Complexity of piecewise convex transformations in two dimensions, with applications to polygonal billiards on surfaces of constant curvature

Eugene Gutkin [*] and Serge Tabachnikov [†]

May 2, 2006

Abstract

We introduce piecewise convex transformations, and develop geometric tools to study their complexity. We apply the results to the complexity of polygonal inner and outer billiards on surfaces of constant curvature.

To A. A. Kirillov on the occasion of the 70th anniversary

Introduction

Transformations that arise in geometric dynamics often fit into the following scheme. The phase space of the transformation, $T : X \to X$, has a finite decomposition $\mathcal{P} : X = X_a \cup X_b \cup \cdots$ into 'nice' subsets. The interiors of $X_a, X_b, \ldots$ are disjoint, and T is discontinuous (or not defined) on the union

[*]UCLA, Los Angeles, CA 90095, USA and IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro RJ, Brasil. *E-mail*: egutkin@math.ucla.edu,gutkin@impa.br

[†]Department of Mathematics, Pennsylvania State University, University Park, PA 16802. *E-mail*: tabachni@math.psu.edu

of their boundaries, $\partial\mathcal{P} = \partial X_a \cup \partial X_b \cup \cdots$. The sets $X_a, X_b, \cdots$ are endowed with a 'rigid structure'; the restrictions $T : X_a \to X, T : X_b \to X, \ldots$ preserve that structure. Set $\mathcal{P}_1 = \mathcal{P}$. The sets $T^{-1}(X_a) \cap X_a, T^{-1}(X_a) \cap X_b, \ldots$ form the decomposition $\mathcal{P}_2$, which is a refinement of $\mathcal{P}_1$; it provides a defining partition for the transformation T^2. Iterating this process, we obtain a tower of decompositions $\mathcal{P}_n, n \geq 1$, where $\mathcal{P}_n$ plays for T^n the same role that $\mathcal{P}$ played for T.

Let $\mathcal{A} = \{a, b, \ldots\}$ be the alphabet labeling the atoms of $\mathcal{P}$. A phase point $x \in X$ is regular if every point of the orbit $x, Tx, T^2x, \ldots$ belongs to the interior of an atom of $\mathcal{P}$. Let $x \in X_a, Tx \in X_b, \ldots$. The word $\sigma(x) = ab\cdots$ on the alphabet $\mathcal{A}$ is the *code* of x. Let $\Sigma(n)$ be the set of words of length n obtained this way. The function $f(n) = |\Sigma(n)|$ is the *complexity* associated with the triple $X, \mathcal{P}, T$. Its behavior as $n \to \infty$ is an important characteristic of the dynamical system in question. We will develop a geometric approach to complexity.

Several classes of transformations (e. g., piecewise isometries, piecewise affine mappings, etc) fit into the scheme above. The following examples have directly motivated our study.

Example A. Let $P \subset \mathbf{R^2}$ be a polygon with sides $a, b, \ldots$, and let X be the phase space of the *billiard map* T_{bil} in P. Its elements are the billiard segments in P. Let $X_a, X_b, \ldots$ be the set of segments that begin in $a, b, \ldots$. The decomposition $\mathcal{P} : X = X_a \cup X_b \cup \cdots$ yields the coding of billiard orbits by the sides that they hit [16]. Basic questions about its complexity are open [10].

Example B. Let $P \subset \mathbf{R^2}$ be a convex polygon with corners $a, b, \ldots$. The complement $X = \mathbf{R^2} \setminus P$ is the phase space of the *outer billiard* T^{out} about P. (It is also called the *dual billiard*. See [16] or [8]). The conical regions bounded by its singularities form the natural decomposition $\mathcal{P} : X = X_a \cup X_b \cup \cdots$. In $X_a, X_b, \ldots$ the mapping T^{out} is the symmetry about $a, b, \ldots$. The decomposition $\mathcal{P}$ yields the coding of outer billiard orbits by the corners that they hit.

We will study the complexity of (2-dimensional) *piecewise convex transformations*. This is a wide class of geometric dynamical systems; in particular, it contains the examples above. Our setting is as follows.

A *chord space* is a topological space such that for any pair of distinct points x_0, x_1 there is a unique *chord* $[x_0, x_1]$ joining them. See examples of chord spaces in section 1. A *convex cell complex* is a cell complex whose closed cells are chord spaces, and the chord structures agree on the intersections. Let $\mathcal{P} : X = \overline{X}_a \cup \overline{X}_b \cup \cdots$ be the decomposition by the closed 2-cells. Suppose that there are homeomorphisms $T_a : \overline{X}_a \to X, T_b : \overline{X}_b \to X, \ldots$ such that for every chord $\gamma \subset X_a$ and any 2-cell $X_b \subset X$ the curve $T_a(\gamma) \cap \overline{X}_b$ is a chord. Then $(X, \mathcal{P}, T)$ is a *piecewise convex transformation* of X.

In section 1 we develop geometric and combinatorial techniques to study the complexity of piecewise convex transformations. Then we apply these techniques to the inner and outer polygonal billiards on surfaces of constant curvature, $\varkappa$. One of our goals is to develop a uniform approach to these dynamical systems. Note that the elliptic ($\varkappa = 1$) and the hyperbolic ($\varkappa = -1$) cases have been studied much less than the parabolic case ($\varkappa = 0$) [16].

Let $P \subset M$ be a geodesic polygon on a surface of any constant curvature. Since the outer billiard about P is a piecewise isometry, it directly fits into the framework of piecewise convex transformations. In section 2 we put the inner billiard into this framework. In order to do this, we modify the definition of the *billiard phase space*. Our phase space, $X = X(P)$, is the quotient of the set of billiard segments in P by an equivalence relation. See Definition 5. Endowed with the quotient topology and the natural chord structure, X is a cell complex; it is also a (finite, branched) covering of the space, $\mathcal{L} = \mathcal{L}(P)$, of rays intersecting P. See Theorem 1. In section 2 we develop a dictionary to translate the statements of section 1 into the language of billiard orbits. See Proposition 4.

In section 3 we investigate the inner billiard complexity. First, we establish the background by considering arbitrary polygons, and any constant curvature. Then we study each of the three cases separately. Below is a sample of our results.

Let $\varkappa = 0$. The side complexity of billiard orbits in any rational polygon grows at most cubically. See Theorem 3.
Let $\varkappa = 1$. The side complexity of billiard orbits grows subexponentially. See Theorem 4.
Let $\varkappa = -1$. The side complexity $f(n)$ of billiard orbits grows exponentially; the exponent in question is the topological entropy h_{top} of the billiard map. More precisely, $f(n)\exp(-h_{\text{top}}n)$ is a temperate function. See Theorem 5.

In section 4 we investigate complexity of the polygonal outer billiard on surfaces of constant curvature. Here are some of our results.

Let $\varkappa = 0$. For an arbitrary (resp. rational) polygon the compexity has poly-

nomial (resp. quadratic) bounds from above and from below. See Theorem 6 and Theorem 7.
Let $\varkappa = 1$. Then the complexity grows subexponentially. See Theorem 8.
Let $\varkappa = -1$. For an arbitrary polygon, the complexity has a sharp linear lower bound; for *large polygons* the complexity grows linearly. See Theorem 9.

Remark 1 The idea to relate the side complexity with the geometry and combinatorics of billard orbits goes back to [5]. The duscussion in [5] concerns the billiard in a convex, euclidean polygon. Thus, claim 2 of Theorem 3 is contained in [5]. The structure of billiard singularities in convex polygons is less complex than in general; certain types of the singularities that we had to account for in the proof of Lemma 2 do not occur for convex polygons. This circumstance allowed the authors of [5] to obtain an analog of our Lemma 4 directly from Euler's identity. See the proof of Lemma 3.1 in [5]. The paper N. Bedaride, *Billiard complexity in rational polyhedra*, Reg. & Chao. Dyn. 8 (2003) contains an attempt to adapt Lemma 3.1 to noncovex polygons.

Remark 2 The connection observed in [5] is a part of a more general phenomenon. The class of transformations put forward here provides a natural framework to study the coding complexity, and, in particular, this phenomenon. This class (of piecewise convex transformations) is general enough to contain the billiard in any polygon on a surface of constant curvature. However, putting the general billiard into the framework of piecewise convex transformations is far from straightforward. See the discussion in section 2. The preprint [12] is a preliminary version of this work.

Acknowledgements. We thank the Research in Pairs program in Oberwolfach and to the Shapiro visiting program at Penn State for their support. The second author is also grateful to the Max-Planck-Institut in Bonn for its hospitality. The first author thanks ETH in Zurich (FIM) for its support and hospitality. The second author was partially supported by an NSF grant.

