive addends $\beta^{\pm}_j$ and some integers $\chi_j \geq 2$ for $j = 1, \ldots, k$, with $k \geq 1$. (There are interesting billiard problems that will require $k < 3$, or even $k = 1$.)

We only assume three hypotheses about these factors and integers:

(A) $0 < \alpha^-_j < \alpha^+_j$ for $j = 1, \ldots, k$;

(B) $\alpha := \alpha^+ > 1$ and $\alpha^+\alpha^- = 1$, where $\alpha^{\pm} = \prod_{j=1}^k \alpha^{\pm}_j$; and

(X) Integers $\chi_2, \ldots, \chi_k \geq 2$ are large enough and $\chi_1 \gg \chi_2, \ldots, \chi_k$.

There is no assumption on the addends. Clearly, all arcs $\Gamma_1, \ldots, \Gamma_k$ of the circular polygon Γ are equally important, so $\chi_1 \gg \chi_2, \ldots, \chi_k$ is a purely technical hypothesis. It is used only once, at the beginning of the proof of Lemma 24. It is needed just to establish the topological transitivity of the subshift map. The rest of Theorem A, as well as Theorems C, D and E do not need it.

We remark two facts related to billiards in circular polygons, although we forget about billiards in the rest of this section. The first remark is a trivial verification.

Lemma 20. *The factors* $\alpha^{\pm}_j$ *defined in Lemma 18 satisfy hypotheses* (**A**) *and* (**B**).

Proof. Hypothesis (**A**) follows from properties $\mu_j \neq 1$. Hypothesis (**B**) follows from the telescopic products

$$\prod_{j=1}^k \frac{\delta_j}{\delta_{j+1}} = 1, \qquad \prod_{j=1}^k \mu_j = \prod_{j=1}^k \sqrt{\frac{r_j}{r_{j+1}}} = 1,$$

which are easily obtained from the cyclic identities $\delta_{k+1} = \delta_1$ and $r_{k+1} = r_1$. □

The second remark is that we prefer to encode in a single symbol all information related to each complete turn around Γ, although we may construct our symbolic dynamics directly with the disjoints sets $\widetilde{\mathcal{K}}^{\varsigma}_{j,n}$ as symbols. That is, if a generic sliding orbit follows, along a complete turn around the boundary Γ, the itinerary

$$\mathcal{K}^{\varsigma_1}_{1,n_1} \subset \mathcal{D}_1, \mathcal{K}^{\varsigma_2}_{2,n_2} \subset \mathcal{D}_2, \dots, \mathcal{K}^{\varsigma_k}_{k,n_k} \subset \mathcal{D}_k,$$

where $\widetilde{\mathcal{K}}^{\varsigma}_{j,n}$ are the pairwise disjoint horizontal slabs described in Corollary 19, then we construct the symbol

$$\boldsymbol{q} = (q_1, \dots, q_k) \in \mathbb{Z}^k, \qquad |q_j| = n_j, \qquad \operatorname{sign}(q_j) = \varsigma_j,$$

which motivates the following definition.

Definition 21. The *alphabet of admissible symbols* is the set

$$\boldsymbol{Q} = \left\{ \boldsymbol{q} = (q_1, \dots, q_k) \in \mathbb{Z}^k : \begin{array}{l} \alpha_j^- |q_j| + \beta_j^- < |q_{j+1}| < \alpha_j^+ |q_j| - \beta_j^+, \ \forall j = 1, \dots, k-1 \\ |q_j| \geq \chi_j, \ \forall j = 1, \dots, k \end{array} \right\}.$$

This alphabet has infinitely many symbols by hypothesis (**A**). Thinking in the billiard motivation behind these symbols, we ask that symbols associated to consecutive turns around Γ satisfy the following admissibility condition.

Definition 22. We say that a finite, one-sided, or two-sided sequence of admissible symbols $\mathfrak{q} = (\boldsymbol{q}^i)_{i \in I} \subset \boldsymbol{Q}$, with $\boldsymbol{q}^i = (q_1^i, \dots, q_k^i)$, is *admissible* if and only if

$$\alpha_k^- |q_k^i| + \beta_k^- < |q_1^{i+1}| < \alpha_k^+ |q_k^i| - \beta_k^+, \qquad \forall i.$$

Admissible sequences are written with Fraktur font: $\mathfrak{q}$. Its vector symbols are written with boldface font and labeled with superscripts: $\boldsymbol{q}^i$. Components of admissible symbols are written with the standard font and labeled with subscripts: q_j^i or q_j.

Definition 23. The *shift spaces of admissible sequences* are

$$\mathfrak{Q}^+ = \left\{ \mathfrak{q} = (\boldsymbol{q}^i)_{i \in \mathbb{N}_0} \in \boldsymbol{Q}^{\mathbb{N}_0} : \mathfrak{q} \text{ is admissible} \right\},$$
$$\mathfrak{Q} = \left\{ \mathfrak{q} = (\boldsymbol{q}^i)_{i \in \mathbb{Z}} \in \boldsymbol{Q}^{\mathbb{Z}} : \mathfrak{q} \text{ is admissible} \right\}.$$

If $\boldsymbol{q} = (q_1, \dots, q_k) \in \boldsymbol{Q}$, then we write $|\boldsymbol{q}| = |q_1| + \dots + |q_k|$. We equip $\mathfrak{Q}$ with the topology defined by the metric

$$d_{\mathfrak{Q}} : \mathfrak{Q} \times \mathfrak{Q} \to [0, +\infty), \qquad d_{\mathfrak{Q}}(\mathfrak{p}, \mathfrak{q}) = \sum_{i \in \mathbb{Z}} \frac{1}{2^{|i|}} \frac{|\boldsymbol{p}^i - \boldsymbol{q}^i|}{1 + |\boldsymbol{p}^i - \boldsymbol{q}^i|}.$$

We want to estimate the size of the maxima (sometimes, the minima as well) of the sets

$$\Xi_j^i(n) = \left\{ n_j^i \in \mathbb{N} : \exists \mathfrak{q} \in \mathfrak{Q} \text{ such that } |q_1^0| = n \text{ and } |q_j^i| = n_j^i \right\}, \quad i \in \mathbb{Z}, \ j = 1, \dots, k, \tag{11}$$

when $n \geq \chi_1$ or $n \gg 1$. We ask in (11) for the existence of some $\mathfrak{q} \in \mathfrak{Q}$ —that is, some two-sided infinite sequence—, but it does not matter. We would obtain exactly the same sets just by asking the existence of some finite sequence $q_1^0, \dots, q_k^0, q_1^1, \dots, q_k^1, \dots, q_1^i, \dots, q_j^i$ that satisfies the corresponding admissibility conditions.

Several estimates about maxima and minima of sets (11) are listed below. Their proofs have been postponed to Appendix A.

Lemma 24. *We assume hypotheses* (**A**), (**B**) *and* (**X**). *Set* $\zeta_j^i(n) = \min \Xi_j^i(n)$, $\xi_j^i(n) = \max \Xi_j^i(n)$. *Let* $\rho^0(n) = n$ *and* $\rho^i(n) = \sum_{j=1}^k \sum_{m=0}^{i-1} \xi_j^m(n)$ *for all* $i \geq 1$.

(a) There are positive constants $\nu < \lambda$, $\nu' < \lambda'$, $\tau < 1$ *and* $\gamma^{\pm}$, *which depend on factors* $\alpha_j^{\pm}$ *and addends* $\beta_j^{\pm}$ *but not on integers* χ_j, *such that the following properties hold.*

i) $\nu n \leq \xi_j^0(n) \leq \lambda n$ *for all* $j = 1, \dots, k$ *and* $n \geq \chi_1$;

ii) $\tau \alpha^{|i|} \xi_j^0(n) \leq \xi_j^i(n) \leq \alpha^{|i|} \xi_j^0(n)$ *for all* $j = 1, \dots, k$, $i \in \mathbb{Z}$ *and* $n \geq \chi_1$;

iii) $\nu' \xi_j^i(n) \leq \rho^i(n) \leq \rho^{i+1}(n) \leq \lambda' \xi_j^i(n)$ *for all* $j = 1, \dots, k$, $i \geq 0$, *and* $n \geq \chi_1$;

iv) $\zeta_1^1(n) \leq \max\{\chi_1, n/\alpha + \gamma^-\} \leq n-1 < n+1 \leq \alpha n - \gamma^+ \leq \xi_1^1(n)$ *for all* $n > \chi_1$ *and* $\zeta_1^1(n) = n < n+1 \leq \alpha n - \gamma^+ \leq \xi_1^1(n)$ *for* $n = \chi_1$*; and*

v) Once fixed any $N \in \mathbb{N}$*, we have that*

$$\chi_1 \leq \zeta_1^1(n) \leq n/\alpha + \gamma^- \leq n - N < n + N \leq \alpha n - \gamma^+ \leq \xi_1^1(n),$$

for all sufficiently large n*.*

(b) $\Xi_j^i(n) = [\zeta_j^i(n), \xi_j^i(n)] \cap \mathbb{N}$ *for all* $j \mod k$*,* $i \in \mathbb{Z}$ *and* $n \geq \chi_1$*; that is,* $\Xi_j^i(n)$ *has no gaps in* $\mathbb{N}$*. Besides,* $\big[\max\{\chi_1, n - |i|\}, n + |i|\big] \cap \mathbb{N} \subset \Xi_1^i(n)$ *for all* $i \in \mathbb{Z}$ *and* $n \geq \chi_1$*.*

Corollary 25. *We assume hypotheses* (**A**)*,* (**B**) *and* (**X**)*.*

(a) Given any $\boldsymbol{q}^-, \boldsymbol{q}^+ \in \boldsymbol{Q}$ *there is an admissible sequence of the form* $\big(\boldsymbol{q}^-, \boldsymbol{q}^1, \ldots, \boldsymbol{q}^l, \boldsymbol{q}^+\big)$ *for some* $l \in \mathbb{N}$ *and* $\boldsymbol{q}^1, \ldots, \boldsymbol{q}^l \in \boldsymbol{Q}$*.*

(b) Given any $N \in \mathbb{N}$*, there is a subset* $\boldsymbol{Q}_N \subset \boldsymbol{Q}$*, with* $\#\boldsymbol{Q}_N = N$*, such that the short sequence* $(\boldsymbol{q}, \boldsymbol{q}')$ *is admissible for all* $\boldsymbol{q}, \boldsymbol{q}' \in \boldsymbol{Q}_N$*.*

(c) $\mathfrak{Q} \neq \emptyset$*.*

Proof. (a) Let $l = |q_1^- - q_1^+| - 1$. Part (b) of Lemma 24 implies that $q_1^+ \in \Xi_1^{l+1}(q_1^-)$. Therefore, we can construct iteratively such a sequence $\boldsymbol{q}^1, \ldots, \boldsymbol{q}^l$.

(b) Fix $N \in \mathbb{N}$. Part (av) of Lemma 24 implies that if $\boldsymbol{q}, \boldsymbol{q}' \in \boldsymbol{Q}$ with $|q_1 - q_1'| \leq N$ and $|q_1|, |q_1'| \gg 1$, then $(\boldsymbol{q}, \boldsymbol{q}')$ is admissible. So, we can take any subset $\boldsymbol{Q}_N = \{\boldsymbol{q}^1, \ldots, \boldsymbol{q}^N\} \subset \boldsymbol{Q}$ such that $|q_1^n| = |q_1^1| + n - 1$ with $|q_1^1| \gg 1$. Clearly, $\#\boldsymbol{Q}_N = N$.

(c) Let $\mathfrak{q} = (\boldsymbol{q}^i)_{i \in \mathbb{Z}}$ with $\boldsymbol{q}^i = (q_1^i, \ldots, q_k^i) \in \boldsymbol{Q}$ such that $|q_1^{i+1} - q_1^i| \leq 1$. Then $\mathfrak{q} \in \mathfrak{Q}$. □

Definition 26. The *shift map* $\sigma : \mathfrak{Q} \to \mathfrak{Q}$, $\mathfrak{p} = \sigma(\mathfrak{q})$, is given by $\mathfrak{p}^i = \mathfrak{q}^{i+1}$ for all $i \in \mathbb{Z}$.

The following proposition tells us some important properties of the shift map. Note that by *topological transitivity* we mean for any nonempty open sets $U, V \subset \mathfrak{Q}$ there is $n \in \mathbb{N}$ such that $\sigma^n(U) \cap V \neq \emptyset$. If $N \in \mathbb{N}$, $\Sigma_N = \{1, \ldots, N\}^{\mathbb{Z}}$ and the shift map $\sigma_N : \Sigma_N \to \Sigma_N$, $(t_i)_{i \in \mathbb{Z}} = \sigma_N\big((s_i)_{i \in \mathbb{Z}}\big)$, is given by $t_i = s_{i+1}$ for all $i \in \mathbb{Z}$, then we say that $\sigma_N : \Sigma_N \to \Sigma_N$ is the *full* N*-shift*. We denote by $h_{\mathrm{top}}(f)$ the *topological entropy* of a continuous self-map f.

Proposition 27. *We assume hypotheses* (**A**)*,* (**B**) *and* (**X**)*.*

The shift map $\sigma : \mathfrak{Q} \to \mathfrak{Q}$ *exhibits topological transitivity and sensitive dependence on initial conditions, has infinite topological entropy, and contains the full* N*-shift as a topological factor for any* $N \in \mathbb{N}$*. Besides, the subshift space of periodic admissible sequences*

$$\mathfrak{P} = \{\mathfrak{q} \in \mathfrak{Q} : \exists p \in \mathbb{N} \textit{ such that } \sigma^p(\mathfrak{q}) = \mathfrak{q}\}$$

is dense in the shift space $\mathfrak{Q}$*.*

Proof. On the one hand, part (a) of Corollary 25 implies that the shift map $\sigma : \mathfrak{Q} \to \mathfrak{Q}$ is equivalent to a transitive topological Markov chain. It is well-known that such objects exhibit topological transitivity, sensitive dependence on initial conditions, and density of periodic points. See, for example, Sections 1.9 and 3.2 of [38].

On the other hand, let $\boldsymbol{Q}_N = \{\boldsymbol{q}^1, \ldots, \boldsymbol{q}^N\}$ be the set provided in part (b) of Corollary 25. Set $\mathfrak{Q}_N = (\boldsymbol{Q}_N)^{\mathbb{Z}}$. We consider the bijection

$$g = (g^i)_{i \in \mathbb{Z}} : \mathfrak{Q}_N \to \Sigma_N, \quad g^i(\boldsymbol{q}^n) = n, \qquad \forall i \in \mathbb{Z},\ \forall n \in \{1, \ldots, N\}.$$

Then $\mathfrak{Q}_N$ is a subshift space of $\mathfrak{Q}$. That is, $\sigma(\mathfrak{Q}_N) = \mathfrak{Q}_N$. Besides, the diagram

$$\begin{array}{ccc} \mathfrak{Q}_N & \xrightarrow{\sigma_{|\mathfrak{Q}_N}} & \mathfrak{Q}_N \\ {\scriptstyle g}\big\downarrow & & \big\downarrow{\scriptstyle g} \\ \Sigma_N & \xrightarrow[\sigma_N]{} & \Sigma_N \end{array}$$

commutes, so $h_{\mathrm{top}}(\sigma) \geq h_{\mathrm{top}}(\sigma_{|\mathfrak{Q}_N}) = h_{\mathrm{top}}(\sigma_N) = \log N$ for all $N \in \mathbb{N}$. This means that σ has infinite topological entropy. □

6 Chaotic motions

In this section, we detail the construction of a domain accumulating on the boundary of the phase space on which the dynamics is semiconjugate to a shift on infinitely many symbols, thus proving Theorem A; in fact, we reformulate Theorem A in the form of Theorem 31 below. The proof uses the method of *stretching along the paths* summarised in Section 3, the Fundamental Corollary obtained in Section 4 and the shift map described in Section 5.

Recall that the quantities $\alpha_j^{\pm}$, $\beta_j^{\pm} = \alpha_j^{\pm} + 1$ and $\chi_j \geq 2$, introduced in Lemma 17, Lemma 18, and Corollary 19 satisfy hypotheses (**A**) and (**B**), and moreover we assume hypothesis (**X**), so Proposition 27 holds.

We now introduce some notation that is convenient for the statements and proofs in this section.

Definition 28. The *partial sums* $s^i, s_j^i : \mathfrak{Q} \to \mathbb{Z}$ for $i \in \mathbb{Z}$ and $j = 1, \ldots, k$ are defined by

$$s^i(\mathfrak{q}) = \begin{cases} \displaystyle\sum_{m=0}^{i-1} \sum_{j=1}^{k} |q_j^m|, & \text{for } i \geq 0, \\ \displaystyle -\sum_{m=i}^{-1} \sum_{j=1}^{k} |q_j^m|, & \text{for } i < 0, \end{cases} \qquad s_j^i(\mathfrak{q}) = s^i(\mathfrak{q}) + \sum_{m=1}^{j-1} |q_m^i|.$$

The partial sums $s^i, s_j^i : \mathfrak{Q}^+ \to \mathbb{N}_0$ are analogously defined for $i \geq 0$ and $j = 1, \ldots, k$.

The following proposition gives the relationship between some types of admissible sequences (two-sided: $\mathfrak{q} \in \mathfrak{Q}$, one-sided: $\mathfrak{q} \in \mathfrak{Q}^+$ and periodic: $\mathfrak{q} \in \mathfrak{P}$) and orbits of f with prescribed itineraries in the set of pairwise disjoint cells $\mathcal{K}_{j,n}^{\varsigma}$ (introduced in Corollary 19) with $j = 1, \ldots, k$, $n \geq \chi_j$ and $\varsigma \in \{-, +\}$. It is the key step in obtaining chaotic properties.