1 Piecewise convex transformations: geometry and combinatorics

1.1 Chord spaces and convex cell complexes

A topological space, X, is a *chord space* if for any pair of distinct points $x_0, x_1 \in X$ there is a unique *chord* $[x_0, x_1] \subset X$ joining them. We require that the chords satisfy the standard properties [3]. Let X, Y be chord spaces. A *mapping of chord spaces* is a continuous mapping $f : X \to Y$ that sends chords to chords. A subset $Y \subset X$ of a chord space is chord-convex[1] if for any $y_0, y_1 \in Y$ we have $[y_0, y_1] \subset Y$. Then Y itself is a chord space. The *convex hull*, $\text{hull}(Y) \subset X$, of any subset $Y \subset X$ is a chord space.

Example 1 i) Let X be a *Hadamard manifold*, i. e., a simply connected, complete riemannian manifold of nonpositive sectional curvature. Set $[x, y]$ be the unique connecting geodesic. Then X is a chord space.
ii) Let $x, y \in \mathbf{RP}^n$ be a pair of distinct points in the real projective space of n dimensions. Let $l = l(x, y)$ be the projective line containing them. Let $X \subset \mathbf{RP}^n$ be a convex subset, disjoint from a hyperplane. Defining $[x, y]$ to be the intersection $l(x, y) \cap X$, makes X a chord space.

We will encounter situations where the chord joining $x_0, x_1 \in X$ does not always exist. For example, this happens if X is a nonconvex subset of a chord space. We will say that X is a *space with a chord structure*. Mappings of spaces with chord structures are defined the same way as the mappings of chord spaces.

Example 2 Let $\mathcal{R}$ be the space of *euclidean rays* (i.e., oriented straight lines) in $\mathbf{R^2}$. Endowed with the natural topology, $\mathcal{R}$ is the infinite cylinder [15].

Two rays l_0, l_1 are parallel (resp. *antiparallel*) if the corresponding lines are parallel, and their directions are the same (resp. opposite). Let $l_0, l_1 \in \mathcal{R}$ be distinct rays. If they are not parallel or antiparallel, then l_0, l_1 intersect at a point, say $o \in \mathbf{R^2}$, forming a cone, $C \subset \mathbf{R^2}$, with the apex at o. The chord $[l_0, l_1]$ consists of the rays passing through o and contained in C. If l_0, l_1 are parallel, then C becomes a strip, and $[l_0, l_1]$ is defined analogously. If l_0, l_1 are antiparallel, then $[l_0, l_1]$ is not defined. This is a chord structure on $\mathcal{R}$. Geometrically, $[l_0, l_1]$ is the *pencil of rays* interpolating between l_0, l_1.

[1] We will simply say *convex* in what follows.

Example 3 Let $\mathcal{R}_{\mathbf{H}} \subset \mathcal{R}$ be the set of rays intersecting the unit disc. We endow $\mathcal{R}_{\mathbf{H}}$ with the induced chord structure. Let $l_0, l_1 \in \mathcal{R}_{\mathbf{H}}$ be distinct rays. By definition, the chord $[l_0, l_1]_{\mathbf{H}}$ exists iff $[l_0, l_1]$ exists and $[l_0, l_1] \subset \mathcal{R}_{\mathbf{H}}$. Let AB, CD be the chords of the unit disc corresponding to l_0, l_1 respectively. Then $[l_0, l_1]_{\mathbf{H}}$ does not exist iff the points A, B, C, D of the unit circle are in a cyclic order. See figure 1.

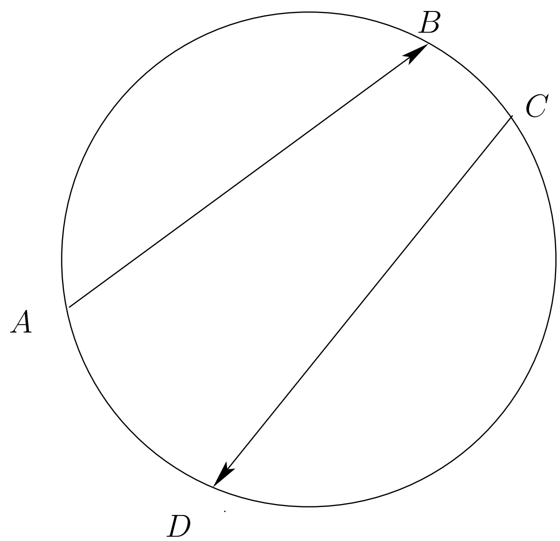

Figure 1: Elements of $\mathcal{R}_{\mathbf{H}}$ that cannot be joined by a chord

For simplicity of exposition, we will restrict our considerations from here on to two dimensions. A *cell complex* is a topological space, X, endowed with a *cell decomposition*. By our assumption, a cell complex has zero-cells (vertices), one-cells (edges), and two-cells (faces). Each cell $C \subset X$ is homeomorphic to the open disc of the same dimension. The boundary $\partial C \subset X$ is homeomorphic to the sphere; it is a finite union of cells of smaller dimension. We will often assume that X is a connected compact, and that every point of X belongs to the closure of a face.

Definition 1 *Let X be a cell complex. Suppose that each closed cell $\overline{C} \subset X$ is a chord space, and that the chord structures agree on the intersections $\overline{C'} \cap \overline{C''}$. Then X is a* convex cell complex.

Let X be a convex cell complex, and let $\Gamma \subset X$ be the union of closures of one-cells. Then Γ is a graph, and its edges are chords; we will say that Γ is a *convex (chord) graph.* We will also speak of spaces with *convex graphs*, and use the notation (X, Γ).

Definition 2 *Let (X, Γ) and (Y, Δ) be convex cell complexes. Let $\varphi : X \to Y$ be a continuous mapping. Suppose that for any cell $D \subset Y$ the preimage $f^{-1}(D) = C_1 \cup \cdots \cup C_n$ is a (nonempty) disjoint union of cells, and that the maps $\varphi : C_i \to D$ are isomorphisms of chord spaces. Then $\varphi : X \to Y$ is a* (branched) covering of convex cell complexes. *The maximal value of $n = n(D)$ is the* degree of the covering.

Let X be a topological space, and let Y be a cell complex. A continuous, surjective mapping $f : X \to Y$ is a *(branched) topological covering* if for any cell $D \subset Y$ we have $f^{-1}(D) = C_1 \cup \cdots \cup C_n$, a disjoint union, and the restrictions $f : C_i \to D$ are homeomorphisms.

Lemma 1 *Let X (resp. (Y, Δ)) be a topological space (resp. a convex cell complex), and let $f : X \to Y$ be a (branched) topological covering. Then there is a unique convex cell complex (X, Γ) such that $f : (X, \Gamma) \to (Y, \Delta)$ is a (branched) covering of convex cell complexes.*

Proof. The representations $f^{-1}(D) = C_1 \cup \cdots \cup C_n$, where $n = n(D)$ and D runs through the cells of Y, define the cells of X, and the unique chord structures on them such that $f|_{C_i} : C_i \to D$ are isomorphisms of chord spaces. Setting $\Gamma = f^{-1}(\Delta)$, we obtain the claim. ∎

We will say that the convex cell complex (X, Γ) of Lemma 1 is *induced by the mapping $f : X \to Y$.*

1.2 Piecewise convex transformations

We will now define a class of dynamical systems that will provide a common framework for several kinds of geometric transformations with singularities.

Definition 3 *Let (X, Γ) be a convex cell complex. Suppose that for each 2-cell $F \subset X$ we have a homeomorphism $T_F : \overline{F} \to X$ such that for any chord $\gamma \subset F$ and any 2-cell $G \subset X$ the curve $T_F(\gamma) \cap \overline{G}$ is a chord. Then (X, Γ, T) is a* piecewise convex self-mapping.

Let (X, Γ, T) and (X, Δ, S) be piecewise convex self-mappings. They are *mutually inverse* if $T \circ S(x) = S \circ T(x) = x$ for all $x \in X \setminus (\Gamma \cup \Delta)$.

Definition 4 *A piecewise convex transformation* (X, Γ, T) *is an invertible piecewise convex self-mapping. We will use the notation* $(X, \Gamma, T)^{-1} = (X, \Gamma_{-1}, T^{-1})$.

If (X, Γ) is a convex cell complex, we denote by $\mathcal{P}(\Gamma)$ the associated representation of X as the union of closed faces of Γ; we will refer to it as a *convex partition*.[2] If $\Gamma', \Gamma'' \subset X$ are convex graphs, their *join* $\Gamma' \vee \Gamma''$ is also a convex graph, and

$$\mathcal{P}(\Gamma' \vee \Gamma'') = \mathcal{P}(\Gamma') \vee \mathcal{P}(\Gamma''). \tag{1}$$

We outline a proof of this identity. A face, F, of $\Gamma' \vee \Gamma''$ is a connected component of $X \setminus (\Gamma' \cup \Gamma'')$. For $x, y \in F$ the chord $[x, y]$ avoids Γ', Γ''. Thus, $[x, y]$ belongs to unique faces F' of Γ' and F'' of Γ'', i. e., $F \subset F' \cap F''$. The converse also holds; thus $F = F' \cap F''$, implying equation (1).

Let (X, Γ, T) be a piecewise convex transformation. Setting $\Gamma_1 = \Gamma$ and $\Gamma_{n+1} = \Gamma_n \vee T^{-1}(\Gamma_n)$, we inductively define an increasing tower Γ_k, $k \geq 1$, of convex graphs on X. By construction, Γ_k is the singular set of T^k, and the piecewise convex transformation (X, Γ_k, T^k) is the kth iteration of T. Let $\mathcal{P}_k = \mathcal{P}(\Gamma_k)$. The set $S_\infty = \cup_{k=1}^{\infty} \Gamma_k$ is a countable (at most) union of chords. The complement $X_\infty = X \setminus S_\infty$ is the set of points $x \in X$ such that $x, Tx, T^2x, \ldots$ belong to open faces of Γ. We will refer to them as *regular points*. Iterating (X, Γ_{-1}, T^{-1}), we obtain the sequence of convex graphs Γ_{-k}, $k \geq 1$ and the piecewise convex transformations (X, Γ_{-k}, T^{-k}), inverse to (X, Γ_k, T^k).

Let $\mathcal{A} = \{a, b, \ldots\}$, $p = |\mathcal{A}|$, be a set labeling the faces of Γ, and let $\mathcal{L}$ be the *full shift space* on the *alphabet* $\mathcal{A}$. Assigning to a point $x \in X_\infty$ the sequence of labels of the faces of Γ containing $x, Tx, T^2x, \ldots$ we obtain the *coding map* $\sigma : X_\infty \to \mathcal{L}$. Set $\Sigma = \sigma(X_\infty)$, and let $\Sigma(n)$ be the set of words of length n that occur in Σ. The function $f(n) = |\Sigma(n)|$ is the *complexity* of (X, Γ, T). The proposition below summarizes the discussion.

Proposition 1 *Let* (X, Γ, T) *be a piecewise convex transformation. Then there is a sequence* Γ_k, $k \geq 1$ *of convex graphs in* X *such that the iterations of* T *correspond to piecewise convex transformations* (X, Γ_k, T^k). *There is a*

[2] It is not a set-theoretic partition of X; convex partitions provide an alternative approach to our material [11].

natural bijection between $\Sigma(n)$ and the set of faces of the convex graph Γ_n; the complexity of (X, Γ, T) satisfies $f(n) = |\mathcal{P}(\Gamma_n)|$.

1.3 Joins of convex graphs: A combinatorial formula

Let X be a compact space with a chord structure. Let $\Gamma', \Gamma'' \subset X$ be convex chord graphs. Set $\Gamma = \Gamma' \vee \Gamma''$.

Denote by $F', F'', F, E', E'', E, V', V'', V$ the sets of faces, edges and vertices of $\Gamma', \Gamma'', \Gamma$ respectively.[3] Let $e' \in E', e'' \in E''$ be arbitrary edges. If they intersect, then either they intersect transversally, or they overlap. The latter can occur in four ways. See figure 2. Denote by $c(\Gamma', \Gamma'')$ the number of overlappings.

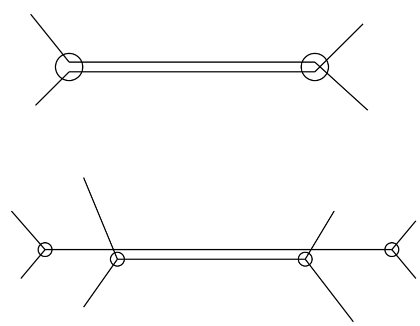

Figure 2: Overlapping of edges of two chord graphs

Lemma 2 *Let $\Gamma', \Gamma'' \subset X$ be convex chord graphs, and let $\chi = \chi(X)$ be the Euler number. Let V'_d, V''_d be the sets of vertices of Γ', Γ'' respectively, disjoint from the other graph. Set $V_{ess} = V \setminus (V'_d \cup V''_d)$. Then*

$$|F| - |F'| - |F''| + \chi = |V_{ess}| - c(\Gamma', \Gamma''). \tag{2}$$

Proof. Any convex chord graph $A \subset X$ satisfies $|F(A)| - |E(A)| + |V(A)| = \chi$. Using this identity, we obtain

$$|F| - |F'| - |F''| + \chi = (|E| - |E'| - |E''|) + (|V'| + |V''| - |V|). \tag{3}$$

[3] Recall that these are the *open cells* of a cell complex.

Denote by e'_i, e''_j the edges of Γ', Γ'' respectively. Let a'_i, a''_j be the number of vertices of Γ'', Γ' inside the edges e'_i, e''_j respectively. Let b'_i, b''_j be the number of times that e'_i, e''_j transversally intersects an edge of Γ'', Γ' respectively. Then e'_i, e''_j contribute $a'_i + b'_i + 1, a''_j + b''_j + 1$ edges to E respectively. Taking the overlapping into account, we obtain

$$|E| = \sum_i a'_i + \sum_i b'_i + |E'| + \sum_j a''_j + \sum_j b''_j + |E''| - c(\Gamma', \Gamma''). \qquad (4)$$

Let V_c be the set of common vertices of Γ', Γ''. Let V'_e, V''_e be the sets of vertices of Γ', Γ'' repectively that belong to edges of the other graph. Let $V_n = V \setminus (V' \cup V'')$ be the set of "new" vertices of Γ. Then

$$|V'| = |V'_e| + |V'_d| + |V_c|, |V''| = |V''_e| + |V''_d| + |V_c|, |V| = |V'| + |V''| - |V_c| + |V_n|. \qquad (5)$$

Besides

$$\sum_i a'_i = |V''_e|, \sum_j a''_j = |V'_e|, \sum_i b'_i = \sum_j b''_j = |V_n|. \qquad (6)$$

From equations (4-6), we have

$$|E| - |E'| - |E''| = |V'_e| + |V''_e| + 2|V_n| - c(\Gamma', \Gamma'').$$

Substituting this into equation (3), and using equation (5), we obtain the claim. ∎

The corollary below concerns a few special cases of Lemma 2.

Corollary 1 *Suppose, in addition to the assumptions of Lemma 2, that* $\chi = 1$*, and that the edges of graphs* Γ', Γ'' *do not overlap. Then*

$$|F| - |F'| - |F''| + 1 = |V_{ess}| \qquad (7)$$

and

$$|V_n| \leq |F| - |F'| - |F''| + 1 \leq |V|. \qquad (8)$$

If $V'_d = V''_d = \emptyset$ *then*

$$|F| - |F'| - |F''| + 1 = |V|. \qquad (9)$$

Proof. The claims are immediate from Lemma 2 and equation (5). ∎

1.4 Complexity of piecewise convex transformations: a geometric formula

Let Σ be a language on a finite alphabet $\mathcal{A}$, and let $\Sigma(n)$ be the set of words of length n in Σ. The *complexity* of Σ is the function $f(n) = |\Sigma(n)|$. Let $\varphi(n) = f(n+1) - f(n)$ and $\psi(n) = \varphi(n+1) - \varphi(n)$ be the first and second "derivatives" of complexity, respectively.

For any $w \in \Sigma$ let $m_l(w), m_r(w), m_b(w)$ be the number of extensions of w of the type aw, wb, awb respectively. Assume that $m_l(w), m_r(w) \geq 1$ for any $w \in \Sigma$. A word is *bispecial* if $m_l(w), m_r(w) > 1$. Let $\mathcal{B} \subset \Sigma$ be the set of bispecial words, and set $\mathcal{B}(n) = \mathcal{B} \cap \Sigma(n)$. Define the *Cassaigne index* [4] by

$$\mu(w) = m_b(w) - m_l(w) - m_r(w) + 1. \tag{10}$$

Note that $\mu(w) = 0$ if w is not bispecial. Then we have [4]

$$\psi(n) = \sum_{w \in \mathcal{B}(n)} \mu(w) = \sum_{w \in \Sigma(n)} \mu(w). \tag{11}$$

Set $\mu(n) = \sum_{w \in \Sigma(n)} \mu(w)$ for $1 \leq n$ and $\mu(n) = 0$ for $n \leq 0$. Set $M(n) = \sum_{k \leq n} \mu(k)$.

Lemma 3 *The complexity of a language satisfies*

$$f(n) = f(1) + (n-1)(f(2) - f(1)) + \sum_{k \leq n-2} M(k). \tag{12}$$

Proof. "Integrate" equation (11). ∎

Let Σ be the coding language of (X, T, Γ). For $w \in \Sigma(n)$ let $X(w) \subset X$ be the corresponding 2-cell of Γ_n. Let $\Gamma(w)$ be the restriction of $\Gamma_{-1} \vee \Gamma_{n+1}$ to $X(w)$. Let $V_{\text{ESS}}(w)$ (resp. $OE(w)$) be the set of essential vertices (resp. edge overlappings) for $\Gamma(w)$.