Proposition 29. *We have the following three versions.*

(**T**) *If $\mathfrak{q} \in \mathfrak{Q}$, then there is $x \in \mathcal{D}_1$ such that*

$$f^{s_j^i(\mathfrak{q})}(x) \in \mathcal{K}_{j,|q_j^i|}^{\operatorname{sign}(q_j^i)}, \qquad \forall i \in \mathbb{Z},\ \forall j = 1, \ldots, k. \tag{12}$$

(**O**) *If $\mathfrak{q} \in \mathfrak{Q}^+$, then there is a path $\gamma \subset \mathcal{D}_1$ such that:*

i) $f^{s_j^i(\mathfrak{q})}(\gamma) \subset \mathcal{K}_{j,|q_j^i|}^{\operatorname{sign}(q_j^i)}$ for all $i \geq 0$ and $j = 1, \ldots, k$; and

ii) γ is horizontal in $\mathcal{D}_1$ (that is, γ connects the left side $\mathcal{L}_1$ with the right side $\mathcal{L}_1^1$).

(**P**) *If $\mathfrak{q} \in \mathfrak{P}$ has period p, then there is a point $x \in \mathcal{D}_1$ such that:*

i) $f^{s_j^i(\mathfrak{q})}(x) \in \mathcal{K}_{j,|q_j^i|}^{\operatorname{sign}(q_j^i)}$ for all $i \in \mathbb{Z}$ and $j = 1, \ldots, k$; and

ii) $f^{s^p(\mathfrak{q})}(x) = x$, so x is a $(p, s^p(q))$-periodic point of f with period $s^p(\mathfrak{q})$ and rotation number $p/s^p(\mathfrak{q})$.

All these billiard orbits are contained in the generic sliding set $\mathcal{S}_0$. In particular, they have no points in the extended singularity set $\mathcal{L} = \bigcup_j \left(\mathcal{L}_j \cup \mathcal{L}_j^{1/2} \cup \mathcal{L}_j^1\right)$. Obviously, these claims only hold for forward orbits in version (**O**).

Proof. It is a direct consequence of Theorem 9, Corollary 19, the definitions of admissible symbols and admissible sequences, and the definition of rotation number. □

To adapt the language of [48, 49, 51] to our setting, one could say that Proposition 29 implies that the billiard map *induces chaotic dynamics on infinitely many symbols*.

Remark 30. The partial sum $s^i(\mathfrak{q})$, with $i \geq 0$, introduced in Definition 28 counts the number of impacts that any of its corresponding sliding billiard trajectories have after the first i turns around Γ. Analogously, $s_j^i(\mathfrak{q})$, with $i \geq 0$, adds to the previous count the number of impacts in the first $j - 1$ arcs at the $(i + 1)$-th turn. There is no ambiguity in these counts, because generic sliding billiard trajectories have no impacts on the set of nodes $\Gamma_\star$, see Remark 14. The partial sums with $i < 0$ store information about the backward orbit.

Let us introduce four subsets of the first fundamental domain that will be invariant under a return map F yet to be defined. First, we consider the *fundamental generic sliding set*

$$\mathcal{R} = \mathcal{S}_0 \cap \mathcal{D}_1 \subset \operatorname{Int} \mathcal{D}_1.$$

Any f-orbit that begins in $\mathcal{R}$ returns to $\mathcal{R}$ after a finite number of iterations of the billiard map f. Let $\tau : \mathcal{R} \to \mathbb{N}$ be the *return time* defined as $\tau(x) = \min\{n \in \mathbb{N} : f^n(x) \in \mathcal{R}\}$. Then $F : \mathcal{R} \to \mathcal{R}$, $F(x) = f^{\tau(x)}(x)$, is the promised *return map*. The return map $F : \mathcal{R} \to \mathcal{R}$ is a homeomorphism since the billiard map $f : \mathcal{M} \to \mathcal{M}$ is a homeomorphism and $\mathcal{R}$ is contained in the interior of the fundamental set $\mathcal{D}_1$.

Next, we define the sets

$$\begin{aligned} \mathcal{I} &= \big\{x \in \mathcal{M} : \exists \mathfrak{q} \in \mathfrak{Q} \text{ such that the prescribed itinerary (12) takes place}\big\}, \\ \mathcal{P} &= \big\{x \in \mathcal{I} : \exists p \in \mathbb{N} \text{ such that } F^p(x) = x\big\} \end{aligned}$$

and the map $h : \mathcal{I} \to \mathfrak{Q}$, $h(x) = \mathfrak{q}$, where $\mathfrak{q}$ is the unique admissible sequence such that the prescribed itinerary (12) takes place. It is well-defined because cells $\mathcal{K}^{\varsigma}_{j,n}$ are pairwise disjoint. This is the topological semiconjugacy we were looking for. Clearly,

$$\tau(x) = s^1(\mathfrak{q}) = |q_1^0| + \cdots + |q_k^0|, \qquad \forall x \in \mathcal{I},\ \mathfrak{q} = h(x) \tag{13}$$

where $\tau(x)$ is the return time and the partial sum $s^1(\mathfrak{q})$ counts the number of impacts after the first turn around Γ of the billiard orbit starting at x.

Theorem 31. *The sets $\mathcal{P}$, $\mathcal{J} := \overline{\mathcal{P}}$, $\mathcal{I}$ and $\mathcal{R}$ are F-invariant:*

$$F(\mathcal{P}) = \mathcal{P}, \qquad F(\mathcal{J}) = \mathcal{J}, \qquad F(\mathcal{I}) \subset \mathcal{I}, \qquad F(\mathcal{R}) = \mathcal{R}.$$

Besides, $\emptyset \neq \mathcal{P} \subsetneq \mathcal{J} \subset \mathcal{I} \subset \mathcal{R}$. The maps $h : \mathcal{I} \to \mathfrak{Q}$, $h_{|\mathcal{J}} : \mathcal{J} \to \mathfrak{Q}$ and $h_{|\mathcal{P}} : \mathcal{P} \to \mathfrak{P}$ are continuous surjections, and the three diagrams

$$\begin{tikzcd} \mathcal{I} \arrow[r, "F_{|\mathcal{I}}"] \arrow[d, "h"'] & \mathcal{I} \arrow[d, "h"] \\ \mathfrak{Q} \arrow[r, "\sigma"'] & \mathfrak{Q} \end{tikzcd} \qquad \begin{tikzcd} \mathcal{J} \arrow[r, "F_{|\mathcal{J}}"] \arrow[d, "h_{|\mathcal{J}}"'] & \mathcal{J} \arrow[d, "h_{|\mathcal{J}}"] \\ \mathfrak{Q} \arrow[r, "\sigma"'] & \mathfrak{Q} \end{tikzcd} \qquad \begin{tikzcd} \mathcal{P} \arrow[r, "F_{|\mathcal{P}}"] \arrow[d, "h_{|\mathcal{P}}"'] & \mathcal{P} \arrow[d, "h_{|\mathcal{P}}"] \\ \mathfrak{P} \arrow[r, "\sigma_{|\mathfrak{P}}"'] & \mathfrak{P} \end{tikzcd} \tag{14}$$

commute. Periodic points of $F_{|\mathcal{J}}$ are dense in $\mathcal{J}$. Given any $\mathfrak{q} \in \mathfrak{P}$ with period p, there is at least one $x \in (h_{|\mathcal{P}})^{-1}(\mathfrak{q}) \in \mathcal{P}$ such that $f^{s^p(\mathfrak{q})}(x) = F^p(x) = x$.

Proof. Properties $F(\mathcal{P}) = \mathcal{P}$, $F(\mathcal{R}) = \mathcal{R}$ and $\mathcal{P} \subset \mathcal{I}$ are trivial, by construction. Inclusion $\mathcal{I} \subset \mathcal{R}$ follows from the definitions of both sets and property (b) of Corollary 19.

Let us prove that $h : \mathcal{I} \to \mathfrak{Q}$ is continuous and surjective. Surjectivity follows directly from version (**T**) of Proposition 29. Choose any $x \in \mathcal{I}$ and $\epsilon > 0$. Choose $l \in \mathbb{N}$ such that $\sum_{|i|>l} 2^{-|i|} < \epsilon$. Let $\mathfrak{q} = (\boldsymbol{q}^i)_{i\in\mathbb{Z}} = h(x)$ with $\boldsymbol{q}^i = (q_1^i, \ldots, q_k^i)$. Using that the compact sets $\mathcal{K}^{\varsigma}_{j,n}$ are mutually disjoint, F is a homeomorphism and condition (12), we can find $\delta_j^i > 0$ for each $|i| \leq l$ and $j = 1, \ldots, k$ such that

$$f^{s_j^i(\mathfrak{q})}\big(\mathcal{B}_{\delta_j^i}(x) \cap \mathcal{I}\big) \subset \mathcal{K}^{\operatorname{sign}(q_j^i)}_{j,|q_j^i|}, \qquad \forall |i| \leq l,\ \forall j = 1, \ldots, k.$$

Here, $\mathcal{B}_\delta(x)$ is the disc of radius δ centered at x. If $\delta = \min\{\delta_j^i : |i| \leq l,\ j = 1, \ldots, k\}$, then

$$d(x, y) < \delta \text{ and } \mathfrak{p} = (\boldsymbol{p}^i)_{i\in\mathbb{Z}} = h(y) \Longrightarrow \boldsymbol{p}^i = \boldsymbol{q}^i \text{ for each } |i| \leq l.$$

Therefore,

$$d_{\mathfrak{Q}}\big(h(y), h(x)\big) = d_{\mathfrak{Q}}(\mathfrak{p}, \mathfrak{q}) = \sum_{|i|>l} \frac{1}{2^{|i|}} \frac{|\boldsymbol{p}^i - \boldsymbol{q}^i|}{1 + |\boldsymbol{p}^i - \boldsymbol{q}^i|} < \sum_{|i|>l} \frac{1}{2^{|i|}} < \epsilon,$$

which implies that $h : \mathcal{I} \to \mathfrak{Q}$ is continuous.

Next, we prove simultaneously that $F(\mathcal{I}) \subset \mathcal{I}$ and that $\sigma \circ h = h \circ F_{|\mathcal{I}}$. Let $x \in \mathcal{I}$, $y = F(x) \in \mathcal{R}$, $\mathfrak{q} = (\boldsymbol{q}^i)_{i\in\mathbb{Z}} = h(x)$ with $\boldsymbol{q}^i = (q_1^i, \dots, q_k^i)$, and $\mathfrak{p} = (\boldsymbol{p}^i)_{i\in\mathbb{Z}} = \sigma(\mathfrak{q}) \in \mathfrak{Q}$ with $\boldsymbol{p}^i = (p_1^i, \dots, p_k^i)$, so $\boldsymbol{p}^i = \boldsymbol{q}^{i+1}$ and $p_j^i = q_j^{i+1}$. The prescribed itinerary (12) and relation (13) imply that

$$f^{s_j^i(\mathfrak{p})}(y) = f^{s_j^i(\sigma(\mathfrak{q}))}(F(x)) = f^{s_j^{i+1}(\mathfrak{q})-s^1(\mathfrak{q})}\big(f^{s^1(\mathfrak{q})}(x)\big) = f^{s_j^{i+1}(\mathfrak{q})}(x) \in \mathcal{K}_{j,|q_j^{i+1}|}^{\mathrm{sign}(q_j^{i+1})} = \mathcal{K}_{j,|p_j^i|}^{\mathrm{sign}(p_j^i)}$$

for all $i \in \mathbb{Z}$ and $j = 1, \dots, k$, so $\sigma(h(x)) = \sigma(\mathfrak{q}) = \mathfrak{p} = h(y) = \sigma(F(x))$ and $F(x) = y \in h^{-1}(\mathfrak{p}) \subset \mathcal{I}$ for all $x \in \mathcal{I}$, as we wanted to prove. Hence, the first diagram in (14) defines a topological semiconjugacy.

Let us check that $\mathcal{J} := \overline{\mathcal{P}} \subset \mathcal{I}$ and $F(\mathcal{J}) = \mathcal{J}$. We have $\mathcal{J} \subset \mathcal{I}$, because $\mathcal{I} = h^{-1}(\mathfrak{Q})$ is closed (continuous preimage of a closed set). Besides, on one hand we have $\mathcal{P} = F(\mathcal{P}) \subset F(\overline{\mathcal{P}}) = F(\mathcal{J})$ implying that $\mathcal{J} = \overline{\mathcal{P}} \subset \overline{F(\mathcal{J})} = F(\mathcal{J})$, while on the other we have $F(\mathcal{J}) = F(\overline{\mathcal{P}}) \subset \overline{F(\mathcal{P})} = \overline{\mathcal{P}} = \mathcal{J}$. To establish that the second diagram in (14) is still a topological semiconjugacy, we must prove that $h(\mathcal{J}) = \mathfrak{Q}$. We clearly have $h(\mathcal{J}) \subset \mathfrak{Q}$ since $\mathcal{J} \subset \mathcal{I}$, and since $h_{|\mathcal{I}}$ is a semiconjugacy. Meanwhile, since $\mathcal{J}$ is compact (closed by definition and contained in the bounded set $\mathcal{D}_1$) so too is $h(\mathcal{J})$; moreover $h(\mathcal{J})$ contains $h(\mathcal{P}) = \mathfrak{P}$ which is dense in $\mathfrak{Q}$ by Proposition 27, and so we obtain $\mathfrak{Q} = \overline{\mathfrak{P}} \subset h(\mathcal{J})$. Therefore $h(\mathcal{J}) = \mathfrak{Q}$.

To complete the proof of the theorem, notice that periodic points of $F_{|\mathcal{J}}$ are dense in $\mathcal{J}$ by construction and the last claim of the Theorem 31 follows from version (**P**) of Proposition 29, which also implies $\mathcal{P} \neq \emptyset$. □

Proposition 27 and Theorem 31 imply Theorem A as stated in the introduction and it is the first step in proving Theorems C, D and E. For instance, upon combining Theorem 31 with the topological transitivity of the shift map guaranteed by Proposition 27, we already obtain the existence of trajectories approaching the boundary asymptotically. It remains to determine the *optimal* rate of diffusion. This is done in Section 7 by analysing the sequences $\mathfrak{q} = (\boldsymbol{q}^i)_{i\geq 0} \in \mathfrak{Q}^+$ for which $s^i(\mathfrak{q})$ increases in the *fastest* possible way as $i \to +\infty$. Lemma 24 plays a role in that analysis.

We end this section with three useful corollaries. First, we prove Corollary B on final sliding motions.

Proof of Corollary B. The clockwise case is a by-product of the counter-clockwise one, because if we concatenate the arcs $\Gamma_1, \dots, \Gamma_k$ of the original circular polygon Γ in the reverse order $\Gamma_k, \dots, \Gamma_1$, then we obtain the reversed circular polygon Γ' with the property that counter-clockwise sliding billiard trajectories in Γ' are in 1-to-1 correspondence with clockwise sliding billiard trajectories in Γ. Thus, it suffices to consider the counter-clockwise case.

Symbols $\boldsymbol{q} = (q_1, \dots, q_k) \in \boldsymbol{Q} \subset \mathbb{Z}^k$ keep track of the proximity of the fundamental quadrilaterals $\mathcal{Q}_{j,|q_j|}$ to the inferior boundary of $\mathcal{M}$. That is, the larger the absolute value $|q_j|$, the smaller the angle of reflection θ for any $x = (\varphi, \theta) \in \mathcal{Q}_{j,|q_j|}$. For this reason, by construction, if one considers a bounded sequence in $\mathfrak{Q}$, (respectively, a sequence $\mathfrak{q} \in \mathfrak{Q}$ such that $\chi_j \leq \min_{i\in\mathbb{Z}} q_j^i < \limsup_{|i|\to+\infty} q_j^i = +\infty$ for all $j = 1, \dots, k$) (respectively, a sequence $\mathfrak{q} \in \mathfrak{Q}$ such that $\lim_{|i|\to+\infty} q_j^i = +\infty$ for all $j = 1, \dots, k$), the corresponding sliding orbit in $\mathcal{J} \subset \mathcal{M}$ belongs to $\mathcal{B}_0^- \cap \mathcal{B}_0^+$ (respectively, $\mathcal{O}_0^- \cap \mathcal{O}_0^+$) (respectively, $\mathcal{A}_0^- \cap \mathcal{A}_0^+$). By considering two-sided sequences $\mathfrak{q} \in \mathfrak{Q}$ which have different behaviors at each side, one can construct trajectories which belong to $\mathcal{X}_0^- \cap \mathcal{Y}_0^+ \neq \emptyset$ for any prescribed choice $\mathcal{X}, \mathcal{Y} = \mathcal{B}, \mathcal{O}, \mathcal{A}$ such that $\mathcal{X} \neq \mathcal{Y}$. The existence of all these sequences comes from part (b) of Lemma 24, since we can control the size of $|q_j^i|$ just from the size of $|q_1^i|$. □

Corollary 32. *With the notation as in Theorem 31, the following properties are satisfied.*

(a) The return map $F|_{\mathcal{J}}$ has infinite topological entropy.

(b) There is a compact F-invariant set $\mathcal{K} \subset \mathcal{J}$ such that $F|_{\mathcal{K}}$ is topologically semiconjugate to the shift $\sigma : \mathfrak{Q} \to \mathfrak{Q}$ via the map $h|_{\mathcal{K}}$ in the sense of (14)*; it is topologically transitive; and it has sensitive dependence on initial conditions.*

Proof. (a) It follows from the fact that $\sigma : \mathfrak{Q} \to \mathfrak{Q}$ has infinite topological entropy and it is a topological factor of $F : \mathcal{J} \to \mathcal{J}$.