Lemma 4 *For any $w \in \Sigma$ we have*

$$\mu(w) = |V_{ess}(w)| - |OE(w)|. \tag{13}$$

Proof. Let Γ', Γ'' be the restrictions of $\Gamma_{n+1}, \Gamma_{-1}$ to $X(w)$. In equation (2) we have $m_b(w) = |F|, m_r(w) = |F'|, m_l(w) = |F''|, \chi = 1$. Lemma 2 yields the claim. ∎

For $n \geq 1$ set

$$V_{\text{ess}}(n) = \cup_{w \in \Sigma(n)} V_{\text{ess}}(w),\ v(n) = |V_{\text{ess}}(n)|,\ V(n) = \sum_{k \leq n} v(k); \tag{14}$$

$$OE(n) = \cup_{w \in \Sigma(n)} OE(w),\ c(n) = |OE(n)|,\ C(n) = \sum_{k \leq n} c(k). \tag{15}$$

For $k \leq 0$ set $V(k) = C(k) = 0$. Thus, $v(n)$ (resp. $c(n)$) is the number of essential vertices (resp. edge overlappings) of the join $\Gamma_{-1} \vee \Gamma_{n+1}$.

Proposition 2 *Let (X, T, Γ) be a piecewise convex transformation and set $\mathcal{P}_k = \mathcal{P}(\Gamma_k)$. Then the complexity of (X, T, Γ) satisfies*

$$f(n) = |\mathcal{P}_1| + (n-1)(|\mathcal{P}_2| - |\mathcal{P}_1|) + \sum_{k \leq n-2} V(k) - \sum_{k \leq n-2} C(k). \tag{16}$$

Proof. Combine Lemma 3 with Lemma 4. ∎

2 Billiard map as a piecewise convex transformation: the dictionary

Let M be a complete riemannian surface, and let $P \subset M$ be a connected, compact domain with a piecewise smooth boundary. The billiard flow in P is a particular case of the geodesic flow of a riemannian manifold (with a boundary and corners, in general). The boundary, ∂P, provides the standard cross-section; the corresponding Poincare map is the *billiard in P*. We refer to [16] for details. We will restrict our attention to the case where the curvature $\varkappa = \varkappa(M)$ is constant,[4] and ∂P is a finite union of geodesic segments. This is the *billiard in a geodesic polygon* on a surface of constant curvature.

[4]We normalize $\varkappa = 0, -1, 1$ and refer to these cases as parabolic, hyperbolic and elliptic respectively.

Let $\tilde{M}$ be the univeral covering,[5] let $q : \tilde{M} \to M$ be the projection, let $F \subset \tilde{M}$ be a fundamental domain, and set $P_F \subset F$ be the preimage of P. The geodesic polygon is *tame* if $q : P_F \to P$ is a bijective isometry. We will assume that P is tame.[6] Hence $P \subset M$ is a geodesic polygon, where $M = \mathbf{R^2}, \mathbf{H^2}, \mathbf{S^2}$ respectively in our three cases.

In order to proceed in a uniform fashion, we will use the *projective models* of the three geometries at hand: in a projective model the geodesics are straight lines. For $\mathbf{H}^2$ this is the Klein-Beltrami model [2]; a projective model of the elliptic geometry is obtained via the central projection of $\mathbf{S^2}$ onto a plane. (We make the technical assumption that P is contained in a hemisphere.) Thus, in all three cases, $P \subset \mathbf{R^2}$ is a euclidean polygon. In the hyperbolic case, we have the extra condition $P \subset D$ where D is the unit disc. Let $\mathcal{L} = \mathcal{L}(P)$ be the set of rays intersecting P. The chord structure on $\mathcal{L}$ is induced by the inclusion $\mathcal{L} \subset \mathcal{R}$.[7] See Examples 2, 3. Let $\mathcal{S}$ (resp. $\mathcal{C}$) be the set of sides (resp. corners) of P. Let $\Lambda \subset \mathcal{L}$ be the set of rays intersecting $\mathcal{C}$.

Proposition 3 *Let $P \subset \mathbf{R^2}$ be an arbitrary p-gon, let $\mathcal{L} = \mathcal{L}(P)$ be the set of rays intersecting P, and let $\Lambda = \Lambda(P) \subset \mathcal{L}$ be the set of rays intersecting $\mathcal{C}$. Then $\mathcal{L}$ is a closed annulus, and $(\mathcal{L}, \Lambda)$ is a convex cell complex. It has at most $p(p-1)$ vertices, at most $2p(p-1)$ edges, and less than $2p(p-1)$ faces.*

Proof. Let $\tilde{P} \subset \mathbf{R^2}$ be the convex hull of P. Then $\mathcal{L} = \mathcal{L}(\tilde{P})$, and for any bounded, convex domain $\Omega \subset \mathbf{R^2}$ the space $\mathcal{L}(\Omega)$ is a topological annulus [15].

For $o \in \mathcal{C}$ let $\Lambda_o \subset \mathcal{L}$ consist of rays containing o. Then $\Lambda_o \subset \Lambda$ is a topological circle; if $o \in \mathcal{C}$, then Λ_o is the union of chords. Thus, $\Lambda = \cup_{o \in \mathcal{C}} \Lambda_o$ is a chord graph. Let $F \subset \mathcal{L}$ be a two-cell, i. e., a connected component of $\mathcal{L} \setminus \Lambda$. It suffices to show that for any $l_0, l_1 \in F$ the chord $[l_0, l_1] \subset F$.

For $o \in \mathcal{C}$ and $l \in \mathcal{L} \setminus \Lambda_o$, we define the *sign of o with respect to l* by $\mu_l(o) = 1$ (resp. $\mu_l(o) = -1$) if o is on the left (resp. right) of l. By continuity, for any $o \in \mathcal{C}$ the sign $\mu_l(o)$ is the same for all $l \in F$.

Suppose that $l_0, l_1 \in F$ are anti-parallel, and let $C \subset \mathbf{R^2}$ be the closed strip between them. The equality $\mu_{l_o}(o) = \mu_{l_1}(o)$ holds iff $o \in C$. The

[5]In the three cases at hand $\tilde{M} = \mathbf{R^2}, \mathbf{H^2}, \mathbf{S^2}$.

[6]We make this assumption for simplicity of exposition; our results remain valid, mutatis mutandis, without it.

[7]In the hyperbolic case we have $\mathcal{L} \subset \mathcal{R}_{\mathbf{H}} \subset \mathcal{R}$. The chord structures on $\mathcal{L}$ induced from $\mathcal{R}_{\mathbf{H}}, \mathcal{R}$ coincide.

absence of corners of P in $\mathbf{R^2} \setminus C$ implies that $l_0, l_1 \notin \mathcal{L}$, contrary to the assumption. Thus, l_0, l_1 cannot be anti-parallel. Either they intersect or are parallel.

Suppose that l_0, l_1 intersect, and let C be the cone defined in Example 2. Now we have $\mu_{l_o}(o) = \mu_{l_1}(o)$ iff $o \in \mathbf{R^2} \setminus C$. Thus, C does not contain corners of P. If l_0, l_1 are parallel, we apply the same argument to the strip between l_0, l_1.

We have shown that $l_0, l_1 \in F$ implies that the rays either intersect or are parallel, and that the cone (resp. strip) between them is free of corners of P. Thus, $[l_0, l_1] \subset F$.

It remains to estimate the numbers of cells of $(\mathcal{L}, \Lambda)$. Let v, e, f be the number of vertices, edges, faces respectively. The vertices correspond to the rays $l \in \Lambda$ passing through a pair of points of $\mathcal{C}$. Since there are $p(p-1)$ ordered pairs of corners, we have $v \leq p(p-1)$. The edges belong to the circles $\Lambda_o, o \in \mathcal{C}$. The edges of Λ_o are separated by the vertices corresponding to the pairs $(o, o'), (o', o)$, where $o' \neq o$. Thus Λ_o consists of at most $2(p-1)$ edges. Since every edge belongs to a unique Λ_o, we have $e \leq 2p(p-1)$. By Euler's formula, $f = e - v$; hence $f < e$. ∎

We will define the billiard phase space by a sequence of steps. A *segment* is an oriented line segment, $x = [b, e\rangle$, $b \neq e$. Let $l \in \mathcal{L}$ be the ray containing $[b, e\rangle$ and having the same direction. A *billiard segment* is a segment such that $[b, e\rangle \subset P$ and $b, e \in \partial P$. Let $X_0 = X_0(P)$ be the set of billiard segments, and set $b = \beta(x), e = \eta(x), l = \lambda_0(x)$. The map $\beta \times \eta \times \lambda_0 : X_0 \to \partial P \times \partial P \times \mathcal{L}$ is injective; from here on we identify X_0 with its image $(\beta \times \eta \times \lambda_0)X_0 \subset \partial P \times \partial P \times \mathcal{L}$. We endow X_0 with the induced topology.