(b) It is a direct consequence of our Theorem 31 and a theorem of Auslander and Yorke. See [51, Item (v) of Theorem 2.1.6] for details. □

Given any integers $1 \leq p < q$, let $\Pi(p,q)$ be the set of (p,q)-periodic billiard trajectories in the circular k-gon Γ. That is, the set of periodic trajectories that close after p turns around Γ and q impacts in Γ, so they have rotation number p/q. The symbol $\#$ denotes the *cardinality* of a set. Let $2^{\mathbb{R}^{n+1}}$ be the power set of $\mathbb{R}^{n+1}$. Let $G_q : 2^{\mathbb{R}^{n+1}} \to \mathbb{N}_0$ be the function

$$G_q(K) = \#\left\{\boldsymbol{x} = (x_1, \ldots, x_{n+1}) \in K \cap \mathbb{Z}^{n+1} : x_1 + \cdots + x_{n+1} = q\right\}$$

that counts the integer points in any subset $K \subset \mathbb{R}^{n+1}$ whose coordinates sum $q \in \mathbb{N}$.

Corollary 33. *Let $\alpha_j^{\pm}$, $\beta_j^{\pm} = \alpha_j^{\pm} + 1$ and χ_j be the quantities defined in Lemma 17 and Corollary 19. If $p, q \in \mathbb{N}$ with $1 \leq p < q$, then*

$$\#\Pi(p,q) \geq 2^{n+1} G_q\big(P^{(p)}\big), \tag{15}$$

where $n+1 = kp$ and

$$P^{(p)} = \left\{ \boldsymbol{x} \in \mathbb{R}^{n+1} : \begin{array}{l} \alpha_j^- x_j + \beta_j^- < x_{j+1} < \alpha_j^+ x_j - \beta_j^+, \quad \forall j = 1, \ldots, n \\ \alpha_{n+1}^- x_{n+1} + \beta_{n+1}^- < x_1 < \alpha_{n+1}^+ x_{n+1} - \beta_{n+1}^+, \\ x_j \geq \chi_j, \quad \forall j = 1, \ldots, n+1 \end{array} \right\} \tag{16}$$

is an unbounded convex polytope of $\mathbb{R}^{n+1}$.

Proof. Let $p, q \in \mathbb{N}$ such that $1 \leq p < q$. Set $n+1 = kp$. Let $\mathfrak{P}_p$ be the set of admissible periodic sequences of period p. We consider the map $\psi_p : \mathfrak{P}_p \to \mathbb{N}^{n+1}$ defined by

$$\psi_p(\mathfrak{q}) = \boldsymbol{x} = (x_1, \ldots, x_{n+1}) = \left(|q_1^0|, \ldots, |q_k^0|, |q_1^1|, \ldots, |q_k^1|, \ldots, |q_1^{p-1}|, \ldots, |q_k^{p-1}|\right),$$

where $\mathfrak{q} = (\boldsymbol{q}^i)_{i \in \mathbb{Z}} \in \mathfrak{P}_p$ and $\boldsymbol{q}^i = (q_1^i, \ldots, q_k^i) \in \boldsymbol{Q}$. Note that $s^p(\mathfrak{q}) = x_1 + \cdots + x_{n+1}$ when $x = \psi_p(\mathfrak{q})$. Besides, $\psi_p(\mathfrak{P}_p) \subset P^{(p)} \cap \mathbb{Z}^{n+1}$ and the map $\psi_p : \mathfrak{P}_p \to P^{(p)} \cap \mathbb{Z}^{n+1}$ is 2^{n+1}-to-1 by construction. Therefore, each point $\boldsymbol{x} \in P^{(p)} \cap \mathbb{Z}^{n+1}$ whose coordinates sum q gives rise to, at least, 2^{n+1} different generic sliding (p,q)-periodic billiard trajectories, see version (**P**) of Proposition 29. □

Lower bound (15) is far from optimal, since it does not take into account the periodic billiard trajectories that are not generic or not sliding. But we think that it captures with great accuracy the growth rate of $\#\Pi(p,q)$ when p/q is relatively small and $q \to +\infty$. It will be the first step in proving Theorem D in Section 8.

7 Optimal linear speed for asymptotic sliding orbits

In this section we establish the existence of uncountably many *points* in the fundamental domain $\mathcal{D}_1$ that give rise to generic asymptotic sliding billiard trajectories (that is, those trajectories in the intersection $\mathcal{A}_0^- \cap \mathcal{A}_0^+ \subset \mathcal{S}_0$ described in the introduction) that approach the boundary asymptotically with optimal uniform linear speed as $|n| \to +\infty$. We also look for trajectories just in $\mathcal{A}_0^+ \subset \mathcal{S}_0$, in which case we obtain uncountably many *horizontal paths* (not points) in $\mathcal{D}_1$. The dynamic feature that distinguishes such trajectories in that they approach the boundary in the fastest way possible among all trajectories that give rise to admissible sequences of symbols.

We believe that the union of all these horizontal paths (respectively, all these points) is a Cantor set times an interval (respectively, the product of two Cantor sets). However, in order to prove it rigorously, we would need to prove that our semiconjugacy $h_{|\mathcal{J}} : \mathcal{J} \to \mathfrak{Q}$, see (14), is, indeed, a full conjugacy. Both sets are F-invariant and they accumulate on the first node of the circular polygon. Obviously, there are similar sets for each one of the other nodes.

The reader must keep in mind the notations listed at the beginning of Section 6, the estimates in Lemma 24, and the interpretation of the partial sums $s^i, s_j^i : \mathfrak{Q} \to \mathbb{N}_0$, with $i \in \mathbb{Z}$ and $j = 1, \ldots, k$, presented in Remark 30.

Definition 34. The uncountably infinite *sign spaces* are

$$\mathfrak{T}^+ = \left\{\mathfrak{t} = (\boldsymbol{t}^i)_{i \geq 0} : \boldsymbol{t}^i = (t_1^i, \ldots, t_k^i) \in \{-,+\}^k\right\},$$
$$\mathfrak{T} = \left\{\mathfrak{t} = (\boldsymbol{t}^i)_{i \in \mathbb{Z}} : \boldsymbol{t}^i = (t_1^i, \ldots, t_k^i) \in \{-,+\}^k\right\}.$$

To avoid any confusion, be aware that the dynamical index of the iterates of asymptotic generic sliding trajectories was called $n \in \mathbb{Z}$ in Theorem C, but it is called $l \in \mathbb{Z}$ in Theorem 35 below.

Theorem 35. *There are constants $0 < d_- < d_+$ such that the following properties hold.*

(a) There are pairwise disjoint paths $\gamma_n^{\mathfrak{t}} \subset \mathcal{K}_{1,n}^{t_1^1} \subset \mathcal{D}_1$ for any $n \geq \chi_1$ and $\mathfrak{t} \in \mathfrak{T}^+$, 'horizontal' since they connect the left side $\mathcal{L}_1$ with the right side $\mathcal{L}_1^1$, such that

$$\left.\begin{array}{c} \Pi_\theta\big(f^l(x)\big) = \Pi_\theta(x), \quad \forall l = 0, \dots, n-1 \\ nd_-\Pi_\theta(x) \leq l\Pi_\theta\big(f^l(x)\big) \leq nd_+\Pi_\theta(x), \quad \forall l \geq n \end{array}\right\} \quad \forall x \in \gamma_n^{\mathfrak{t}},\ \forall n \geq \chi_1,\ \forall \mathfrak{t} \in \mathfrak{T}^+.$$

(b) There are pairwise distinct points $x_n^{\mathfrak{t}} \in \mathcal{K}_{1,n}^{t_1^1} \subset \mathcal{D}_1$ for any $n \geq \chi_1$ and $\mathfrak{t} \in \mathfrak{T}$ such that

$$\left.\begin{array}{c} \Pi_\theta\big(f^l(x_n^{\mathfrak{t}})\big) = \Pi_\theta\big(x_n^{\mathfrak{t}}\big), \quad \forall l = 0, \dots, n-1 \\ \Pi_\theta\big(f^l(x_n^{\mathfrak{t}})\big) = \Pi_\theta\big(f^{-1}(x_n^{\mathfrak{t}})\big), \quad \forall l = -1, \dots, -m \\ nd_-\Pi_\theta(x_n^{\mathfrak{t}}) \leq |l|\Pi_\theta\big(f^l(x_n^{\mathfrak{t}})\big) \leq nd_+\Pi_\theta(x_n^{\mathfrak{t}}), \quad \forall l \geq n \text{ or } l < -m \end{array}\right\} \forall n \geq \chi_1,\ \forall \mathfrak{t} \in \mathfrak{T},$$

where $m = -\xi_k^{-1}(n) \in \mathbb{N}$.

Proof. (a) Identity $\Pi_\theta\big(f^l(x)\big) = \Pi_\theta(x)$ for all $x \in \mathcal{Q}_{1,n}$ and $l = 0, \dots, n-1$ is trivial, because these first impacts are all over the first arc Γ_1, so the angle of reflection remains constant. Henceforth, we just deal with the case $l \geq n$.

Fix $n \geq \chi_1$ and $\mathfrak{t} = (\boldsymbol{t}^i)_{i\geq 0} \in \mathfrak{T}^+$ with $\boldsymbol{t}^i = (t_1^i, \dots, t_k^i)$. Let $\mathfrak{n} = (\boldsymbol{n}^i)_{i\geq 0} \in \big(\mathbb{N}^k\big)^{\mathbb{N}_0}$ with $\boldsymbol{n}^i = (n_1^i, \dots, n_k^i) \in \mathbb{N}^k$ be the sequence given by

$$n_j^i := \xi_j^i(n) = \max \Xi_j^i(n),$$

where $\Xi_j^i(n) \subset \mathbb{N}$ is the set (11). We view $n_1^0 = n$ as the 'starting' value, since the sequence $\mathfrak{n}$ is completely determined by n. However, we do not make this dependence on n explicit for the sake of brevity. Let $\rho^0 = n$,

$$\rho^i = s^i(\mathfrak{n}) = \sum_{m=0}^{i-1} \sum_{j=1}^{k} n_j^m, \quad \forall i > 0,$$

$$\rho_j^i = s_j^i(\mathfrak{n}) = s^i(\mathfrak{n}) + \sum_{m=1}^{j-1} n_m^i, \quad \forall j \mod k,\ \forall i \geq 0.$$

Note that $\rho_1^i = \rho^i$. We use the convention $\rho_{k+1}^i = \rho^{i+1}$.

There is $\mathfrak{q} = (\boldsymbol{q}^i)_{i\geq 0} \in \mathfrak{Q}^+$ with $\boldsymbol{q}^i = (q_1^i, \dots, q_k^i)$ such that $\operatorname{sign}(\mathfrak{q}) = \mathfrak{t}$ and $|\mathfrak{q}| = \mathfrak{n}$ by definition. Note that $s_j^i(\mathfrak{q}) = \rho_j^i$ for any $i \geq 0$ and $j = 1, \dots, k$. Version (**O**) of Proposition 29 implies that there is a path $\gamma_n^{\mathfrak{t}} \in \mathcal{D}_1$, horizontal in the sense that it connects the left side $\mathcal{L}_1$ with the right side $\mathcal{L}_1^1$, such that

$$f^{\rho_j^i}(x) \subset \mathcal{K}_{j,n_j^i}^{t_j^i}, \qquad \forall x \in \gamma_n^{\mathfrak{t}}, \quad \forall i \geq 0, \quad \forall j = 1, \dots, k.$$

In particular, $\gamma_n^{\mathfrak{t}} \subset \mathcal{K}_{1,n}^{t_1^1}$. The paths $\gamma_n^{\mathfrak{t}}$ are pairwise disjoint, because the cells $\mathcal{K}_{j,n}^{\mathfrak{s}}$ are.

Fix $x = (\varphi, \theta) \in \gamma_n^{\mathfrak{t}}$ and $l \geq n$. Set $(\varphi_l, \theta_l) = f^l(\varphi, \theta)$. Our goal is to prove that

$$nd_- \leq l\theta_l/\theta \leq nd_+, \tag{17}$$

for some constants $0 < d_- < d_+$ that do no depend on the choices of the starting value $n \geq \chi_1$, the sign sequence $\mathfrak{t} \in \mathfrak{T}^+$, the point $x \in \gamma_n^{\mathfrak{t}}$ or the forward iterate $l \geq n$.

Let $i \geq 0$ be the number of complete turns around Γ that this billiard trajectory performs from the 0-th impact to the l-th impact, and let $j \in \{1, \dots, k\}$ be the arc index where the l-th impact lands, so $\rho^i \leq \rho_j^i \leq l < \rho_{j+1}^i \leq \rho^{i+1}$. Set $r = \rho_j^i$. Then $(\varphi_r, \theta_r) \in \mathcal{K}_{j,n_j^i}^{t_j^i} \subset \mathcal{Q}_{j,n_j^i}$, and so, since the orbit segment $(\varphi_r, \theta_r), (\varphi_{r+1}, \theta_{r+1}), \dots, (\varphi_{l-1}, \theta_{l-1}), (\varphi_l, \theta_l)$ remains in the circular arc Γ_j without crossing the singularity segment $\mathcal{L}_{j+1}$, we have

$$\frac{\delta_j}{2n_j^i + 2} = \min_{y \in Q_{j,n_j^i}} \Pi_\theta(y) \leq \theta_l = \theta_r \leq \max_{y \in Q_{j,n_j^i}} \Pi_\theta(y) = \frac{\delta_j}{2n_j^i - 2},$$

see Lemma 17. From $x = (\varphi, \theta) \in \gamma_n^{\mathfrak{t}} \subset \mathcal{K}_{1,n}^{t_1^1} \subset \mathcal{Q}_{1,n}$, we also have

$$\frac{\delta_1}{2n+2} = \min_{y \in Q_{1,n}} \Pi_\theta(y) \le \theta \le \max_{y \in Q_{1,n}} \Pi_\theta(y) = \frac{\delta_1}{2n-2}.$$

By combining the last three displayed sets of inequalities, we get that

$$\frac{\delta_j}{\delta_1} \frac{n-1}{n_j^i+1} \rho^i \le l\theta_l/\theta \le \frac{\delta_j}{\delta_1} \frac{n+1}{n_j^i-1} \rho^{i+1}. \tag{18}$$

Let $\nu' < \lambda'$ be the positive constants that appear in Part (aiii) of Lemma 24, so

$$\nu' n_j^i \le \rho^i \le \rho^{i+1} \le \lambda' n_j^i. \tag{19}$$

Bound (17) follows from (18) and (19) if we take

$$\begin{aligned}
d_+ &= \frac{\lambda'}{\delta_1} \max\{\delta_1, \dots, \delta_k\} \max\left\{ \frac{(n+1)n_j^i}{(n_j^i-1)n} : n \ge \chi_1,\ n_j^i \ge \chi_j,\ j = 1, \dots, k \right\} \\
&= \frac{\lambda'}{\delta_1} \max\{\delta_1, \dots, \delta_k\} \max\left\{ \frac{(\chi_1+1)\chi_j}{(\chi_j-1)\chi_1} : j = 1, \dots, k \right\}, \\
d_- &= \frac{\nu'}{\delta_1} \min\{\delta_1, \dots, \delta_k\} \min\left\{ \frac{(n-1)n_j^i}{(n_j^i+1)n} : n \ge \chi_1,\ n_j^i \ge \chi_j,\ j = 1, \dots, k \right\} \\
&= \frac{\nu'}{\delta_1} \min\{\delta_1, \dots, \delta_k\} \min\left\{ \frac{(\chi_1-1)\chi_j}{(\chi_j+1)\chi_1} : j = 1, \dots, k \right\}.
\end{aligned}$$

(b) The proof is similar, but using version (**T**) of Proposition 29. We omit the details. We just stress that if $x \in \mathcal{Q}_{1,n}$, $h(x) = \mathfrak{q} = (\boldsymbol{q}^i)_{i \in \mathbb{Z}}$ and $m := |q_k^{-1}| = -\xi_k^{-1}(n)$, then the first m backward iterates of the point x impact on the last arc Γ_k. □

The constants $0 < a < b$ in Theorem C are directly related to the constants $0 < d_- < d_+$ in Theorem 35. To be precise, we can take

$$\begin{aligned}
a &= \min_{n \ge \chi_1} \frac{1}{n d_+ \max_{x \in \mathcal{Q}_{1,n}} \Pi_\theta(x)} = \min_{n \ge \chi_1} \frac{2n-2}{n d_+ \delta_1} = \frac{2\chi_1 - 2}{\chi_1 \delta_1 d_+} > 0, \\
b &= \max_{n \ge \chi_1} \frac{1}{n d_- \min_{x \in \mathcal{Q}_{1,n}} \Pi_\theta(x)} = \max_{n \ge \chi_1} \frac{2n+2}{n d_- \delta_1} = \frac{2\chi_1 + 2}{\chi_1 \delta_1 d_-} > a.
\end{aligned}$$

The sequences $(\cup_{\mathfrak{t} \in \mathfrak{T}^+} \gamma_n^{\mathfrak{t}})_{n \ge \chi_1}$ and $(\cup_{\mathfrak{t} \in \mathfrak{T}^+} x_n^{\mathfrak{t}})_{n \ge \chi_1}$ are composed by uncountable sets of horizontal paths and points, respectively, with the desired optimal uniform linear speed. The index $n \ge \chi_1$ of the sequence counts the number of impacts that the corresponding billiard trajectories have in the first arc Γ_1 at the beginning. The fundamental quadrilaterals $\mathcal{Q}_{1,n}$ tend to the first node as $n \to +\infty$: $\lim_{n \to +\infty} \mathcal{Q}_{1,n} = (a_1, 0)$, so we conclude that both sequences accumulate on that node when $n \to +\infty$.

Let us justify the optimality of linear speed.

Proposition 36. *There is no billiard trajectory in a circular polygon such that*

$$\lim_{n \to +\infty} n\theta_n = 0.$$

Proof. We have already proved that all asymptotic billiard trajectories that give rise to admissible sequences of symbols satisfy an upper bound of the form

$$1/\theta_n \le b|n|, \qquad \forall |n| \gg 1$$

for some uniform constant $b > 0$ The problem is that there could be some *slightly faster* billiard trajectories that *do not* give rise to admissible sequences.

For instance, if we look at the fundamental quadrilateral $\mathcal{Q}_{j,n}$ displayed in Figure 5 and its image $f^n(Q_{j,n})$ displayed in Figure 6, we see that all points $x \in \mathcal{Q}_{j,n}$ close enough to $\mathcal{L}_{j+1}^{-n+1/2}$ have an image $f^n(x)$ below the lowest admissible fundamental quadrilateral $\mathcal{Q}_{j+1,m}$ with $m = \max\{n' \ge$

$\chi_1 : (n, n') \in \Xi_j\}$. Therefore, since we only deal with admissible sequences of symbols, we have 'lost' the lower non-admissible portion of the red quadrilateral with parabolic shape in Figure 6.

However, part (c) of Lemma 11 shows that, once we fix any $\epsilon \in \big(0, \min\{\mu_1, \dots, \mu_k\}\big)$, we have

$$\Pi_\theta\big(f^n(x)\big) \geq (\mu_j - \epsilon)\Pi_\theta(x), \qquad \forall x \in \mathcal{Q}_{j,n}, \quad \forall j \mod k, \quad \forall n \gg 1,$$

provided $\mu_j < 1$, so these lower non-admissible portions can not be much lower than the ones that we have already taken into account. This means that if we repeat the computations of all constants that appear along our proofs, but replacing μ_j with $\mu_j - \epsilon$ provided $\mu_j < 1$, then we obtain a new uniform constant $\hat{b} \in (b, +\infty)$ such that

$$1/\theta_n \leq \hat{b}|n|, \qquad \forall |n| \gg 1$$

for all billiard trajectories, with no exceptions. □

8 On the number of periodic trajectories

In this section, we construct exponentially large (in q) lower bounds on the number of periodic trajectories of period q, thus proving Theorem D. The strategy of the proof is to use the lower bound (15) provided in Corollary 33. In Section 8.1 we state the main results. Then Section 8.2 contains the proof of a general polynomial lower bound from which we deduce the asymptotic exponential lower bound in Section 8.3.

8.1 Statement of the results

Recall that factors $\alpha_j^\pm$ from the fundamental lemma satisfy hypotheses (**A**) and (**B**) (see Lemma 20). Throughout this section we do not increase the size of χ_j. Indeed, we no longer need the estimates contained in Lemma 24, although we still need those contained in Lemma 17. So, we may consider significantly smaller integers χ_j. For instance, we may take (7) when $\mu_j < 1$. Recall also the unbounded convex polytope $P^{(p)} \subset \mathbb{R}^{n+1}$ introduced in Corollary 33, with $p \in \mathbb{N}$ and $n + 1 = kp$.

Let $\Pi(p, q)$ be the set of (p, q)-periodic billiard trajectories for any $1 \leq p < q$. Let $\Pi(q) = \cup_{1\leq p<q}\Pi(p, q)$ be the set of all periodic trajectories with period q. We state three lower bounds on the number of periodic billiard trajectories in the theorem below. First, a polynomial general lower bound of $\#\Pi(p, q)$. Second, an exponential asymptotic lower bound of $\#\Pi(q)$ as $q \to +\infty$. Third, a polynomial asymptotic lower bound of $\#\Pi(p, q)$ as $q \to +\infty$, for any fixed $p \in \mathbb{N}$. The symbol $\#$ denotes the *cardinality* of a set. The *floor* and *ceiling* functions are denoted with symbols $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$.

Theorem 37. *If Γ is a circular k-gon and $p \in \mathbb{N}$, there are constants $a_\star, b_\star, h_\star, x_\star, M_\star, c_\star(p) > 0$ such that the following three lower bounds hold:*

(a) $\#\Pi(p, q) \geq 2\left(a_\star q/kp - b_\star\right)^{kp-1}/kp$ *for all* $q > b_\star kp/a_\star$.

(b) $\#\Pi(q) \geq \#\Pi(p, q) \geq M_\star \mathrm{e}^{h_\star q}/q$ *when* $p = \lfloor x_\star q/k\rfloor$ *and* $q \to +\infty$.

(c) $\#\Pi(p, q) \geq c_\star q^{kp-1} + \mathrm{O}(q^{kp-2})$ *as* $q \to +\infty$ *for any fixed* $p \in \mathbb{N}$.

Remark 38. We give explicit expressions for all involved constants. We can take

$$\begin{aligned}
a_\star &= 4\min\left\{\frac{(\alpha_1 - \alpha_1^-)A_1}{(1+\alpha_1^-)A}, \frac{(\alpha_1^+ - \alpha_1)A_1}{(1+\alpha_1^+)A}, \dots, \frac{(\alpha_k - \alpha_k^-)A_k}{(1+\alpha_k^-)A}, \frac{(\alpha_k^+ - \alpha_k)A_k}{(1+\alpha_k^+)A}\right\},\\
b_\star &= 6 + 4\max\{\chi_1, \dots, \chi_k\},\\
h_\star &= a_\star W_0(b_\star/\mathrm{e})/b_\star,\\
x_\star &= a_\star W_0(b_\star/\mathrm{e})/((1 + W_0(b_\star/\mathrm{e}))b_\star),\\
M_\star &= 2(a_\star/x_\star - b_\star)^{-k-1}/x_\star,\\
c_\star(p) &= 2(a_\star)^{kp-1}/(kp)^{kp},
\end{aligned}$$

where $\alpha_j = \sqrt{\alpha_j^-\alpha_j^+}$, $A_j = \prod_{i=1}^{j-1}\alpha_i$, $A = \frac{1}{k}\sum_{j=1}^k A_j$ and $W_0 : [-1/\mathrm{e}, +\infty) \to [-1, +\infty)$ is the real part of the principal branch of the Lambert W function. Note that $A_{k+1} = A_1 = 1$ by hypothesis (**B**). Therefore, $A_j = A_{j \mod k}$. Function $W_0(x)$ is implicitly determined by relations $W_0(x\mathrm{e}^x) = x$ for all $x \geq -1$ and $W_0(x)\mathrm{e}^{W_0(x)} = x$ for all $x \geq -1/\mathrm{e}$, see [20].

The exponent $h_\star = a_\star W_0(b_\star/\mathrm{e})/b_\star > 0$ in the exponentially small lower bound is the most important constant in Theorem 37. It is 'proportional' to $a_\star$. We note that there is $i \in \{1,\dots,k\}$ such that $A = \frac{1}{k}\sum_{j=1}^k A_j \geq A_i$, so $a_\star < 4$. The exponent $h_\star$ also depends on $b_\star$ through the Lambert function W_0, but we believe that this is due to the techniques used and does not come from any fundamental characteristic of the problem. It is known that $W_0(x)/x$ is decreasing for $x > 0$, $\lim_{x\to 0^+} W_0(x)/x = W'(0) = 1$ and $W_0(x)/x$ is asymptotic to $\frac{\log x}{x}$ as $x \to +\infty$. Hence, $h_\star < a_\star/\mathrm{e} < 4/\mathrm{e}$ for any Γ. We conclude that the expression $h_\star = a_\star W_0(b_\star/\mathrm{e})/b_\star$ is, by no means, optimal. If Γ tends to a circle, then α_j^- and α_j^+ become closer and closer, so $h_\star$ tends to zero.

The optimal constant $c_\star(p)$ that satisfies the third bound can be way bigger than the crude value $c_\star(p) = 2(a_\star)^{kp-1}/(kp)^{kp}$ obtained directly from the first bound. We give a way to compute the optimal value $c_\star(p) = 2^{kp}\lim_{q\to+\infty} q^{1-kp}G_q\big(P^{(p)}\big)$ in Proposition 39, whose proof is postponed to Appendix B. If P is a Jordan measurable set of $\mathbb{R}^n$, let $\mathrm{V}(P)$ be its *n-dimensional volume*. Let $H_{n+1} = \big\{\boldsymbol{x} \in \mathbb{R}^{n+1} : x_1 + \cdots + x_{n+1} = 1\big\}$. Let $\Pi_{n+1} : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be the projection

$$\boldsymbol{x} = (x_1,\dots,x_{n+1}) \mapsto \tilde{\boldsymbol{x}} = (x_1,\dots,x_n).$$

Projected objects onto $\mathbb{R}^n$ are distinguished with a tilde. Recall that $n+1 = pk$.

Proposition 39. *(a) If Γ is a circular k-gon and $p \in \mathbb{N}$, then*

$$\#\Pi(p,q) \geq 2^{kp}G_q\big(P^{(p)}\big) \geq 2^{kp}V\big(\tilde{K}_\infty^{(p)}\big)q^{kp-1} + \mathrm{O}(q^{kp-2}) \quad \text{as } q \to +\infty,$$

where $\tilde{K}_\infty^{(p)} = \overline{\lim_{q\to+\infty}\tilde{P}_q^{(p)}}$ is the closure of the limit of the bounded convex polytopes

$$\tilde{P}_q^{(p)} = \Pi_{n+1}\big(P_q^{(p)}\big), \qquad P_q^{(p)} = P^{(p)}/q \cap H_{n+1}, \qquad P^{(p)}/q = \{\boldsymbol{x}/q : \boldsymbol{x} \in P^{(p)}\},$$

which are computed by q-contraction, section with hyperplane H_{n+1} and projection by Π_{n+1} of the unbounded convex polytope $P^{(p)}$ defined in (16).

(b) This lower bound is optimal in the sense that

$$\lim_{q\to+\infty} q^{1-kp}G_q\big(P^{(p)}\big) = \mathrm{V}\left(\tilde{K}_\infty^{(p)}\right).$$

(c) The half-space representation of the limit compact convex polytope is

$$\tilde{K}_\infty^{(p)} = \overline{\lim_{q\to+\infty}\tilde{P}_q^{(p)}} = \left\{\tilde{\boldsymbol{x}} \in \mathbb{R}^n : \begin{array}{l} \alpha_j^- x_j \leq x_{j+1} \leq \alpha_j^+ x_j, \quad \forall j = 1,\dots,n-1 \\ \alpha_n^- x_n \leq 1 - \varsigma(\tilde{\boldsymbol{x}}) \leq \alpha_n^+ x_n \\ \alpha_{n+1}^-(1-\varsigma(\tilde{\boldsymbol{x}})) \leq x_1 \leq \alpha_{n+1}^+(1-\varsigma(\tilde{\boldsymbol{x}})) \\ x_j \geq 0, \quad \forall j = 1,\dots,n \\ \varsigma(\tilde{\boldsymbol{x}}) \leq 1 \end{array}\right\}, \tag{20}$$

where $\varsigma(\tilde{\boldsymbol{x}}) = x_1 + \cdots + x_n$.

There exist several algorithms to compute the volume of compact convex polytopes from their half-space representations, so expression (20) can be used to compute $V\big(\tilde{K}_\infty^{(p)}\big)$.

8.2 Proof of the polynomial general lower bound

Recall that $P^{(p)}$ is the unbounded convex polytope (16). We will introduce a cube

$$K = \{\boldsymbol{x} \in \mathbb{R}^{n+1} : |\boldsymbol{x} - \boldsymbol{o}|_\infty \leq t\}, \tag{21}$$

which is the ball centered at the point $\boldsymbol{o} \in \mathbb{R}^{n+1}$ of radius t in the infinity norm $|\cdot|_\infty$. Its center $\boldsymbol{o} = (o_1,\dots,o_{n+1})$ will have three key properties: 1) $\boldsymbol{o} \in P^{(p)}$, 2) $\sum_{j=1}^{n+1} o_j = q$, and 3) $o_j = o_{j \mod k}$. Then, radius t is taken as the largest value such that $K \subset P^{(p)}$. For convenience, we will not make explicit the dependence of K on the integers $1 \leq p < q$.

Lemma 40. *Let $k,n,p,q \in \mathbb{N}$ such that $1 \leq p < q$ and $n+1 = kp$. Recall constants listed in Remark 38. If $\kappa_\star = a_\star/4$, $\tau_\star = \max\{\chi_1,\dots,\chi_k\}$, $0 \leq t < t_\star = \kappa_\star q/(n+1) - \tau_\star$ and*

$$\boldsymbol{o} = (o_1,\dots,o_{n+1}) \in \mathbb{R}^{n+1}, \qquad o_j = \frac{qA_{j \mod k}}{(n+1)A},$$

then $o_1 + \cdots + o_{n+1} = q$ and $K = \{\boldsymbol{x} \in \mathbb{R}^{n+1} : |\boldsymbol{x} - \boldsymbol{o}|_\infty \leq t\} \subset P^{(p)}$.

Proof. Clearly, $o_1 + \cdots + o_{n+1} = \frac{q}{(n+1)A}\sum_{j=1}^{n+1} A_j = \frac{qp}{(n+1)A}\sum_{j=1}^{k} A_j = \frac{kp}{n+1}q = q$.

If $\boldsymbol{x} \in K$, then $\boldsymbol{x} = \boldsymbol{o} + t\boldsymbol{u}$ for some $\boldsymbol{u} \in \mathbb{R}^{n+1}$ such that $|\boldsymbol{u}|_\infty \le 1$. With the suitable choice of the radius t, the point $\boldsymbol{x}$ satisfies the following three sets of inequalities that define the unbounded convex polytope $P^{(p)}$ given in (16):

- *First set (with $2n$ inequalities).* Since $o_{j+1} = \alpha_j o_j$ for all $j = 1, \dots, n$, we see that

$$\alpha_j^- x_j + \beta_j^- < x_{j+1} < \alpha_j^+ x_j - \beta_j^+ \Leftrightarrow \begin{cases} (\alpha_j^- u_j - u_{j+1})t < (\alpha_j - \alpha_j^-)o_j - \beta_j^- \\ (u_{j+1} - \alpha_j^+ u_j)t < (\alpha_j^+ - \alpha_j)o_j - \beta_j^+ \end{cases}$$

- *Second set (with 2 inequalities).* Since $A_1 = 1$ and $A_{n+1} = A_k = \prod_{j=1}^{k-1}\alpha_j = 1/\alpha_k$, we get that $o_1 = \alpha_k o_{n+1} = \alpha_k o_k$. Besides, $\beta_{n+1}^\pm = \beta_k^\pm$ and $\alpha_{n+1}^\pm = \alpha_k^\pm$. Hence,

$$\alpha_{n+1}^- x_{n+1} + \beta_{n+1}^- < x_1 < \alpha_{n+1}^+ x_{n+1} - \beta_{n+1}^+ \Leftrightarrow \begin{cases} (\alpha_k^- u_{n+1} - u_1)t < (\alpha_k - \alpha_k^-)o_k - \beta_k^- \\ (u_1 - \alpha_k^+ u_{n+1})t < (\alpha_k^+ - \alpha_k)o_k - \beta_k^+ \end{cases}$$

- *Third set (with $n+1$ inequalities).* $x_j \ge \chi_j \Leftrightarrow -u_j t \le o_j - \chi_j$.

Let us analyse the RHS and LHS of the above $3n+3$ inequalities. Coordinates o_j can be as big as needed if we take $q/(n+1) \gg 1$, because quotients $A_{j \mod k}/A$ do not depend on p, q or n. Thus, using that $\alpha_j^- < \alpha_j < \alpha_j^+$ for all $j = 1, \dots, k$, all RHS can be made positive if we take $q/(n+1) \gg 1$. On the other hand, we can bound the LHS as follows:

$$(\alpha_j^- u_j - u_{j+1})t \le (1 + \alpha_j^-)t, \qquad (u_{j+1} - \alpha_j^+ u_j)t \le (1 + \alpha_j^+)t, \qquad -u_j t \le t,$$

because $|u_j| \le |\boldsymbol{u}|_\infty \le 1$ for all $j = 1, \dots, n+1$ and $t \ge 0$. Therefore, these $3n+3$ inequalities hold when we take any $t \in [0, t_\star)$ with

$$\begin{aligned} t_\star &= \min\left\{ \frac{(\alpha_j - \alpha_j^-)o_j - \beta_j^-}{1 + \alpha_j^-}, \frac{(\alpha_j^+ - \alpha_j)o_j - \beta_j^+}{1 + \alpha_j^+}, o_j - \chi_j : j = 1, \dots, n+1 \right\} \\ &= \min\left\{ \kappa_j^- q/(n+1) - \tau_j^-, \kappa_j^+ q/(n+1) - \tau_j^+, \kappa_j q/(n+1) - \tau_j : j = 1, \dots, k \right\}, \end{aligned}$$

where

$$\kappa_j^\pm = \frac{|\alpha_j - \alpha_j^\pm| A_j}{(1 + \alpha_j^\pm)A}, \qquad \kappa_j = \frac{A_j}{A}, \qquad \tau_j^\pm = \frac{\beta_j^\pm}{1 + \alpha_j^\pm} = 1, \qquad \tau_j = \chi_j.$$

All these arguments imply that $K \subset P^{(p)}$ provided that $0 \le t < t_\star := \kappa_\star q/(n+1) - \tau_\star$, where

$$\kappa_\star = \min\left\{\kappa_1^-, \kappa_1^+, \kappa_1, \dots, \kappa_k^-, \kappa_k^+, \kappa_k\right\} > 0, \quad \tau_\star = \max\left\{\tau_1^-, \tau_1^+, \tau_1, \dots, \tau_k^-, \tau_k^+, \tau_k\right\} > 0.$$

These constants $\kappa_\star$ and $\tau_\star$ do not depend on p, q or n.