Let $X_1 \subset \partial P \times \partial P \times \mathcal{L}$ be the closure of $(\beta \times \eta \times \lambda_0)(X_0)$. Elements of the boundary $\partial X_0 = X_1 \setminus X_0$ arise from sequences of billiard segments that converge to degenerate limits. The mapping λ_0 extends to X_1 by continuity; we denote this extension by $\lambda_1 : X_1 \to \mathcal{L}$. In the representation $X_1 \subset \partial P \times \partial P \times \mathcal{L}$ the map λ_1 is the projection on the last coordinate. It is continuous and surjective.

A billiard segment may (properly) contain other billiard segments. For $x', x'' \in X_0$ set $x' \sim x''$ iff x', x'' are contained in the same billiard segment, see figure 3. This is an equivalence relation on X_0; it extends, by continuity, to X_1.

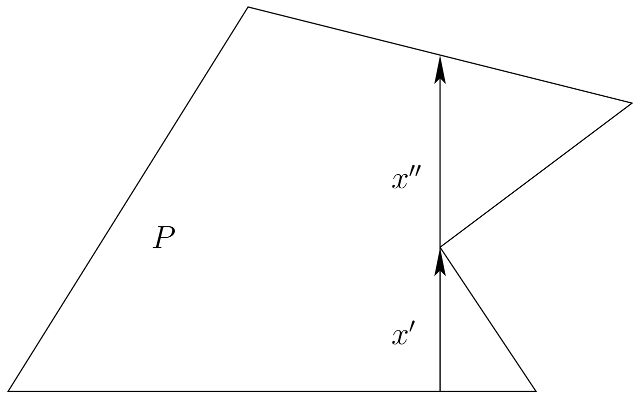

Figure 3: Equivalence relation for billiard segments

Definition 5 *The quotient of X_1 by this equivalence relation, endowed with the quotient topology, is the* billiard phase space $X = X(P)$.

For $x \in X_1$ we denote by $\{x\} \in X$ its equivalence class. By definition of the equivalence relation, the mapping $\lambda_1 : X_1 \to \mathcal{L}$ descends to a continuous, surjective mapping $\lambda : X \to \mathcal{L}$. We will now turn to the billiard map and its inverse.

Let $x_0 = [b, e\rangle \in X_0$ be a billiard segment. The (inverse) billiard map $x_0 \mapsto T_{\text{bil}}(x_0) = [e, e_1\rangle$ (resp. $x_0 \mapsto T_{\text{bil}}^{-1}(x_0) = [b_{-1}, b\rangle$) is defined, unless $e \in \mathcal{C}$ (resp. $b \in \mathcal{C}$) or b, e belong to the same side of P. Let Γ_0^+, Γ_0^- denote these sets. We view X_0 as a subset of X_1, and let $\Gamma_1^+ = \overline{\Gamma_0^+}, \Gamma_1^- = \overline{\Gamma_0^-} \subset X_1$ be their closures. We view $T_{\text{bil}}, T_{\text{bil}}^{-1}$ as self-mappings of X_1 not defined on Γ_1^+, Γ_1^- respectively; besides, they are not defined on $X_1 \setminus X_0$.

Let $q : X_1 \to X$ be the projection. Set $\Gamma_+ = q(\Gamma_1^+), \Gamma_- = q(\Gamma_1^-)$; let $\Gamma_q = \{x \in X : |q^{-1}(x)| > 1\}$ and $\Gamma_\partial = q(X_1 \setminus X_0)$. Let $T, T^{-1} : X \to X$ be the push-downs of $T_{\text{bil}}, T_{\text{bil}}^{-1}$ respectively. By definition, T (resp. T^{-1}) is not defined on $\Gamma_+ \cup \Gamma_q \cup \Gamma_\partial$ (resp. $\Gamma_- \cup \Gamma_q \cup \Gamma_\partial$). Set $\Gamma = \Gamma_+ \cup \Gamma_- \cup \Gamma_q \cup \Gamma_\partial$.

Theorem 1 *Let $P \subset M$ be a geodesic polygon on a surface of constant curvature, let $X = X(P)$ be the billiard phase space.*
1. The pair (X, Γ) is a convex cell complex, and $\lambda : (X, \Gamma) \to (\mathcal{L}, \Lambda)$ is a (branched) covering of cell complexes.

2. The (partially defined) mappings T, T^{-1} yield piecewise convex transformations (X, Γ, T) and (X, Γ, T^{-1}). The set $\Gamma \subset X$ is the union of the discontinuities of T and T^{-1}.

Proof. 1. The mapping $\lambda : X \to \mathcal{L}$ is continuous, surjective, and $\Gamma = \lambda^{-1}(\Lambda)$. Indeed, Λ is the set of rays passing through $\mathcal{C}$; by definition, $x \in \Gamma$ iff there is a (possibly degenerate) billiard segment in $q^{-1}(x)$ that contains a corner of P. This holds iff $\lambda(x) \in \Lambda$. Hence $x \in \Gamma$ iff $\lambda(x) \in \Lambda$. By Lemma 3 and Lemma 1, it suffices to show that λ is a branched covering of topological spaces. For $l \in \mathcal{L} \setminus \Lambda$ the intersection $l \cap P$ is a finite disjoint union of billiard segments $x_i = [b_i, e_i\rangle, 1 \leq i \leq n(l)$. Projecting them to X, we obtain $n(l)$ phase points; we denote them by x_i as well. Thus, $\lambda^{-1}(l) = \{x_1, \ldots, x_{n(l)}\} \subset X \setminus \Gamma$. The number $n(l)$ is constant on connected components of $\mathcal{L} \setminus \Lambda$. Hence, for any two-cell, $G \subset \mathcal{L}$, we have $\lambda^{-1}(G) = \cup_{i=1}^{n(G)} F_i$; the restriction of λ to every F_i is a homeomorphism, $\lambda : F_i \to G$.

Let now $E \subset \Lambda$ be an edge. A similar argument shows that $\lambda^{-1}(E) = \cup_{i=1}^{n(E)} e_i$, a disjoint union, where $e_i \subset \Gamma$, and the restriction of λ to every e_i is a homeomorphism, $\lambda : e_i \to E$. This verifies the assumptions of Lemma 1, hence the claim.

2. Let $G \subset \mathcal{L}$ be a two-cell, and let $F \subset X$ be one of the components of $\lambda^{-1}(G)$. In the proof of claim 1 we have identified F with a subset of X_0. Let $F = \{x = [b(x), e(x)\rangle\}$. Moreover, all points $b(x)$ (resp. $e(x)$) belong to the interior of a side, $s_{b(F)}$ (resp. $s = s_{e(F)}$) of P; furthermore, $s_{b(F)} \neq s_{e(F)}$. Let $\sigma_s \in \mathrm{Iso}(M)$ be the geodesic reflection about s. The billiard segments $T_{\mathrm{bil}}(x) = [e(x), e_1\rangle$ are well defined, unless the reflected ray $l_1 = \sigma_s(\lambda(x_0))$ passes through a *nonconvex corner* of P. See figure 4. In this case, the billiard segment $T_{\mathrm{bil}}(x_0)$ is not defined, and T_{bil} is discontinuous at x_0. Nevertheless, the phase point $\{T_{\mathrm{bil}}(x_0)\} = T(x_0) = x_1 \in X$ is well defined. By definition of the quotient topology, T is continuous at x_0. This proves the continuity of T on F, and hence on $X \setminus \Gamma$.

The same argument works for T^{-1}, but we will present an alternative proof. The claim will follow from a special symmetry of the billiard. Let $x_0 = [b, e\rangle \in X_0$. We define the *direction reversing involution* $\rho_0 : X_0 \to X_0$ by $\rho_0(x_0) = [e, b\rangle \in X_0$. It is continuous, and it extends by continuity to $\rho_1 : X_1 \to X_1$, which descends to the involution $\rho : X \to X$. For $l \in \mathcal{L}$ let l' be the same line with the opposite direction. Then $l \mapsto l'$ defines the involution $r : \mathcal{L} \to \mathcal{L}$ on the space of rays. The mappings λ_0, λ_1, and $\lambda : X \to \mathcal{L}$ commute with the respective involutions. In particular, $\lambda \circ \rho = r \circ \lambda$. The

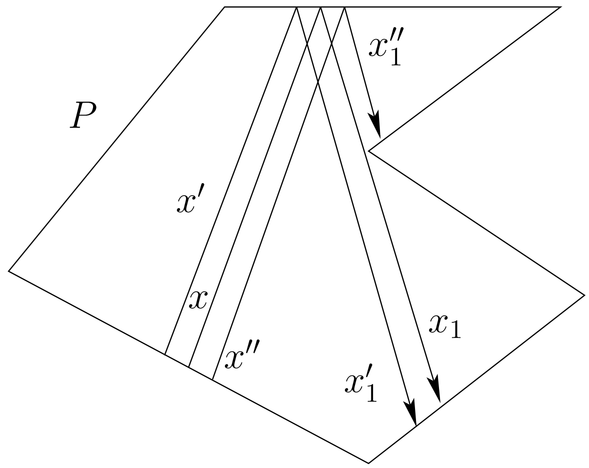

Figure 4: Pushing the billiard map down to the quotient phase space; the phase points x_1', x_1'', x_1 are close to each other

billiard symmetry relates the maps T_{bil}^{-1}, T_{bil}, and ρ_0. It says that whenever both expressions are defined, we have $T_{\text{bil}}^{-1} = \rho_0 \circ T_{\text{bil}} \circ \rho_0$.