We note that $(\alpha_j^+ - \alpha_j)/(1 + \alpha_j^+) < 1$. Hence, $\kappa_j^+ < \kappa_j$ and we can take

$$\kappa_\star = \min\left\{\kappa_1^-, \kappa_1^+, \dots, \kappa_k^-, \kappa_k^+\right\} = a_\star/4,$$

see Remark 38. □

We look for a lower bound on $G_q(K)$, where K is a cube of the form (21) such that $\sum_{j=1}^{n+1} o_j = q$. Note that $G_q\big(\{0,1\}^n\big) = \binom{n}{q}$, so $G_q\big(\{0,1\}^{2q}\big) = \binom{2q}{q} \ge 4^q/(2q+1)$ grows exponentially fast as $q \to +\infty$. We want to generalise this idea. Since there is no standard notation for the generalised binomial coefficients that we need —for instance, symbols $\binom{n,m}{q}$ and $\binom{n}{q}^{(m)}$ can be found in [47, 42]—, we use our own notation. Set

$$[0..m] := \mathbb{Z} \cap [0, m] = \{0, 1, \dots, m-1, m\}.$$

Then $G_q\big([0..m]^n\big)$ counts the number of *weak compositions* of q into n parts with no part exceeding m. Note that $G_q\big([0..m]^n\big) = 0$ for any $q \notin [0..nm]$. It is well know [25, section I.3] that

$$\sum_{q=0}^{\infty} G_q\big([0..m]^n\big) x^q = (1 + x + x^2 + \cdots + x^m)^n.$$

Using this polynomial identity, Andrews [2] deduced that, once $m, n \in \mathbb{N}$ are fixed, the sequence $G_q\big([0..m]^n\big)$ is unimodal in q and reaches its maximum at $q = \lfloor nm/2 \rfloor$.

Lemma 41. $G_{\lfloor nm/2\rfloor}\big([0..m]^n\big) \geq \frac{(m+1)^n}{nm+1} \geq \frac{(m+1)^{n-1}}{n}$ *for all* $m, n \in \mathbb{N}$.

Proof. It follows from $\#[0..nm] = nm+1$, $\sum_{q=0}^{nm} G_q\big([0..m]^n\big) = \#\left([0..m]^n\right) = (m+1)^n$, and inequalities $G_q\big([0..m]^n\big) \leq G_{\lfloor nm/2\rfloor}\big([0..m]^n\big)$ for all $q \in [0..mn]$. □

Now we are ready to establish the lower bound on $G_q(K)$ that we are looking for.

Lemma 42. *Let* $n, q \in \mathbb{N}$ *and* $t > 0$. *If* K *is a cube of the form (21) such that* $\sum_{j=1}^{n+1} o_j = q$ *and* $t \geq 3/2$, *then*

$$G_q(K) \geq \frac{(2t-3)^n}{n+1}. \tag{22}$$

Proof. There exists an integer point $\boldsymbol{o}' \in \mathbb{Z}^{n+1}$ such that $|\boldsymbol{o} - \boldsymbol{o}'|_\infty \leq 1$ and $\sum_{j=1}^{n+1} o'_j = q$. If $\boldsymbol{o} \in \mathbb{Z}^{n+1}$, we take $\boldsymbol{o}' = \boldsymbol{o}$. If $\boldsymbol{o} \notin \mathbb{Z}^{n+1}$, we can take, for instance,

$$o'_j = \begin{cases} \lfloor o_j \rfloor + 1, & \text{for } j \leq i, \\ \lfloor o_j \rfloor, & \text{otherwise,} \end{cases}$$

where $i = q - \sum_{j=1}^{n+1} \lfloor o_j \rfloor \in [1..n]$, so that $\sum_{j=1}^{n+1} o'_j = i + \sum_{j=1}^{n+1} \lfloor o_j \rfloor = q$.

Set $m = \lfloor t \rfloor - 1 \in \mathbb{N} \cup \{0\}$ and $v_j = o'_j - m$. Clearly, $[v_j, v_j + 2m] \subset [o_j - t, o_j + t]$. Hence, given any $\boldsymbol{y} \in [0..2m]^{n+1}$ such that $\sum_{j=1}^{n+1} y_j = (n+1)m$, the sum of the components of the vector $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{v} \in K \cap \mathbb{Z}^{n+1}$ is equal to

$$\sum_{j=1}^{n+1} x_j = (n+1)m + \left(\sum_{j=1}^{n+1} o'_j\right) - (n+1)m = q.$$

Besides, the correspondence $[0..2m]^{n+1} \ni \boldsymbol{y} \mapsto \boldsymbol{x} = \boldsymbol{y} + \boldsymbol{v} \in K \cap \mathbb{Z}^{n+1}$ is injective, which implies that

$$G_q(K) \geq G_{(n+1)m}\big([0..2m]^{n+1}\big) = G_{\lfloor (n+1)2m/2\rfloor}\big([0..2m]^{n+1}\big) \geq \frac{(2m+1)^n}{n+1} \geq \frac{(2t-3)^n}{n+1}.$$

We have used Lemma 41 and $m = \lfloor t \rfloor - 1 \geq t - 2$ in the last two inequalities. □

To end, we prove the first lower bound stated in Theorem 37.

Proof of the polynomial general lower bound. This bound follows from bound (15), the inclusion $K \subset P^{(p)}$, bound (22), condition $0 \leq t < t_\star := \kappa_\star q/(n+1) - \tau_\star$ required in Lemma 40, and the identities $a_\star = 4\kappa_\star$, $b_\star = 4\tau_\star + 6$ and $n + 1 = kp$. Namely,

$$\begin{aligned} \#\Pi(p,q) \geq 2^{n+1} G_q(P^{(p)}) &\geq \max_{t \in [3/2, t_\star)} \left\{2^{n+1} G_q(K)\right\} \geq \max_{t \in [3/2, t_\star)} \frac{2(4t-6)^n}{n+1} = \frac{2(4t_\star - 6)^n}{n+1} \\ &= \frac{2}{n+1}\left(\frac{4\kappa_\star q}{n+1} - 4\tau_\star - 6\right)^n = \frac{2}{n+1}\left(\frac{a_\star q}{n+1} - b_\star\right)^n = \frac{2}{kp}\left(\frac{a_\star q}{kp} - b_\star\right)^{kp-1}. \end{aligned}$$

Note that $[3/2, t_\star) \neq \emptyset$ since $q > b_\star kp/a_\star$ implies that $t_\star = \kappa_\star q/(n+1) - \tau_\star > 3/2$. □

8.3 Proof of the two asymptotic lower bounds

We describe the exponentially fast growth of $\#\Pi(p,q)$ when $p = \lfloor xq/k \rfloor$ and $q \to +\infty$ for some fixed limit ratio $x > 0$. We shall also determine the limit ratio $x_\star > 0$ that gives the largest exponent $h_\star$ in the exponential bound.

Lemma 43. *Let* $0 < a < b$ *and* $k \in \mathbb{N}$. *If*

- $M(x) = 2(a/x - b)^{-k-1}/x$ *for* $0 < x < a/b$;
- $h(x) = x \log(a/x - b)$ *for* $0 < x < a/b$; *and*
- $G(p,q) = 2(aq/kp - b)^{kp-1}/kp$ *for* $p, q \in \mathbb{N}$ *such that* $0 < kp/q < a/b$,

then

$$G(\lfloor xq/k\rfloor, q) \geq M(x)\mathrm{e}^{h(x)q}/q, \qquad \forall q \geq (1+b)kp/a,\ \forall x \in \big(0, a/(b+1)\big].$$

The exponent $h : (0, a/b) \to \mathbb{R}$ *reaches its maximum value* $h_\star = h(x_\star) = aW_0(b/\mathrm{e})/b > 0$ *at the point* $x_\star = aW_0(b/\mathrm{e})/((1+W_0(b/\mathrm{e}))b)$.

Proof. If $q \in \mathbb{N}$, $x \in (0, a/(b+1)]$ and $p = \lfloor xq/k\rfloor$, then $aq/kp - b \geq 1$, $xq - k < kp \leq xq$ and

$$G(p,q) = \frac{2}{kp}\left(\frac{aq}{kp} - b\right)^{kp-1} \geq \frac{2}{kp}\left(\frac{aq}{kp} - b\right)^{xq-k-1} \geq \frac{2}{xq}\left(\frac{a}{x} - b\right)^{xq-k-1} = \frac{1}{q}M(x)\mathrm{e}^{h(x)q}.$$

Next, we look for the global maximum of $h(x)$. After the changes of variable

$$(0,+\infty) \ni \hat{x} \leftrightarrow x = \frac{a}{b+\mathrm{e}^{\hat{x}}} \in \big(0, a/(b+1)\big),$$
$$(0,+\infty) \ni \hat{h} \leftrightarrow h = a\hat{h} \in (0,+\infty),$$

we get that $h(x) = x\log(a/x - b) = x\hat{x} = a\hat{x}/(b+\mathrm{e}^{\hat{x}})$, so $\hat{h}(\hat{x}) = \hat{x}/(b+\mathrm{e}^{\hat{x}})$. We have reduced the search of the global maximum point $x_\star \in (0, a/(b+1))$ of $h(x)$ to the search of the global maximum point $\hat{x}_\star > 0$ of $\hat{h}(\hat{x})$. Since

$$\frac{\mathrm{d}\hat{h}}{\mathrm{d}\hat{x}}(\hat{x}) = \frac{b + (1-\hat{x})\mathrm{e}^{\hat{x}}}{(b+\mathrm{e}^{\hat{x}})^2} = 0 \Leftrightarrow (\hat{x}-1)\mathrm{e}^{\hat{x}-1} = b/\mathrm{e} \Leftrightarrow \hat{x} = \hat{x}_\star := 1 + W_0(b/\mathrm{e}),$$

we deduce that $\hat{h}(\hat{x})$ reaches its maximum value

$$\hat{h}_\star = \hat{h}(\hat{x}_\star) = \frac{\hat{x}_\star}{b+\mathrm{e}^{\hat{x}_\star}} = \frac{1}{\mathrm{e}^{\hat{x}_\star}} = \frac{1}{\mathrm{e}}\frac{1}{\mathrm{e}^{W_0(b/\mathrm{e})}} = \frac{W_0(b/\mathrm{e})}{b}$$

at the point $\hat{x} = \hat{x}_\star$. In order to compute $\hat{x}_\star$, we have used that $W_0(b/\mathrm{e})\mathrm{e}^{W_0(b/\mathrm{e})} = b/\mathrm{e}$. Expressions for $x_\star$ and $h_\star$ are obtained by undoing both changes of variable. □

We can prove now the second and third lower bounds stated in Theorem 37.

Proof of both asymptotic lower bounds. The second bound of Theorem 37 follows from the first one by applying Lemma 43 with $a = a_\star$ and $b = b_\star$. Analogously, the third bound of Theorem 37 follows from the first one by taking $c_\star(p) = 2(a_\star)^{kp-1}/(kp)^{kp}$. □

9 The length spectrum of circular polygons

The purpose of this section is to prove Theorem E, which shows an unusual feature of the length spectrum of billiards in circular polygons. In particular, it shows that the well-known results of Marvizi-Melrose [44] fail to hold for circular polygons. This was expected because there are so many periodic billiard trajectories inside circular polygons —as we have seen in the previous section— that we can construct sequences of them whose lengths have rather different asymptotic behaviors.

Let $|\Gamma|$ be the length of Γ. Let $\kappa(s)$ be the curvature of Γ as a function of an arc-length parameter $s \in [0, |\Gamma|)$. If $z, z' \in \Gamma$ are any two consecutive impact points of a billiard trajectory g, then the segment $[z, z'] \subset \mathbb{R}^2$ is a *link* of g and $\int_z^{z'} \mathrm{d}s$ is the distance from z to z' along Γ. Note that $|z' - z| < \int_z^{z'} \mathrm{d}s$ by convexity. If $g = \{z_0, \ldots, z_{q-1}\} \subset \Gamma$ is a q-periodic billiard trajectory, let $L(g) = |z_1 - z_0| + \cdots + |z_{q-1} - z_0|$ be its *length*. Let

$$\underline{L}_q = \inf\{L(g) : g \in \Pi(1,q)\}, \qquad \overline{L}_q = \sup\{L(g) : g \in \Pi(1,q)\}.$$

To begin with, let us recall the Marvizi-Melrose results for smooth ovals. A *smooth oval* is a regular, simple, closed, oriented C^∞ curve with positive curvature everywhere.

Theorem 44 (Marvizi & Melrose [44])**.** *Let* Γ *be any smooth oval.*

(a) $\lim_{q\to+\infty} q^i\big(\overline{L}_q - \underline{L}_q\big) = 0$ *for all* $i \in \mathbb{N}$.

(b) There are asymptotic coefficients $c_i \in \mathbb{R}$ *such that if* $g_q \in \Pi(1,q)$*, then*

$$L(g_q) \asymp |\Gamma| + \sum_{i=1}^{\infty} \frac{c_i}{q^{2i}}, \quad \text{as } q \to +\infty.$$

(c) $c_1 = -\frac{1}{24}\left[\int_\Gamma \kappa^{2/3}(s)\,\mathrm{d}s\right]^3 < 0$.

(d) If $[z, z']$ *is a link of* $g_q \in \Pi(1, q)$*, then*

$$\int_z^{z'} \mathrm{d}s \asymp \frac{1}{q}|\Gamma|, \quad \textit{uniformly as } q \to +\infty.$$

The symbol $\asymp$ means that the RHS is asymptotic to the LHS. The first property implies that the Melrose-Marvizi asymptotic coefficients c_i do not depend on the choice of the sequence of periodic trajectories $(g_p)_q$. All of them can be explicitly written in terms of the curvature. For instance, the formulas for c_1, c_2, c_3, and c_4 can be found in [55]. Property (d) means that not only are the lengths of g_q asymptotically well-behaved, but as $q \to +\infty$, the distribution of the points in g_q is asymptotically well-behaved with respect to any one point. Hence, property (d) is like a weak local version of property (b). There is also a strong local version in [44, Theorem 5.9].

From now on, with Γ denoting a circular polygon, we use the notation introduced in previous sections, although we are now only interested in $(1, q)$-periodic trajectories, so $1 \le j \le k$ and we no longer need to consider j modulo k. Recall that $p = 1$ along this section.

We will check that none of properties (a)–(d) of Theorem 44 hold for circular polygons.

First, we consider the simplest periodic sliding trajectories in a circular polygon, wich are the trajectories that impact *all its nodes* in such a way that the angle of reflection remains *constant* along the whole trajectory. These *nodal* sliding periodic trajectories can only take place in certain circular polygons, which we call *rational*.

Definition 45. We say that a circular polygon Γ is *rational* when all its central angles are rational multiples of π, so

$$\delta_j = m_j \delta,$$

for some $\delta = \gcd(\delta_1, \ldots, \delta_k)$ and $m_j = \delta_j/\delta \in \mathbb{N}$. Set $M = \sum_{j=1}^k m_j$. Then $M\delta = 2\pi$. A billiard trajectory inside a rational circular polygon is *nodal* when it impacts all nodes (interspersed with possibly many other non-nodal impacts) in the counter-clockwise ordering.

Squared pseudo-ellipses and Moss's eggs are rational circular polygons, see Section 2. Any nodal orbit in a rational circular polygon has constant angle of reflection, is sliding, and is periodic with a rotation number of the form $1/q$, q being the period.

Nodal billiard trajectories give the simplest examples of sequences of sliding periodic billiard trajectories in circular polygons where properties (c) and (d) of Theorem 44 fail, because it is really easy to compute their lengths.

Proposition 46. *Let* Γ *be a rational circular* k*-gon with arcs* Γ_j*, radii* r_j *and central angles* δ_j*. Set* $\delta = \gcd(\delta_1, \ldots, \delta_k)$*. Fix some* $\psi = \delta/2i$ *with* $i \in \mathbb{N}$*. Let* g_q *be the billiard trajectory generated by*

$$(\varphi_n, \theta_n) = f^n(a_1, \psi), \qquad \forall n \in \mathbb{Z}.$$

(a) The billiard trajectory g_q *is nodal and* $g_q \in \Pi(1, q)$ *with period* $q = Mi$*.*

(b) $L(g_q) = |\Gamma| - \pi^2|\Gamma|/6q^2 + \mathrm{O}(1/q^4)$ *as* $q = Mi \to +\infty$*.*

(c) If $[z, z']$ *is a circular link of* g_q *associated to the arc* Γ_j*, then*

$$\int_z^{z'} \mathrm{d}s = \frac{1}{q}2\pi r_j \neq \frac{1}{q}|\Gamma|.$$

Proof. (a) Once we fix the index $i \in \mathbb{N}$, we deduce that $\varphi_{im_1} = a_1 + 2im_1\psi = a_1 + \delta_1 = b_1 = a_2$. That is, the first im_1 links of the billiard trajectory connect both nodes of the first arc. In particular, the angle of reflection does not change when we enter the second arc, so the next im_2 links connect both its nodes, and so on. This means that the orbit is nodal and periodic with rotation number $1/q$ and period $q = Mi$.