Note that T, T^{-1} are defined and continuous on $X \setminus \Gamma$, and that ρ preserves Γ; hence, we have the identity

$$\rho \circ T = T^{-1} \circ \rho. \tag{17}$$

We have shown that the dicontinuities of both T and T^{-1} belong to Γ. It remains to show the opposite inclusion. Recall that $\Gamma = \Gamma_+ \cup \Gamma_- \cup \Gamma_q \cup \Gamma_\partial$. Neither T nor T^{-1} are defined on Γ_∂. The map T (resp. T^{-1}) is not defined on Γ_+ (resp. Γ_-). Let $x \in \Gamma_q$, i. e., $|q^{-1}(x)| > 1$. Then there is a billiard segment $[b, e\rangle \in q^{-1}(x)$ such that either $b \in \mathcal{C}$ or $e \in \mathcal{C}$. Hence, either $T_{\text{bil}}([b, e\rangle)$ or $T_{\text{bil}}^{-1}([b, e\rangle)$ is not defined. Therefore, at least one of T, T^{-1} is not defined on x, yielding a discontinuity.

Let now $F \subset X \setminus \Gamma$ be a two-cell, and let $x_0, x_1 \in F$. Set $G = \lambda(F) \subset \mathcal{L}$, and $l_0 = \lambda(x_0), l_1 = \lambda(x_1) \in G$. Then $\lambda^{-1}(G) = \cup_{i=1}^{n(G)} F_i$, and without loss of generality, $F = F_1$. By the proof of claim 1, l_0, l_1 either intersect or are parallel. In either case we have $[l_0, l_1] \subset G$; the rays $l_t \in [l_0, l_1]$ intersect P forming $n(G)$ disjoint curves; each curve is a chord, γ_i, in $F_i, 1 \leq i \leq n(G)$. The chord γ_1 joins x_0 with x_1. Denote by $[b_t, e_t\rangle$ the corresponding family of billiard segments. For $0 \leq t \leq 1$ the points e_t belong to the interior of a side,

$s \in \mathcal{S}$. The reflected pencil $\{\sigma_s(l_t) : 0 \leq t \leq 1\}$ is the chord $[\sigma_s(l_0), \sigma_s(l_1)]$. Applying the preceding argument to $[\sigma_s(l_0), \sigma_s(l_1)]$, we see that the intersection of $T(\gamma_1) \subset X$ with any face of X is a chord. Since F is an arbitrary face of X, we have established that T is a piecewise convex transformation (X, Γ, T). Equation (17) yields the claim for T^{-1}. ■

Lemma 5 *Let P be a polygon with p corners, and let (X, Γ) be the corresponding convex cell complex. Then (X, Γ) has at most $2p^3$ faces.*

Proof. A typical ray $l \in \mathcal{L}$ intersects ∂P in at most p points. These points partition l into at most $p+1$ intervals; at most $p-1$ of them are billiard segments. Hence, the degree of the covering $\lambda : X \to \mathcal{L}$ is not greater than $p-1$. By Proposition 3, the number of faces of $\mathcal{L}$ is less than $2p(p-1)$. The product $(p-1) \times 2p(p-1)$ provides the desired estimate. ■

Let $P \subset M$ be a polygon. Theorem 1 associates with the billiard in P a piecewise convex transformation (X, Γ, T). We will establish a dictionary between the billiard in P and (X, Γ, T). In particular, we will interprete properties of billiard orbits in terms of the notions introduced in section 1.

A billiard segment $s = [b, e\rangle$ is *regular* (resp. *singular*) if it does not contain (resp. contains) corners of P. A billiard orbit is a finite sequence of billiard segments; their number is the *length* of the orbit. Let $s_0, s_1, \ldots, s_n$ be a sequence of billiard segments. It is a billiard orbit iff the segments $s_i, 0 < i < n$, are regular and $s_{i+1} = T_{\text{bil}}(s_i), 0 \leq i < n$. To distinguish billiard orbits from arbitrary sequences of billiard segments, we will use the notation like $\omega = (s_0, s_1, \ldots, s_n)$. The orbit $(s_0, s_1, \ldots, s_n)$ is regular (resp. singular; resp. strongly singular) if the segments s_0, s_n are also regular (resp. at least one of s_0, s_n is singular, resp. both s_0, s_n are singular).[8]

Analogously, a sequence $(x_0, x_1, \ldots, x_n) \subset X$ is a regular (resp. singular, resp. strongly singular) *phase orbit* if all x_i are regular and $T(x_i) = x_{i+1}$, $0 \leq i \leq n$ (resp. either x_0 or x_n is singular, and $x_0 = T^{-1}(x_1)$ in the former case; resp. both x_0, x_n are singular, with the convention above). We will also say that $(x_0, x_1, \ldots, x_n)$ is an *orbit of* (X, Γ, T).

A *generalized diagonal* $\omega = (s_0, s_1, \ldots, s_n)$ is a strongly singular billiard orbit such that $b(s_0)$ (resp. $e(s_n)$) is a corner of P, and neither segment contains other corners.

[8]Note that, by definition, a strongly singular billiard orbit has length at least three.

Let $\omega(0) = (s_0(0), s_1(0), \dots, s_n(0))$ and $\omega(1) = (s_0(1), s_1(1), \dots, s_n(1))$ be two billiard orbits of the same length. Suppose that for $0 \le i \le n$ the chord $\gamma_i = [s_i(0), s_i(1)]$ exists, and for $0 \le i \le n-1$ we have $T_{\text{bil}}(\gamma_i) = \gamma_{i+1}$. This defines the *chord of billiard orbits* $[\omega(0), \omega(1)]$ joining $\omega(0), \omega(1)$.

Let $\omega(0) = (s_0(0), s_1(0), \dots, s_n(0)), \omega(1) = (s_0(1), s_1(1), \dots, s_n(1))$ be generalized diagonals such that the chord $[\omega(0), \omega(1)]$ exists, and $[\omega(0), \omega(1)] = \{\omega(t) : 0 \le t \le 1, \}$ where all $\omega(t)$ are generalized diagonals. Then $[\omega(0), \omega(1)]$ is a *chord of generalized diagonals.*

Definition 6 *A generalized diagonal ω is* isolated *if it is not contained in a chord of generalized diagonals.*

A chord is *maximal* if it is not properly contained in a longer chord. Every chord of generalized diagonals is contained in a unique *open maximal chord* $\{\omega(t) : 0 < t < 1\}$. Since the limits $\omega(0), \omega(1)$ contain corners in their interior, they are not generalized diagonals.

In what follows we refer to the correspondence established in the discussion above as the *dictionary* (between billiard orbits and phase space orbits).

Proposition 4 *Let $P \subset M$ be a geodesic polygon on a surface of constant curvature, let X be the billiard phase space, and let (X, Γ, T) be the associated piecewise convex transformation. Let $n \ge 1$. Then the following holds:*
1. The dictionary establishes a bijection between the set $V_{ess}(n) \subset X$ and the set of isolated generalized diagonals of length $(n+2)$ in P;
2. The dictionary establishes a bijection between the set $OE(n) \subset X$ and the set of maximal chords of generalized diagonals of length $(n+2)$ in P.

Proof. Let $x_0 \in X \setminus \Gamma$. Then x_0 does not contribute to $V_{\text{ess}}(n)$ or to $OE(n)$ unless the following conditions hold:
i) The points $x_i = T^i(x_0) \in X \setminus \Gamma$ for $1 \le i \le n-1$;
ii) The points $x_n = T^n(x_0), x_{-1} = T^{-1}(x_0) \in \Gamma$.

Suppose that these conditions hold. Denote by $\text{ch}_+ \subset \Gamma$ (resp. $\text{ch}_- \subset \Gamma$) the chord containing x_n (resp. x_{-1}). Let $\gamma_n \subset T^{-n}\text{ch}_+$ (resp. $\gamma_{-1} \subset T\text{ch}_-$) be the maximal chords contained in the respective pull-backs, and containing x_0. Then $x_0 \in V_{\text{ess}}(n)$ (resp. x_0 contributes to $OE(n)$, increasing $OE(n)$ by 1) iff the chords γ_{-1}, γ_n intersect transversally (resp. the chords γ_{-1}, γ_n overlap).