(b) The length of each link in the j-th arc is equal to $\ell_j = 2r_j \sin\psi$, so

$$\begin{aligned} L(g_q) &= \sum_{j=1}^k im_j\ell_j = \sum_{j=1}^k 2im_jr_j\left[\psi - \frac{1}{6}\psi^3 + \mathrm{O}\left(\psi^5\right)\right] \\ &= \sum_{j=1}^k m_j\delta r_j - \frac{\delta^2}{24}\left[\sum_{j=1}^k m_j\delta r_j\right]\frac{1}{i^2} + \mathrm{O}\left(i^{-4}\right) = |\Gamma| - \frac{\pi^2|\Gamma|}{6q^2} + \mathrm{O}(1/q^4), \end{aligned}$$

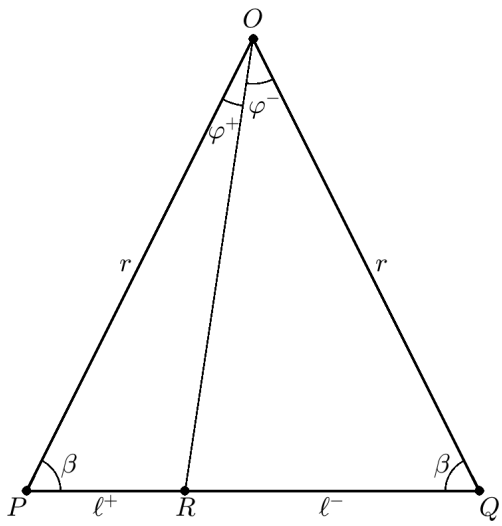

Figure 7: Isosceles triangle ΔOPQ and a point $R \in [P,Q] \setminus \{P,Q\}$ such that $\angle QPO = \angle OQP = \beta$, $|\vec{OP}| = |\vec{OQ}| = r$, $\angle POQ = 2\psi = \varphi^+ + \varphi^- = \angle POR + \angle ROQ$, and $|\vec{PQ}| = \ell = \ell^+ + \ell^- = |\vec{PR}| + |\vec{RQ}|$.

where we have used that $\delta_j = m_j\delta$, $|\Gamma| = \sum_{j=1}^k \delta_j r_j$, $q = Mi$, and $M\delta = 2\pi$.

(c) If two consecutive impact points z and z' belong to the arc Γ_j, then

$$\int_z^{z'} \mathrm{d}s = 2\psi r_j = \frac{1}{i}\delta r_j = \frac{1}{q}M\delta r_j = \frac{1}{q}2\pi r_j \neq \frac{1}{q}|\Gamma|. \qquad \square$$

The previous proposition has been obtained without the heavy machinery developed in this paper, but it needs a rather special type of circular polygons. Next, we deal with general circular polygons, where the computations are more involved.

Remark 47. Corollary 33 implies that there are at least 2^k generic sliding periodic billiard trajectories $g_q \in \Pi(1,q)$ with *exactly* $x_j \in \mathbb{N}$ impacts in the arc Γ_j, $j = 1,\dots,k$, for any integer point $\boldsymbol{x} = (x_1,\dots,x_k) \in P^{(1)} \cap \mathbb{Z}^k = P^{(1)} \cap \mathbb{N}^k$ such that $x_1 + \cdots + x_k = q$. Here, $P^{(1)}$ is the unbounded convex polytope of $\mathbb{R}^k$ defined in (16) for $p = 1$.

We need a couple of technical results before tackling the proof of Theorem E.

First, we compute the lengths of generic sliding $(1,q)$-periodic billiard trajectories. By definition, they impact all arcs but no nodes. Angles similar to $\varphi_j^\pm$ and ψ_j below were considered in the proof of part (c) in Lemma 11.

Lemma 48. *If $g_q \in \Pi(1,q)$ is a generic sliding periodic billiard trajectory inside Γ, then*

$$L(g_q) = \sum_{j=1}^k \left(\ell_j^- + (x_j - 1)\ell_j + \ell_j^+\right), \qquad \ell_j = 2r_j \sin\psi_j, \qquad \ell_j^\pm = \frac{r_j \sin\varphi_j^\pm}{\cos(\psi_j - \varphi_j^\pm)}, \tag{23}$$

where

(i) $x_j \in \mathbb{N}$ is the number of impact points in Γ_j;

(ii) $\psi_j > 0$ is the constant angle of reflection along the x_j impacts in Γ_j; and

(iii) $\varphi_j^\pm \in (0, 2\psi_j)$ are the impact angles such that $[z(b_j - \varphi_j^+), z(a_{j+1} + \varphi_{j+1}^-)]$ is the transition link connecting Γ_j and Γ_{j+1}.

Besides, $\varphi_j^- + 2(x_j - 1)\psi_j + \varphi_j^+ = \delta_j$ for all $j = 1,\dots,k$.

Proof. If we apply the law of sinus to the three triangles ΔPOQ, ΔPOR and ΔROQ displayed in Figure 7 (see its caption for the definition of each quantity), we get that

$$\frac{\ell}{2\sin\psi\cos\psi} = \frac{\ell}{\sin 2\psi} = \frac{r}{\sin\beta} = \frac{r}{\cos\psi}, \qquad \frac{\ell^\pm}{\sin\varphi^\pm} = \frac{r}{\sin(\pi - \beta - \varphi^\pm)} = \frac{r}{\cos(\psi - \varphi^\pm)},$$

since $\beta = \pi/2 - \psi$, $\angle ORP = \pi - \beta - \varphi^+$ and $\angle QRO = \pi - \beta - \varphi^-$. Therefore,

$$\ell = 2r\sin\psi, \qquad \ell^{\pm} = \frac{r\sin\varphi^{\pm}}{\cos(\psi - \varphi^{\pm})}.$$

We deduce (23) from those three formulas.

If g_q has x_j impacts in Γ_j with constant angle of reflection ψ_j, then it has $x_j - 1$ circular links with a certain constant length ℓ_j. Each one of these circular links $[P, Q]$ is the base of an isosceles triangle ΔOPQ like the one displayed in Figure 7, with $O = O_j$, $r = r_j$ and $\psi = \psi_j$. Hence $\ell_j = 2r_j \sin\psi_j$.

Let us consider the transition link $[z(b_j - \varphi_j^+), z(a_{j+1} + \varphi_{j+1}^-)]$ connecting Γ_j and Γ_{j+1} and the isosceles triangle ΔOPQ with

$$O = O_j, \qquad P = z(b_j - \varphi_j^+) \in \Gamma_j, \qquad Q = O_j + r_j \mathrm{e}^{\mathrm{i}(b_j - \varphi_j^+ + 2\psi_j)}.$$

We stress that Q is an auxiliary point: $Q \notin \Gamma$. Let $R = [P, Q] \cap [O, z(b_j)]$. Then $r = r_j$, $\varphi^+ = \varphi_j^+$ and $\ell^+ = \ell_j^+$. Therefore, $\ell_j^+ = r_j \sin\varphi_j^+ / \cos(\psi_j - \varphi_j^+)$. The formula for ℓ_{j+1}^- is deduced in a similar way, but taking

$$O = O_{j+1}, \qquad P = O_{j+1} + r_{j+1}\mathrm{e}^{\mathrm{i}(a_{j+1} - \varphi_{j+1}^- - 2\psi_{j+1})}, \qquad Q = z(a_{j+1} + \varphi_{j+1}^-) \in \Gamma_{j+1},$$

$R = [P, Q] \cap [O, z(a_{j+1})]$, $r = r_{j+1}$, $\varphi^- = \varphi_{j+1}^-$ and $\ell^- = \ell_{j+1}^-$. (In this case, the auxiliary point is P: $P \notin \Gamma$). By construction, the transition link $[z(b_j - \varphi_j^+), z(a_{j+1} + \varphi_{j+1}^-)]$ has length $\ell_j^+ + \ell_{j+1}^+$. This proves (23).

Finally, relation $\varphi_j^- + 2(x_j - 1)\psi_j + \varphi_j^+ = \delta_j$ is geometrically evident. □

Next, we need a technical result about the extreme values of the differentiable strictly concave function (25) over a bounded convex polytope $P_\infty^{(1)}$ related to the unbounded convex polytope $P^{(1)}$ of $\mathbb{R}^k$ defined in (16) for $p = 1$. Recall that $\Delta_{k-1} = \{\boldsymbol{x} \in \mathbb{R}^k : \boldsymbol{x} > 0,\ x_1 + \cdots + x_k = 1\}$ is the open $(k-1)$-simplex and $H_k = \{\boldsymbol{x} \in \mathbb{R}^k : x_1 + \cdots + x_k = 1\}$.

Lemma 49. *The bounded convex polytope $P_\infty^{(1)} = \lim_{q\to\infty}\big(\{\boldsymbol{x}/q : \boldsymbol{x} \in P^{(1)}\} \cap H_k\big)$ is given by*

$$P_\infty^{(1)} = \left\{ \boldsymbol{x} \in \mathbb{R}^k : \begin{array}{ll} \alpha_j^- x_j < x_{j+1} < \alpha_j^+ x_j, & \forall j = 1, \ldots, k-1 \\ \alpha_k^- x_k < x_1 < \alpha_k^+ x_k & \\ x_j > 0, \quad \forall j = 1, \ldots, k & \\ x_1 + \cdots + x_k = 1 & \end{array} \right\} \tag{24}$$

and its compact closure $K_\infty^{(1)}$ is contained in the open simplex Δ_{k-1}. Let

$$h : \Delta_{k-1} \to (-\infty, 0), \qquad h(\boldsymbol{y}) = -\frac{1}{24}\sum_{j=1}^{k} \frac{\delta_j^3 r_j}{y_j^2}. \tag{25}$$

Set $I_1 = h\big(P_\infty^{(1)}\big)$, $c_1^- = \inf I_1$ and $c_1^+ = \sup I_1$. Then $c_1^+ \in I_1$ and

$$-\infty < c_1^- \le -\pi^2|\Gamma|/6 < c_1^+ = \frac{1}{24}\left[\int_\Gamma \kappa^{2/3}(s)\,\mathrm{d}s\right]^3 < 0. \tag{26}$$

Proof. Expression (24) is trivial. We check that $K_\infty^{(1)} \subset \Delta_{k-1}$ by a reductio ad absurdum argument. Let us assume that $\boldsymbol{x} = (x_1, \ldots, x_k) \in K_\infty^{(1)}$ and $x_i = 0$ for some i. Then inequalities $\alpha_j^- x_j \le x_{j+1} \le \alpha_j^+ x_j$ for $j = 1, \ldots, k-1$ and $\alpha_k^- x_k \le x_1 \le \alpha_k^+ x_k$ imply that

$$x_{i+1} = \cdots = x_k = x_1 = \cdots = x_{i-1} = 0,$$

so identity $x_1 + \cdots + x_k = 1$ fails. Contradiction.

The image of a compact convex set by a continuous function that only takes negative values is a compact interval of $(-\infty, 0)$, so $\overline{I_1} = h\big(K_\infty^{(1)}\big) = [c_1^-, c_1^+]$ for some numbers $-\infty < c_1^- \le c_1^+ < 0$. Let us estimate the minimum value c_1^-, compute exactly the maximum value c_1^+, prove that $c_1^- < c_1^+$, and check that $c_1^+ \in I_1$.

We claim that function (25) attains its maximum value only at $\boldsymbol{y} = \boldsymbol{w}(1/3)$, where

$$\boldsymbol{w}(\xi) = \frac{1}{S(\xi)}\big(s_1(\xi), \dots, s_k(\xi)\big) \in \Delta_{k-1}, \qquad s_j(\xi) = \delta_j r_j^{\xi}, \qquad S(\xi) = \sum_{j=1}^{k} \delta_j r_j^{\xi}$$

for all $\xi \in \mathbb{R}$. On the one hand, the gradient of $\mathbb{R}_+^k \ni \boldsymbol{y} \mapsto \sum_{j=1}^k \delta_j^3 r_j / y_j^2$ is the vector with components $-2\delta_j^3 r_j / y_j^3$, so $\boldsymbol{y} \in \Delta_{k-1}$ is a critical point of (25) if and only if

$$\big(s_i(1/3)/y_i\big)^3 = \delta_i^3 r_i / y_i^3 = \delta_j^3 r_j / y_j^3 = \big(s_j(1/3)/y_j\big)^3, \qquad \forall i \neq j.$$

This means that $\boldsymbol{w}(1/3)$ is the only critical point of (25). On the other hand, $\sum_{j=1}^k \delta_j^3 r_j / y_j^2$ is a nonnegative weighted sum of convex terms $1/y_j^2$, so $-\frac{1}{24}\sum_{j=1}^k \delta_j^3 r_j / y_j^2$ is a strictly concave function on $\mathbb{R}_+^k$ and (25) is a differentiable strictly concave function. Hence, the local maximum $\boldsymbol{w}(1/3)$ is a strict global maximum. This proves the claim. Furthermore,

$$h(\boldsymbol{w}(\xi)) = -\frac{1}{24}\sum_{j=1}^{k} \frac{\delta_j^3 r_j}{\big(s_j(\xi)/S(\xi)\big)^2} = -\frac{1}{24} S(\xi)^2 \sum_{j=1}^{k} \delta_j r_j^{1-2\xi} = -\frac{1}{24} S(\xi)^2 S(1-2\xi).$$

In particular, $h(\boldsymbol{w}(0)) < h(\boldsymbol{w}(1/3))$ and

$$h(\boldsymbol{w}(0)) = -S(0)^2 S(1)/24 = -(2\pi)^2|\Gamma|/24 = -\pi^2|\Gamma|/6,$$
$$h(\boldsymbol{w}(1/3)) = -S(1/3)^3/24 = -\frac{1}{24}\left[\sum_{j=1}^{k} \delta_j r_j^{1/3}\right]^3 = -\frac{1}{24}\left[\int_\Gamma \kappa^{2/3}(s)\,\mathrm{d}s\right]^3.$$

Here we have used that $|\Gamma_j| = \delta_j r_j$ and $\int_{\Gamma_j} \kappa^{2/3}(s)\,\mathrm{d}s = |\Gamma_j|/r_j^{2/3} = \delta_j r_j^{1/3}$ since Γ_j is a circular arc of radius r_j and central angle δ_j.

Hence, property $c_1^+ \in I_1$ and inequalities (26) hold provided $\boldsymbol{w}(0) \in K_\infty^{(1)}$ and $\boldsymbol{w}(1/3) \in P_\infty^{(1)}$. It turns out that $\boldsymbol{w}(\xi)$ satisfies the $3k+1$ conditions listed in (24), so that $\boldsymbol{w}(\xi) \in P_\infty^{(1)}$, for all $\xi \in (0, 1/2]$. For instance, $\boldsymbol{w}(\xi)$ satisfies the first $2k-2$ inequalities:

$$\xi \in (0,1/2] \Rightarrow r_j^{\xi} \min\left\{1, \sqrt{r_{j+1}/r_j}\right\} < r_{j+1}^{\xi} < r_j^{\xi} \max\left\{1, \sqrt{r_{j+1}/r_j}\right\}$$
$$\Rightarrow \alpha_j^- w_j(\xi) < w_{j+1}(\xi) < \alpha_j^+ w_j(\xi) \text{ for } j = 1, \dots, k-1.$$

Inequalities $\alpha_k^- w_k(\xi) < w_1(\xi) < \alpha_k^+ w_k(\xi)$ are proved in a similar way. Inequalities $w_j(\xi) > 0$ and identity $w_1(\xi) + \dots + w_k(\xi) = 1$ are trivial. Finally, $\boldsymbol{w}(0) = \lim_{\xi \to 0^+} \boldsymbol{w}(\xi) \in K_\infty^{(1)}$ □

The main result of this section is nothing more than a reformulation of Theorem E.

Theorem 50. *Let $P_\infty^{(1)} \subset \Delta_{k-1}$, $I_1 = h\big(P_\infty^{(1)}\big) \subset (-\infty, 0)$, $c_1^- = \inf I_1$ and $c_1^+ = \max I_1$ be the open bounded convex polytope of $\mathbb{R}^k$, the image interval, and the extreme values introduced in Lemma 49, respectively. Extreme values $c_1^\pm$ satisfy inequalities* (26)*. For any fixed $c \in [c_1^-, c_1^+]$ there exist a period $q_0 \in \mathbb{N}$ and a sequence $(g_q)_{q \geq q_0}$ of generic sliding periodic billiard trajectories $g_q \in \Pi(1, q)$ such that*

$$L(g_q) = |\Gamma| + c/q^2 + \mathrm{O}(1/q^3), \quad \textit{as } q \to +\infty.$$

Consequently, there exist a sequence $(h_q)_q$, with $h_q \in \Pi(1, q)$, such that

$$c_1^- = \liminf_{q \to +\infty} \big((L(h_q) - |\Gamma|)q^2\big) < \limsup_{q \to +\infty} \big((L(h_q) - |\Gamma|)q^2\big) = c_1^+, \quad \textit{as } q \to +\infty.$$

Proof. If $c \in (c_1^-, c_1^+]$, then $c = h(\boldsymbol{y})$ for some $\boldsymbol{y} \in P_\infty^{(1)}$. If $q \in \mathbb{N}$ is big enough, then there exists a point $\boldsymbol{x} = (x_1, \dots, x_k) \in \mathbb{N}^k$ such that $|q\boldsymbol{y} - \boldsymbol{x}|_\infty \leq 1$ and $\boldsymbol{x} \in P^{(1)} \cap qH_k$, where $P^{(1)}$ is the unbounded convex polytope defined in (16) for $p = 1$. Let us prove this claim.

First, we observe that $y_j > 0$, so $qy_j \geq 1$ when $q \gg 1$. If $q\boldsymbol{y} \in \mathbb{N}^k$, then we take $\boldsymbol{x} = q\boldsymbol{y}$. If $q\boldsymbol{y} \notin \mathbb{N}^k$, then we can take, for instance,

$$x_j = \begin{cases} \lfloor qy_j \rfloor + 1, & \text{for } j \leq i, \\ \lfloor qy_j \rfloor, & \text{otherwise,} \end{cases}$$

where $i = q - \sum_{j=1}^k \lfloor qy_j \rfloor \in \{1, \ldots, k-1\}$, so that $\sum_{j=1}^k x_j = i + \sum_{j=1}^k \lfloor qy_j \rfloor = q$. This means that $\boldsymbol{x} \in qH_k$. To end the proof of the claim, we deduce that $\boldsymbol{x} \in P^{(1)}$ from limits $\lim_{q\to+\infty} \boldsymbol{x}/q = \boldsymbol{y} \in P_\infty^{(1)}$ and $P_\infty^{(1)} = \lim_{q\to\infty} \left(\left\{\boldsymbol{x}/q : \boldsymbol{x} \in P^{(1)}\right\} \cap H_k\right)$. Recall that $P_\infty^{(1)}$ is an open set in H_k.