By Theorem 1, $x_0, \dots, x_{n-1}$ defines a regular billiard orbit, $(s_0, \dots, s_{n-1})$. The billiard segments $s_{-1} = T_{\text{bil}}^{-1}(s_0), s_n = T_{\text{bil}}(s_{n-1})$ are singular. By choosing them appropriately, we can assume without loss of generality that s_{-1} (resp. s_n) begins (resp. ends) in a corner.

Let $p_{-1}(t), -\alpha < t < \beta$, and $q_n(t), -\gamma < t < \delta$, be the chords of billiard segments corresponding to ch_-, ch_+ respectively. We normalize the parameter t so that $p_{-1}(0) = s_{-1}, q_n(0) = s_n$. Applying the billiard map, we obtain two chords of billiard orbits: $(p_{-1}(t), p_0(t)), -\alpha < t < \beta$, and $(q_0(t), q_1(t), \dots, q_n(t)), -\gamma < t < \delta$. By construction, $p_0(0) = q_0(0) = s_0$. We denote the two chords of billiard orbits by $[p], [q]$ respectively.

The phase space cord $\gamma_{-1} \subset T(\text{ch}_-)$ (resp. $\gamma_n \subset T^{-n}(\text{ch}_+)$) corresponds to the billiard segment chord $p_0(t), -\alpha < t < \beta$, (resp. $q_0(t), -\gamma < t < \delta$). The chords γ_{-1}, γ_n are transversal at x_0 iff the billiard segment chords $p_0(t), -\alpha < t < \beta$, $q_0(t), -\gamma < t < \delta$, that coincide at $t = 0$, have no other common billiard segments. Equivalently, the billiard orbit chords $[p], [q]$ fit together forming the generalised diagonal ω only at $t = 0$, and not for $t \neq 0$. This happens iff ω is isolated.

The phase space chords γ_{-1}, γ_n overlap at x_0 iff we have $p_0(t) = q_0(t)$ for $-\varepsilon < t < \varepsilon$. Equivalently, the billiard orbit chords $[p], [q]$ fit together for $-\varepsilon < t < \varepsilon$, forming a chord of generalised diagonals. ■

3 Complexity of the billiard in a geodesic polygon

We will now apply the preceding material to the complexity of billiards in geodesic polygons on surfaces of constant curvature. First, we consider the three cases simultaneously, emphasizing their similarities.

3.1 Arbitrary curvature, any polygon

Let $P \subset M$ be a geodesic polygon. Let $G = G(P) \subset \text{Iso}(M)$ be the group generated by the geodesic reflections in the sides of P. Any $g \in G$ is represented by a word whose letters are the generators; the *length* is the minimal number of letters [7]. Let $G^{(n)} \subset G$ be the set of elements of length at most n. Then $G^{(n)} \subset G^{(n+1)}$, and $G = \cup_{n=0}^{\infty} G^{(n)}$.

The operation of *unfolding* sends billiard orbits in P into geodesics in M [16]. See figure 5. Let $\gamma = (x_0, x_1, \dots, x_n)$ be a billiard orbit. Its unfolding is the geodesic $\tilde{\gamma} = (x_0, \tilde{x}_1, \dots, \tilde{x}_n)$; there is $g = g(\gamma) \in G^{(n)}$ such that $\tilde{x}_n = g \cdot x_n$.

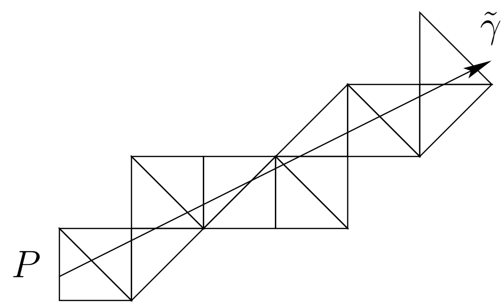

Figure 5: Unfolding a billiard trajectory

Lemma 6 1. *If $\varkappa(M) \le 0$, then all generalized diagonals in P are isolated.* 2. *Let $\varkappa(M) > 0$. Let $\gamma = (x_0, x_1, \dots, x_n)$ be a generalized diagonal, and let $c_0, c_n \in \mathcal{C}$ be its endpoints. Let $\tilde{\gamma} = (x_0, \tilde{x}_1, \dots, \tilde{x}_n)$ be the unfolding of γ, and let $\tilde{c}_n = g(\gamma) \cdot c_n \in M$ be the ending point of $\tilde{\gamma}$. Then γ belongs to a chord of generalized diagonals iff the points $c_0, \tilde{c}_n$ coincide or are antipodal.*

Proof. Let $\gamma(t) = (x_0(t), x_1(t), \dots, x_n(t))$, $-\varepsilon \le t \le \varepsilon$, be a chord of generalized diagonals, and let $\tilde{\gamma}(t) = (x_0(t), \tilde{x}_1(t), \dots, \tilde{x}_n(t))$, $-\varepsilon \le t \le \varepsilon$, be the corresponding chord of geodesics. The elements of G produced by the unfolding do not depend on the chord parameter. In particular, $g(\gamma(t)) = g$ for all t. Thus, $\tilde{c}_n(t) = g \cdot c_n \in M$, hence all geodesics $\tilde{\gamma}(t)$ begin at c_0 and end at $g \cdot c_n$.

The preceding argument is reversible. Therefore, a family $\gamma(t)$ of $(n+1)$-segment generalized diagonals is a chord iff the unfolded family $\tilde{\gamma}(t)$ is a beam of geodesics in M, emanating from $c_0 \in \mathcal{C}$ and refocusing at $g \cdot c_n \in M$. If $\varkappa \le 0$, this is impossible, implying claim 1. Let $\varkappa > 0$, i.e., M is the sphere. A beam emanating from a point refocuses at the antipodal point and at the initial point. ∎

We denote by $gd(k)$ (resp. $cgd(k)$) the number of isolated (resp. maximal chords of) k-segment generalized diagonals in P. Set

$$GD(n) = \sum_{3 \le k \le n} gd(k),\ CGD(n) = \sum_{3 \le k \le n} cgd(k).$$

Thus, $GD(n)$ (resp. $CGD(n)$) is the number of isolated (resp. maximal chords of) generalized diagonals of length at most n. (Recall that the minimal length of a generalized diagonal is three.)

Theorem 2 *Let $P \subset M$ be a geodesic polygon, and let $\varkappa$ be the curvature of M. Let (X, T, Γ) be the piecewise convex transformation associated with the billiard in P, and let $f_\Gamma(\cdot)$ be the corresponding complexity. Then there are integers q_1, q_2 depending on P, so that the following holds.*
1. *For $\varkappa = 1$ we have*

$$f_\Gamma(n) = q_1 + q_2 n + \sum_{k \le n} GD(k) - \sum_{k \le n} CGD(k). \tag{18}$$

2. *If $\varkappa \le 0$, then*

$$f_\Gamma(n) = q_1 + q_2 n + \sum_{k \le n} GD(k). \tag{19}$$

Proof. By Theorem 1 and Proposition 4, equation (16) yields the first claim. Combining it with Lemma 6, we obtain equation (19). ∎

It is traditional to code billiard orbits in P by the sides that they visit. This is the *side coding*, and the corresponding complexity is the *side complexity*, $f_{\text{side}}(n)$. We will define it. Recall that $\mathcal{S}$ is the set of sides of P. Let $\omega = (s_1, \dots, s_n)$ be a regular billiard orbit, and let $\sigma(\omega) = a_0 a_1 \dots a_n$ be the word of the *side coding language*; its letters are the elements of $\mathcal{S}$ that ω visited; here $a_0 = b(s_1), a_1 = b(s_2), \dots, a_n = e(s_n)$. The word $a_0 a_1 \dots a_n \in \Sigma_{\text{side}}(n+1)$ is a side code. We set $f_{\text{side}}(n) = |\Sigma_{\text{side}}(n+1)|$.

Proposition 5 *Let $P \subset M$ be a geodesic polygon, and let $\varkappa$ be the curvature of M. 1. If $\varkappa = 1$, then there are integers q_1, q_2 (depending on P) such that*

$$f_{side}(n) \le q_1 + q_2 n + \sum_{k \le n} GD(k) - \sum_{k \le n} CGD(k). \tag{20}$$

2. *If $\varkappa \leq 0$, then there are constants c_1, c_2 such that*

$$f_{side}(n) \leq c_1 + c_2 n + \sum_{k \leq n} GD(k). \tag{21}$$

3. *If P is a convex polygon, equations (20, 21) are equalities.*

Proof. Denote by $X_0(a,b) \subset X_0$ the set of regular billiard segments $[p,q\rangle$ such that $p \in a, q \in b$. Let $X(a,b) \subset X$ be the corresponding subset of the phase space. Every face of Γ is contained in some $X(a,b)$. Thus, any code by the faces of Γ defines a side code, yielding a surjection $\Sigma(n) \to \Sigma_{\text{side}}(n+1)$. Therefore $f_{\text{side}}(n) \leq f_\Gamma(n)$, and Theorem 2 yields claims 1, 2. Let now P be convex. The discussion of section 2 shows that $\lambda : X \to \mathcal{L}$ is an isomorphism of convex cell complexes; the faces of Γ are the sets $X(a,b) \subset X$, where $a, b \in \mathcal{S}$ are different sides of P. Thus, $f_{\text{side}}(n) = f_\Gamma(n)$, and claim 3 follows from Theorem 2. ■

We will now consider the parabolic, elliptic and hyperbolic cases.