As we have explained before, see Remark 47, if $q \gg 1$ then there are at least 2^k generic sliding periodic billiard trajectories $g_q \in \Pi(1,q)$ with exactly $x_j \in \mathbb{N}$ impacts on the arc Γ_j and length (23). The numbers $x_j \in \mathbb{N}$, the constant angles of reflection $\psi_j > 0$ and the impact angles $\varphi_j^\pm \in (0, 2\psi_j)$ described in Lemma 48 satisfy identity $\varphi_j^- + 2(x_j - 1)\psi_j + \varphi_j^+ = \delta_j$ and uniform estimates $x_j = qy_j + \mathrm{O}(1)$, $\varphi_j^\pm = \mathrm{O}(1/q)$ and $\psi_j = \delta_j/2(x_j - 1) + \mathrm{O}(1/q^2) = \mathrm{O}(1/q)$ as $q \to +\infty$. Therefore,

$$\ell_j^\pm = \frac{r_j \sin \varphi_j^\pm}{\cos(\psi_j - \varphi_j^\pm)} = r_j \varphi_j^\pm + \mathrm{O}\left((\varphi_j^\pm)^3, \varphi_j^\pm |\psi_j - \varphi_j^\pm|^2\right) = r_j \varphi_j^\pm + \mathrm{O}(1/q^3)$$

and

$$\begin{aligned}
(x_j - 1)\ell_j &= 2r_j(x_j - 1)\sin\psi_j = 2r_j(x_j - 1)\Big(\psi_j - \psi_j^3/6 + \mathrm{O}\left(\psi_j^5\right)\Big) \\
&= 2r_j(x_j - 1)\psi_j - r_j(x_j - 1)\psi_j^3/3 + \mathrm{O}(1/q^4) \\
&= 2r_j(x_j - 1)\psi_j - \frac{\delta_j^3 r_j}{24(x_j - 1)^2} + \mathrm{O}(1/q^3) \\
&= 2r_j(x_j - 1)\psi_j - \frac{\delta_j^3 r_j}{24 y_j^2}\frac{1}{q^2} + \mathrm{O}(1/q^3).
\end{aligned}$$

Finally, we estimate the total length (23) as follows:

$$\begin{aligned}
L(g_q) &= \sum_{j=1}^k \left(\ell_j^- + (x_j - 1)\ell_j + \ell_j^+\right) \\
&= \sum_{j=1}^k r_j(\varphi_j^- + 2(x_j - 1)\psi_j + \varphi_j^+) + h(\boldsymbol{y})/q^2 + \mathrm{O}(1/q^3) \\
&= |\Gamma| + c/q^2 + \mathrm{O}(1/q^3).
\end{aligned}$$

We have used that $\varphi_j^- + 2(x_j - 1)\psi_j + \varphi_j^+ = \delta_j$, $|\Gamma| = \sum_{j=1}^k \delta_j r_j$ and $c = h(\boldsymbol{y})$ in the last line. Function $h(\boldsymbol{y})$ was defined in (25). This ends the proof of the case $c \in (c_1^-, c_1^+]$.

The case $c = c_1^-$ can be obtained from the case $c \in (c_1^-, c_1^+]$ by using a classical diagonalisation argument about sequences of sequences of lengths.

Finally, sequence $(h_q)_q$ is constructed by interleaving two sequences of generic sliding $(1,q)$-periodic billiard trajectories associated with the asymptotic coefficients c_1^- and c_1^+ respectively. □

As a by-product of Theorem 50, we get that generic sliding $(1,q)$-periodic billiard trajectories inside circular polygons are *asymptotically shorter* than the ones inside smooth ovals, since c_1^+ in (26) has the same formula that constant c_1 in part (c) of Theorem 44. The relation between the left endpoint c_1^- of the interval I_1 and $-\pi^2|\Gamma|/6$ is an open problem. If $c_1^- = -\pi^2|\Gamma|/6$ were true, nodal billiard trajectories would be the *asymptotically shortest* sliding $(1,q)$-periodic billiard trajectories as $q \to +\infty$. This is one of the reasons to take them into account.

Identity $c_1^- = \pi^2|\Gamma|/6$ holds for squared pseudo-ellipses. Let $\Gamma = E_{\pi/2,r,R}$ be a squared pseudo-ellipse of radii r and $R > r$, see Section 2. That is, $\Gamma = E_{\pi/2,r,R}$ is the circular 4-gon with radii $r_1 = r_3 = r$ and $r_2 = r_4 = R$, and central angles $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \pi/2$. A tedious

computation that we omit for the sake of brevity shows that

$$c_1^- = -\pi^2 |E_{\pi/2,r,R}|/6 = -\pi^3(R+r)/6,$$
$$c_1^+ = -\frac{1}{24}\left[\int_{E_{\pi/2,r,R}} \kappa^{2/3}(s)\,\mathrm{d}s\right]^3 = -\frac{1}{24}\left[\sum_{j=1}^{k} \delta_j \sqrt[3]{r_j}\right]^3 = -\frac{\pi^3}{24}\left[\sqrt[3]{R}+\sqrt[3]{r}\right]^3.$$

These two expressions above coincide when $R = r$. In general, $c_1^+ - c_1^-$ tends to zero when Γ tends to a circle of finite radius.

Acknowledgments

A. C. has received funding for this project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 757802). R. R.-R. was supported in part by the grant PID-2021-122954NB-100 which was funded by MCIN/AEI/10.13039/501100011033 and "ERDF: A way of making Europe". Thanks to Aida Chaikh and Pau Martín for useful and stimulating conversations. We are also indebted to the referee for several suggestions that helped us to improve the exposition.

A Proof of Lemma 24

Throughout this proof we shall freely use the natural convention that the objects $\Xi_{k+1}^i, \zeta_{k+1}^i, \xi_{k+1}^i$ should be identified with $\Xi_1^{i+1}, \zeta_1^{i+1}, \xi_1^{i+1}$, respectively.

(a) The key observation is that functions $\zeta_j^i(n)$ and $\xi_j^i(n)$ can be recursively bounded. To be precise, since $\zeta_{j+1}^i(n)$ is the smallest integer such that $\zeta_{j+1}^i(n) > \alpha_j^- \zeta_j^i(n) + \beta_j^-$, $\xi_{j+1}^i(n)$ is the largest integer such that $\xi_{j+1}^i(n) < \alpha_j^+ \xi_j^i(n) - \beta_j^+$ and $\beta_j^+ = \alpha_j^+ + 1 > 1$, we deduce that

$$\left.\begin{array}{rl} \zeta_{j+1}^0(n) \le \alpha_j^- \zeta_j^0(n) + \beta_j^- + 1, & \forall j = 1.\ldots,k \\ \alpha_j^+ \xi_j^i(n) - \beta_j^+ - 1 \le \xi_{j+1}^i(n) \le \alpha_j^+ \xi_j^i(n), & \forall j = 1,\ldots,k\ \forall i \ge 0. \end{array}\right\} \tag{27}$$

(A comment is in order. The careful reader may notice that, by the definition of the alphabet $\boldsymbol{Q}$, $\zeta_j^0(n) \ge \chi_j$. Thus, it looks like we should have written the bound

$$\zeta_{j+1}^0(n) \le \max\left\{\chi_{j+1}, \alpha_j^- \zeta_j^0(n) + \beta_j^- + 1\right\}, \qquad \forall j = 1,\ldots,k$$

instead of the first bound in (27). However, we do not need it when $\chi_1 \gg \chi_2,\ldots,\chi_k$. Under this assumption, which is the second part of hypothesis (**X**), we know that the first $k-1$ minima $\zeta_2^0(n),\ldots,\zeta_k^0(n)$ are not affected by restrictions $\xi_j^0(n) \ge \chi_j$, and, if necessary, we replace $\zeta_1^1(n) = \zeta_{k+1}^0(n)$ —which is the last minimum that we need to take care of— by χ_1.) If we apply recursively k times the bounds (27), we get the cyclic bounds

$$\left.\begin{array}{l} \zeta_1^1(n) \le \max\{\chi_1, \zeta_1^0(n)/\alpha + \gamma_1^-\} \\ \alpha\xi_j^i(n) - \gamma_j^+ \le \xi_j^{i+1}(n) \le \alpha\xi_j^i(n), \quad \forall j = 1,\ldots,k, \forall i \ge 0 \end{array}\right\} \tag{28}$$

where $\alpha = \prod_{j=1}^k \alpha_j^+$, $1/\alpha = \prod_{j=1}^k \alpha_j^-$ and

$$\gamma_j^\pm = \sum_{m=1}^{k}\left(\prod_{l=1}^{m-1} \alpha_{j-l}^\pm\right)\left(\beta_{j-m}^\pm + 1\right), \qquad \forall j = 1,\ldots,k.$$

i) If we apply recursively $j-1$ times the bounds for the maxima in the second line of equation (27), we get

$$\lambda_j n - \gamma_j = \lambda_j \xi_1^0(n) - \gamma_j \le \xi_j^0(n) \le \lambda_j \xi_1^0(n) = \lambda_j n, \qquad \forall j = 1,\ldots,k, \tag{29}$$

where $\lambda_1 = 1$, $\gamma_1 = 0$, and

$$\lambda_j = \prod_{l=1}^{j-1} \alpha_l^+, \qquad \gamma_j = \sum_{m=1}^{j-1}\left(\prod_{l=1}^{m-1} \alpha_{j-l}^\pm\right)\left(\beta_{j-m}^\pm + 1\right), \qquad \forall j = 2,\ldots,k.$$

If $\lambda = \max\{\lambda_1, \dots, \lambda_k\}$ and we choose any ν such that $0 < \nu < \min\{\lambda_1, \dots, \lambda_k\}$, then (29) implies that $\nu n \le \xi_j^0(n) \le \lambda n$ for all $j = 1, \dots, k$ provided that χ_1 is large enough, as it is assumed in hypothesis (**X**). To be precise, if we assume that $n \ge \chi_1 \gg 1$, then $\nu n \le \lambda_j n - \gamma_j$ for all $j = 1, \dots, k$. It suffices to take

$$\chi_1 \ge \max\left\{\gamma_j/(\lambda_j - \nu) : j = 1, \dots, k\right\}.$$

ii) We assume that $i \ge 0$. The upper bound $\xi_j^i(n) \le \alpha^i \xi_j^0(n)$ follows directly from (28). The lower bound $\xi_j^i(n) \ge \tau\alpha^i\xi_j^0(n)$ for some $\tau \in (0,1)$ is more tricky. First, we realise that if we choose any $\kappa \in (1, \alpha)$, then (28) implies the weaker lower bound $\xi_j^i(n) \ge \kappa^i \xi_j^0(n)$ provided $n \ge \chi_1 \ge \max\left\{\gamma_1^+, \dots, \gamma_k^+\right\}/(\alpha - \kappa)$. This means that $\xi_j^i(n)$ grows geometrically as $i \to +\infty$. Second, we know that

$$\xi_j^i(n) \ge \alpha\xi_j^{i-1}(n) - \gamma_j^+ = \left(1 - \frac{\gamma_j^+}{\alpha\xi_j^{i-1}(n)}\right)\alpha\xi_j^{i-1}(n) \ge \dots \ge \tau_{i,j}\alpha^i\xi_j^0(n),$$

where

$$0 < \prod_{l=0}^{+\infty}\left(1 - \frac{\gamma_j^+}{\alpha\xi_j^l(n)}\right) =: \tau_j < \tau_{i,j} = \prod_{l=0}^{i-1}\left(1 - \frac{\gamma_j^+}{\alpha\xi_j^l(n)}\right) < 1, \qquad \forall i \ge 0.$$

The above infinite product converges to a non-zero value τ_j because

$$\sum_{l=0}^{+\infty}\frac{\gamma_j^+}{\alpha\xi_j^l(n)} \le \frac{\gamma_j^+}{\alpha\xi_j^0(n)}\sum_{l=0}^{+\infty}\kappa^{-l} < +\infty.$$

If we set $\tau = \min\{\tau_1, \dots, \tau_k\}$, then $\xi_j^i(n) \ge \tau\alpha^i\xi_0^i(n)$. This ends the proof for the forward case $i \ge 0$. The backward case $i < 0$ is proved in a similar way.

iii) Inequality $\rho^i(n) \le \rho^{i+1}(n)$ is trivial. Using the already proved parts (ai) and (aii) of this lemma and the formula for geometric sums, we get

$$\begin{aligned}
\frac{\rho^{i+1}(n)}{\xi_j^i(n)} &\le \frac{\sum_{j=1}^k\sum_{m=0}^i \alpha^m\xi_j^0(n)}{\tau\alpha^i\xi_j^0(n)} \le \frac{\sum_{j=1}^k\sum_{m=0}^i \alpha^m\lambda n}{\tau\alpha^i\nu n} = \frac{k\lambda(\alpha^{i+1}-1)}{\tau\nu(\alpha-1)\alpha^i} \\
&\le \frac{k\lambda\alpha}{\tau\nu(\alpha-1)} =: \lambda', \\
\frac{\rho^i(n)}{\xi_j^i(n)} &\ge \frac{\sum_{j=1}^k\sum_{m=0}^{i-1} \tau\alpha^m\xi_j^0(n)}{\alpha^i\xi_j^0(n)} \ge \frac{\sum_{j=1}^k\sum_{m=0}^{i-1} \tau\alpha^m\nu n}{\alpha^i\lambda n} = \frac{k\tau\nu(\alpha^i-1)}{\lambda(\alpha-1)\alpha^i} \\
&\ge \frac{k\tau\nu}{\lambda\alpha} =: \nu'.
\end{aligned}$$

iv) The inequalities $n/\alpha + \gamma^- \le n - 1 < n + 1 \le \alpha n - \gamma^+$ for all $n \ge \chi_1$ follow from hypotheses (**B**) and (**X**). It suffices to take

$$\chi_1 \ge \max\{(1+\gamma^+)/(\alpha-1), (1+\gamma^-)/(1-1/\alpha)\}.$$

Set $\gamma^\pm = \gamma_1^\pm$. The inequalities $\zeta_1^1(n) \le \max\{\chi_1, n/\alpha + \gamma^-\}$ and $\alpha n - \gamma^+ \le \xi_1^1(n)$ follow directly by taking $i = 0$ in (28), because $\zeta_1^0(n) = n = \xi_1^0(n)$ by definition.

v) If we take $n \ge \max\{(\chi_1 - \gamma^-)\alpha, (N+\gamma^+)/(\alpha-1), (N+\gamma^-)/(1-1/\alpha)\}$, then $\chi_1 \le n/\alpha + \gamma^- \le n - N < n + N \le \alpha n - \gamma^+$.

(b) Let us check that the sets $\Xi_j^i(n)$ have no gaps in $\mathbb{N}$. That is, we want to check that $[n^-, n^+] \cap \mathbb{N} \subset \Xi_j^i(n)$ for all $n^\pm \in \Xi_j^i(n)$ such that $n^- \le n^+$.

First, we consider the forward case $i \ge 0$. We prove it by induction on the ordering

$$\Xi_1^0, \dots, \Xi_k^0, \Xi_1^1 = \Xi_{k+1}^0, \dots, \Xi_k^1, \Xi_1^2 = \Xi_{k+1}^1, \dots, \Xi_k^2, \dots, \Xi_1^i = \Xi_{k+1}^{i-1}, \dots, \Xi_j^i, \Xi_{j+1}^i, \dots.$$

The base case is trivial: $\Xi_1^0(n) = \{n\}$. Let us perform now the inductive step. We assume that $\Xi_j^i(n)$ has no holes in $\mathbb{N}$ for some $i \ge 0$ and $1 \le j \le k$. The next set is

$$\Xi_{j+1}^i(n) = \left\{n'' \in \mathbb{N} : n'' \ge \chi_{j+1},\ \exists n' \in \Xi_j^i(n) \text{ s. t. } \alpha_j^- n' + \beta_j^- < n'' < \alpha_j^+ n' - \beta_j^+\right\}.$$

If $\Xi^i_{j+1}(n)$ has a hole in $\mathbb{N}$, there is $n' \geq \chi_j$ such that $\alpha^+_j n' - \beta^+_j \leq \alpha^-_j(n'+1) + \beta^-_j$, which is impossible by hypotheses (**A**) and (**X**). It suffices to take

$$\chi_j > (\alpha^-_j + \beta^-_j + \beta^+_j)/(\alpha^+_j - \alpha^-_j).$$

Property $\big[\max\{\chi_1, n - |i|\}, n + |i|\big] \cap \mathbb{N} \subset \Xi^i_1(n)$ for all $i \in \mathbb{Z}$ and $n \geq \chi_1$ follows by induction from part (aiv) of this lemma and the fact that $\Xi^i_1(n)$ has no gaps in $\mathbb{N}$.

This ends the proof for the forward case $i \geq 0$. The backward case $i < 0$ is similar. □

B Proof of Proposition 39

Fix any $p \in \mathbb{N}$. We look for the optimal value of $c_\star(p) > 0$ such that

$$\#\Pi(p,q) \geq 2^{kp} G_q\big(P^{(p)}\big) \geq c_\star(p) q^n + \mathrm{O}(q^{n-1}) \quad \text{as } q \to +\infty.$$

Therefore, we want to count as many integer points as possible in $P^{(p)} \subset \mathbb{R}^{n+1}$ whose coordinates sum $q \in \mathbb{N}$. We shall put these points in a 1-to-1 correspondence with the integer points of a q-dilated bounded convex polytope of $\mathbb{R}^n$ by means of a projection.

We shall use a lower bound established by Wills [57]. Let us briefly describe it. If $t > 0$ and $P \subset \mathbb{R}^n$, then $tP = \{y\boldsymbol{x} : \boldsymbol{x} \in P\}$ and $P/t = \{\boldsymbol{x}/t : \boldsymbol{x} \in P\}$ are the *t-dilation* and *t-contraction* of P. The *inradius* $\varrho(K)$ of a proper compact convex set $K \subset \mathbb{R}^n$ is the biggest number $\varrho > 0$ such that K contains a ball of radius ϱ. Note that $0 < \varrho(K) < \infty$ for any proper compact K.

Lemma 51. *If K is a proper compact convex subset of $\mathbb{R}^n$, then*

$$\#(tK \cap \mathbb{Z}^n) \geq \mathrm{V}(K)\big(t - \sqrt{n}/2\varrho(K)\big)^n, \qquad \forall t \geq \sqrt{n}/2\varrho(K).$$

Proof. The case $t = 1$ is proved in [57], assuming that $\varrho(K) \geq \sqrt{n}/2$. The general case follows directly from this case since tK is a proper compact convex subset of $\mathbb{R}^n$, $\mathrm{V}(tK) = t^n\,\mathrm{V}(K)$, and $\varrho(tK) = t\varrho(K) \geq 2/\sqrt{n}$ if $t \geq \sqrt{n}/2\varrho(K)$. □

The convex polytope (16) is not closed, so the convex polytopes $\tilde{P}^{(p)}_q$ defined in Proposition 39 are not closed either. However, they are the projection of some convex polytopes contained in the open simplex $\Delta_n = \{\boldsymbol{x} \in \mathbb{R}^{n+1} : \boldsymbol{x} > 0,\ x_1 + \cdots + x_n = 1\}$, which implies that they are bounded. Hence we need to extend Lemma 51 to proper bounded convex subsets of $\mathbb{R}^n$.

Corollary 52. *If P is a proper bounded convex subset of $\mathbb{R}^n$ and $K = \bar{P}$, then*

$$\#(tP \cap \mathbb{Z}^n) \geq \mathrm{V}(K)\big(s - \sqrt{n}/2\varrho(K)\big)^n, \qquad \forall t > s \geq \sqrt{n}/2\varrho(K).$$

Proof. The closure $K = \bar{P}$ is compact. Let $\bar{B}$ be a closed ball of radius $\varrho(K) > 0$ contained in K. Let $B = \operatorname{Int} \bar{B}$. Given any point $-\boldsymbol{x} \in B$, we have that $s(\boldsymbol{x} + K) \subset t(\boldsymbol{x} + P)$ for all $t > s > 0$. If $t > \sqrt{n}/2\varrho(K)$, then there is a point $-\boldsymbol{x}_t \in B$ such that $t\boldsymbol{x}_t \in \mathbb{Z}^n$. Then

$$\begin{aligned}\#(tP \cap \mathbb{Z}^n) &= \#\big((t\boldsymbol{x}_t + tP) \cap \mathbb{Z}^n\big) = \#\big(t(\boldsymbol{x}_t + P) \cap \mathbb{Z}^n\big) \geq \#\big(s(\boldsymbol{x}_t + K) \cap \mathbb{Z}^n\big) \\ &\geq \mathrm{V}(\boldsymbol{x}_t + K)\big(s - \sqrt{n}/2\varrho(\boldsymbol{x}_t + K)\big)^n = \mathrm{V}(K)\big(s - \sqrt{n}/2\varrho(K)\big)^n,\end{aligned}$$

for all $t > s \geq \sqrt{n}/2\varrho(K)$. □

Proof of Proposition 39. (a) Let $H_{n+1} = \big\{\boldsymbol{x} \in \mathbb{R}^{n+1} : x_1 + \cdots + x_{n+1} = 1\big\}$. The cardinality

of a finite set is invariant under q-dilations, q-contractions, and 1-to-1 projections. Thus,

$$\begin{aligned}
\#\Pi(p,q) &\geq 2^{kp} G_q\big(P^{(p)}\big) \\
&= 2^{kp} \#\Big\{\boldsymbol{x} = (x_1,\dots,x_{n+1}) \in P^{(p)} \cap \mathbb{Z}^{n+1} : x_1 + \cdots + x_{n+1} = q\Big\} \\
&= 2^{kp} \#\big(P^{(p)} \cap \mathbb{Z}^{n+1} \cap qH_{n+1}\big) \\
&= 2^{kp} \#\big((P^{(p)}/q) \cap (\mathbb{Z}^{n+1}/q) \cap H_{n+1}\big) \\
&= 2^{kp} \#\big(P_q^{(p)} \cap (\mathbb{Z}^{n+1}/q)\big) \\
&= 2^{kp} \#\big(qP_q^{(p)} \cap \mathbb{Z}^{n+1}\big) \\
&= 2^{kp} \#\big(q\tilde{P}_q^{(p)} \cap \mathbb{Z}^{n}\big) \\
&\geq 2^{kp}\, \mathrm{V}\,\big(\tilde{K}_q^{(p)}\big)\Big(q - 1 - \sqrt{n}/2\varrho\big(\tilde{K}_q^{(p)}\big)\Big)^n \\
&\geq 2^{kp} V\big(\tilde{K}_\infty^{(p)}\big) q^{kp-1} + \mathrm{O}(q^{kp-2}) \quad \text{ as } q \to +\infty,
\end{aligned}$$

where $\tilde{K}_q^{(p)}$ is the closure of $\tilde{P}_q^{(p)}$. We have used Corollary 52 with $t = q$ and $s = q - 1$ in the second to last inequality. In the last inequality, we have used estimates

$$V\big(\tilde{K}_q^{(p)}\big) = V\big(\tilde{K}_\infty^{(p)}\big) + \mathrm{O}(1/q), \qquad \varrho\big(\tilde{K}_q^{(p)}\big) = \varrho\big(\tilde{K}_\infty^{(p)}\big) + \mathrm{O}(1/q).$$

These estimates follow from the fact that each facet of the limit compact polytope $\tilde{K}_\infty^{(p)}$ is at an $\mathrm{O}(1/q)$-distance of the corresponding facet of the polytope $\tilde{K}_q^{(p)}$, which can be easily seen by comparing the half-space representation (20) of $\tilde{K}_\infty^{(p)}$ with the half-space representation

$$\tilde{K}_q^{(p)} = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^n : \begin{array}{l} \alpha_j^- x_j + \beta_j^-/q \leq x_{j+1} \leq \alpha_j^+ x_j - \beta_j^+/q, \quad \forall j = 1,\dots,n-1 \\ \alpha_n^- x_n + \beta_n^-/q \leq 1 - \varsigma(\tilde{\boldsymbol{x}}) \leq \alpha_n^+ x_n - \beta_n^+/q \\ \alpha_{n+1}^-(1 - \varsigma(\tilde{\boldsymbol{x}})) + \beta_{n+1}^-/q \leq x_1 \leq \alpha_{n+1}^+(1 - \varsigma(\tilde{\boldsymbol{x}})) - \beta_{n+1}^+/q \\ x_j \geq \chi_j/q, \quad \forall j = 1,\dots,n \\ \varsigma(\tilde{\boldsymbol{x}}) \leq 1 - \chi_{n+1}/q \end{array} \right\}. \tag{30}$$

(b) All convex bodies are Jordan measurable and

$$\mathrm{V}(J) = \lim_{q\to+\infty} q^{-n} \#\left(J \cap (\mathbb{Z}^n/q)\right) = \lim_{q\to+\infty} q^{-n} \#\left(qJ \cap \mathbb{Z}^n\right)$$

for any Jordan measurable set $J \subset \mathbb{R}^n$, see [30, section 7.2]. Therefore,

$$\begin{aligned}
\lim_{q\to+\infty} q^{-n} G_q\big(P^{(p)}\big) &\leq \lim_{q\to+\infty} q^{-n} \#\big(q\tilde{K}_q^{(p)} \cap \mathbb{Z}^n\big) \\
&\leq \lim_{q\to+\infty} q^{-n} \#\big(q\tilde{K}_\infty^{(p)} \cap \mathbb{Z}^n\big) = \mathrm{V}\,\big(\tilde{K}_\infty^{(p)}\big), \\
\lim_{q\to+\infty} q^{-n} G_q\big(P^{(p)}\big) &\geq \lim_{q\to+\infty} \Big(V\big(\tilde{K}_\infty^{(p)}\big) + \mathrm{O}(1/q)\Big) = \mathrm{V}\,\big(\tilde{K}_\infty^{(p)}\big).
\end{aligned}$$

We have used that $\tilde{K}_q^{(p)} \subset \tilde{K}_\infty^{(p)}$ —compare half-space representations (30) and (20)— in the first line and the lower bound obtained at the beginning of this proof in the second one.

(c) It is a simple computation using that $x_{n+1} = 1 - \varsigma(\tilde{\boldsymbol{x}})$ when $\boldsymbol{x} = (\tilde{\boldsymbol{x}}, x_{n+1}) \in H_{n+1}$. □

References

[1] C. Alsina and R. B. Nelsen. *Icons of Mathematics: An Exploration of Twenty Key Images*, volume 56 of *The Dolciani Mathematical Expositions*. Mathematical Association of America, Washington, DC, 2011.

[2] G. E. Andrews. A theorem on reciprocal polynomials with applications to permutations and compositions. *The American Mathematical Monthly*, 82(8):830–833, 1975.

[3] G. Arioli and P. Zgliczyński. Symbolic dynamics for the Hénon–Heiles Hamiltonian on the critical level. *Journal of Differential Equations*, 171(1):173–202, 2001.

[4] P. Bálint, M. Halász, J. A. Hernández-Tahuilán, and D. P. Sanders. Chaos and stability in a two-parameter family of convex billiard tables. *Nonlinearity*, 24(5):1499, 2011.

[5] T. Banchoff and P. Giblin. On the geometry of piecewise circular curves. *The American Mathematical Monthly*, 101(5):403–416, 1994.

[6] G. Benettin and J.-M. Strelcyn. Numerical experiments on the free motion of a point mass moving in a plane convex region: Stochastic transition and entropy. *Physical Review A*, 17(2):773, 1978.

[7] M. Berger. Seules les quadriques admettent des caustiques. *Bulletin de la Société Mathématique de France*, 123(1):107–116, 1995.

[8] M. Bialy and B. Youssin. Numerical non-integrability of hexagonal string billiard. `arXiv:2202.06801 [math.DS]`, 2022.

[9] G. D. Birkhoff. Surface transformations and their dynamical applications. *Acta Mathematica*, 43:1–119, 1922.

[10] L. Buhovsky and V. Kaloshin. Nonisometric domains with the same Marvizi-Melrose invariants. *Regular and Chaotic Dynamics*, 23(1):54–59, 2018.

[11] L. Bunimovich, H.-K. Zhang, and P. Zhang. On another edge of defocusing: hyperbolicity of asymmetric lemon billiards. *Communications in Mathematical Physics*, 341(3):781–803, 2016.

[12] L. A. Bunimovich. On the ergodic properties of nowhere dispersing billiards. *Communications in Mathematical Physics*, 65(3):295–312, 1979.

[13] L. A. Bunimovich. Elliptic flowers: simply connected billiard tables where chaotic (non-chaotic) flows move around chaotic (non-chaotic) cores. *Nonlinearity*, 35(6):3245, 2022.

[14] J.-P. Burelle and R. Kirk. Piecewise circular curves and positivity. `arXiv:2108.08680 [math.DG]`, 2021.

[15] P. S. Casas and R. Ramírez-Ros. Classification of symmetric periodic trajectories in ellipsoidal billiards. *Chaos*, 22:026110, 2012.

[16] J. Chen, L. Mohr, H.-K. Zhang, and P. Zhang. Ergodicity of the generalized lemon billiards. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(4):043137, 2013.

[17] N. Chernov and R. Markarian. *Chaotic Billiards*, volume 127 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[18] A. Clarke. Generic properties of geodesic flows on analytic hypersurfaces of Euclidean space. *Discrete and Continuous Dynamical Systems*, 42(12):5839–5868, 2022.

[19] A. Clarke and D. Turaev. Arnold diffusion in multidimensional convex billiards. *Duke Mathematical Journal*, 172(10):1813–1878, 2023.

[20] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.

[21] R. Dixon. *Mathographics*. Dover Publications Inc., New York, 1987.

[22] R. Douady. *Applications du théoreme des tores invariants. Thèse de 3ème cycle*. PhD thesis, Université Paris VII, 1982.

[23] H. R. Dullin, P. H. Richter, and A. Wittek. A two-parameter study of the extent of chaos in a billiard system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 6(1):43–58, 1996.

[24] H. L. Fetter. Numerical exploration of a hexagonal string billiard. *Physica D: Nonlinear Phenomena*, 241(8):830–846, 2012.

[25] P. Flajolet and R. Sedgewick. *Combinatorics*. Cambridge University Press, Cambridge, UK, 2009.

[26] M. Gidea and P. Zgliczyński. Covering relations for multidimensional dynamical systems–II. *Journal of Differential Equations*, 202(1):59–80, 2004.

[27] G. Gonçalves, S. Oliffson Kamphorst, and S. Pinto-de Carvalho. Symmetric periodic orbits in symmetric billiards. *Nonlinearity*, 37(1):015005, 2024.

[28] P. M. Gruber. Convex billiards. *Geometriae Dedicata*, 33(2):205–226, 1990.

[29] P. M. Gruber. Only ellipsoids have caustics. *Mathematische Annalen*, 303(1):185–194, 1995.

[30] P. M. Gruber. *Convex and Discrete Geometry*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heildeberg, 2007.

[31] M. Guardia, P. Martín, J. Paradela, and T. M. Seara. Hyperbolic dynamics and oscillatory motions in the 3 body problem. `arXiv:2207.14351 [math.DS]`, 2022.

[32] B. Halpern. Strange billiard tables. *Transactions of the American Mathematical Society*, 232:297–305, 1977.

[33] A. Hayli and T. Dumont. Expériences numériques sur des billards C^1 formés de quatre arcs de cercles. *Celestial Mechanics*, 38(1):23–66, 1986.

[34] M. Hénon and J. Wisdom. The Benettin-Strelcyn oval billiard revisited. *Physica D: Nonlinear Phenomena*, 8(1-2):157–169, 1983.

[35] A. Hubacher. Instability of the boundary in the billiard ball problem. *Communications in Mathematical Physics*, 108(3):483–488, 1987.

[36] X. Jin and P. Zhang. Hyperbolicity of asymmetric lemon billiards. *Nonlinearity*, 34(1):92, 2020.

[37] A. Katok. The growth rate for the number of singular and periodic orbits for a polygonal billiard. *Communications in Mathematical Physics*, 111(1):151–160, 1987.

[38] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Number 54. Cambridge University Press, Cambridge, UK, 1995.

[39] J. Kennedy and J. Yorke. Topological horseshoes. *Transactions of the American Mathematical Society*, 353(6):2513–2530, 2001.

[40] V. V. Kozlov and D. V. Treschev. *Billiards: A Genetic Introduction to the Dynamics of Systems with Impacts*, volume 89 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1991.

[41] V. F. Lazutkin. The existence of caustics for a billiard problem in a convex domain. *Mathematics of the USSR-Izvestiya*, 7(1):185, 1973.

[42] J. Li. Asymptotic estimate for the multinomial coefficients. *Journal of Integer Sequences*, 23:20.1.3, 2020.

[43] H. Makino. Bifurcation and anomalous spectral accumulation in an oval billiard. *Progress of Theoretical and Experimental Physics*, 2019(8):083A01, 2019.

[44] S. Marvizi and R. Melrose. Spectral invariants of convex planar regions. *Journal of Differential Geometry*, 17(3):475–502, 1982.

[45] J. N. Mather. Glancing billiards. *Ergodic Theory and Dynamical Systems*, 2(3-4):397–403, 1982.

[46] J. N. Mather. Variational construction of orbits of twist diffeomorphisms. *Journal of the American Mathematical Society*, 4(2):207–263, 1991.

[47] T. Neuschel. A note on extended binomial coefficients. *Journal of Integer Sequences*, 17:14.10.4, 2014.

[48] D. Papini and F. Zanolin. Fixed points, periodic points, and coin-tossing sequences for mappings defined on two-dimensional cells. *Fixed Point Theory and Applications*, 2004(2):113–134, 2004.

[49] D. Papini and F. Zanolin. On the periodic boundary value problem and chaotic-like dynamics for nonlinear hill's equations. *Advanced Nonlinear Studies*, 4(1):71–91, 2004.

[50] S. Pinto-de Carvalho and R. Ramírez-Ros. Billiards with a given number of (k, n)-orbits. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):026109, 2012.

[51] M. Pireddu. *Fixed points and chaotic dynamics for expansive-contractive maps in Euclidean spaces, with some applications*. PhD thesis, Università di Udine, 2009.

[52] M. Pireddu and F. Zanolin. Cutting surfaces and applications to periodic points and chaotic-like dynamics. *Topological Methods in Nonlinear Analysis*, 30:279–319, 2007.

[53] M. Pireddu and F. Zanolin. Chaotic dynamics in the Volterra predator-prey model via linked twist maps. *Opuscula Mathematica*, 28(4):567–592, 2008.

[54] Y. G. Sinai. Dynamical systems with elastic reflections. *Russian Mathematical Surveys*, 25(2):137, 1970.

[55] A. Sorrentino. Computing Mather's β-function for Birkhoff billiards. *Discrete and Continuous Dynamical Systems*, 35(10):5055–5082, 2015.

[56] S. Tabachnikov. *Billiards*, volume 1 of *Panoramas et Synthèses*. American Mathematical Society, Providence, RI, 1995.

[57] J. M. Wills. Zut Gitterpunktanzahl konvexer Mengen. *Elemente der Mathematik*, 28(3):57–63, 1973.

[58] P. Zgliczyński and M. Gidea. Covering relations for multidimensional dynamical systems. *Journal of Differential Equations*, 202(1):32–58, 2004.