3.2 The euclidean case

Billiard dynamics in euclidean polygons is a classical subject. Still, many basic questions remain open [10, 16]. The side complexity is subexponential [11];[9] it is plausible that the complexity has a polynomial upper bound. A subexponential upper bound on complexity has not been proved for general polygons. A euclidean polygon is *rational* if all of its angles are rational multiples of π. Billiard dynamics in rational polygons is of interest on its own; also, it has been used to study the billiard in generic polygons [16].

Theorem 3 *Let $P \subset \mathbf{R}^2$ be a rational euclidean polygon.* 1. *There exists $c = c(P) > 0$ such that*

$$f_{side}(n) < cn^3. \tag{22}$$

2. *Suppose, in addition, that P is convex. Then there exist positive numbers c_1, c_2 such that*

$$c_1 n^3 < f_{side}(n) < c_2 n^3. \tag{23}$$

[9]Implying that the billiard in any euclidean polygon has zero topological entropy.

Proof. By a theorem of Masur [14], for any rational polygon there exist positive constants c_1', c_2' such that

$$c_1' n^2 < GD(n) < c_2' n^2.$$

Proposition 5 implies both equation (22) and equation (23). ■

3.3 The elliptic case

The theorem below holds for all spherical polygons P; for simplicity of exposition, we will prove it under the assumption that P is an admissible polygon.

Theorem 4 *For any spherical polygon the side complexity grows subexponentially.*

Proof. Let $\mathrm{Geo}(\mathbf{S^2})$ be the space of oriented great circles, and let $\varphi : \mathrm{Geo}(\mathbf{S^2}) \to \mathbf{S^2}$ be the standard diffeomorphism.[10] It endows $\mathrm{Geo}(\mathbf{S^2})$ with an invariant riemannian metric. The geodesics in this metric are the chords in $\mathrm{Geo}(\mathbf{S^2})$. Let (X, T, Γ) be the associated piecewise convex transformation. By the proof of Lemma 3, the branched covering $\lambda : X \to \mathrm{Geo}(\mathbf{S^2})$ induces metrics on the faces of Γ; every face is a convex polygon.

The billiard map T_{bil} is a piecewise geodesic reflection; hence (X, T, Γ) is a piecewise isometry on a convex partition. By Theorem 4.2 of [11], $f_\Gamma(n)$ has subexponential growth. By the proof of Proposition 5, $f_{\mathrm{side}}(n) \leq f_\Gamma(n)$.■

By Lemma 6, a spherical polygon may have chords of generalised diagonals. The examples below illustrate other special features of the spherical billiard.

Example 4 Let $P \subset \mathbf{S^2}$ be a polygon such that $G(P)$ is a finite group.[11] Then every billiard orbit in P is periodic. Since the prime periodic orbits yield a finite number of symbolic codes, f_{side} is bounded.

Example 5 Let P_α be the "bigon", bounded by two geodesics, a, b connecting the North and the South poles, where α is the angle between them. (Note

[10] See section 4.2 for details.

[11] These polygons are classified. See, e.g., [6].

that bigons are not admissible polygons.) For any n the set $\Sigma_{\text{side}}(n)$ consists of 2 elements: $abab\dots$ and $baba\dots$. Thus, $f_{\text{side}}(n) = 2$. For any α the equator provides a 2-segment periodic orbit in P_α. If α is π-rational, then P_α fits Example 4, hence every regular orbit is periodic.

Claim. Let α be π-irrational. Then this equator is the only prime periodic orbit in P_α.

Proof. Denote by ρ_α the rotation by α about the z-axis. Let γ be a periodic orbit. We can assume that it has an even number, $2m$, of segments, and that its symbolic code is $ba\dots ba$. Then the isometry $g(\gamma)$ obtained by unfolding γ is $\rho_\alpha^{2m} \neq 1$.

Let $\ell(\gamma)$ be the spherical geodesic corresponding to γ. (Note that $\ell(\gamma)$ differs, in general, from the unfolding $\tilde{\gamma}$ which is a geodesic segment along $\ell(\gamma)$.) By periodicity of γ, $\ell(\gamma)$ is invariant under $g(\gamma)$. The only geodesic invariant under ρ_α^{2m} is the equator. ■

By convention, a periodic billiard orbit in P does not pass through its corners. In particular, it cannot trace the boundary of P. It is not known whether every euclidean polygon has a periodic orbit [10]. Our next example is a spherical polygon without periodic orbits.

Example 6 For $0 < \alpha < 2\pi$ let $Q = Q_\alpha$ be the isosceles spherical triangle with angle α and two right base angles.

Claim. If α is π-rational then every billiard orbit in Q is periodic. If α is irrational, then Q has no periodic billiard orbits.

Proof. The bigon P of Example 5 is obtained by doubling Q about the equator. Every billiard orbit, γ, in Q lifts to a billiard orbit $\tilde{\gamma}$ in P; the orbit γ is periodic iff so is $\tilde{\gamma}$. If α is π-rational, the claim holds, by Example 5. Let α be irrational, and let γ be a periodic orbit in Q. Since, by Example 5, $\tilde{\gamma}$ runs along the equator, γ traces the boundary of Q. ■

3.4 The hyperbolic case

It is not surpising that the complexity of the billiard in a hyperbolic polygon grows exponentially. We will obtain a more precise result in this direction. A positive function, $s(\cdot)$, of natural argument is *temperate* if for any $h > 0$ and all sufficiently large n we have $e^{-hn} < s(n) < e^{hn}$.

Theorem 5 *Let $P \subset \mathbf{H}^2$ be a geodesic polygon, and let h_{top} be the* topological entropy *of the billiard in P.*

Then $h_{\text{top}} > 0$; there exists a temperate function $s(\cdot)$ such that

$$f_{side}(n) = s(n)e^{h_{\text{top}}n}.$$

Proof. The billiard flow of P is (uniformly) hyperbolic [9].[12] Thus, the metric entropy of the billiard flow (with respect to the Liouville measure) is positive. By Abramov's formula [1], the metric entropy of the billiard map in P (with respect to the canonical measure) is positive as well. By the maximum principle, $h_{\text{top}} > 0$.

Let (X, T, Γ) be the piecewise convex transformation associated with the billiard in P, and let $\mathcal{P} = \mathcal{P}(\Gamma)$ be the corresponding convex partition of X. Let $\mathcal{Q}$ be the (possibly nonconvex) partition of X defined by the sides of P.

Let $\alpha(t), \beta(t)$ be infinite billiard orbits that visit the same sides of P for $-\infty < t < \infty$. Then their unfoldings, $\tilde{\alpha}(t), \tilde{\beta}(t)$ are infinite geodesics in $\mathbf{H}^2$; the distance between $\tilde{\alpha}(t), \tilde{\beta}(t)$ is bounded, as $-\infty < t < \infty$. Hence $\tilde{\alpha}(-\infty) = \tilde{\beta}(-\infty)$ and $\tilde{\alpha}(\infty) = \tilde{\beta}(\infty)$ implying $\tilde{\alpha} = \tilde{\beta}$. Therefore, $\alpha = \beta$. Hence $\mathcal{Q}$ is a generating partition.

Since $\mathcal{P} \prec \mathcal{Q}$, the defining partition $\mathcal{P}$ is generating as well. By [11], the complexity of (X, T, Γ) satisfies $f_\Gamma(n) = s_1(n)e^{h_{\text{top}}n}$, where s_1 is a temperate function. By the proof of Proposition 5, $f_{\text{side}}(n) \le s_1(n)e^{h_{\text{top}}n}$.

Let $\mathcal{P}_n = \mathcal{P}(\Gamma_n)$ (resp. $\mathcal{Q}_n$), $1 \le n$, be the sequence of partitions of X defined by (X, T, Γ) (resp. corresponding to the side coding). Then $\mathcal{P}_n \prec \mathcal{Q}_n$, and $f_{\text{side}}(n) = |\mathcal{Q}_n|, f_\Gamma(n) = |\mathcal{P}_n|$. Let A be an atom of $\mathcal{Q}_n$. Denote by $r_n(A)$ the number of atoms of $\mathcal{P}_n$ that partition A, and let $r_n = \max_{A \in \mathcal{Q}_n} r_n(A)$. Then $f_\Gamma(n) \le r_n f_{\text{side}}(n)$.

Let $(s_1, \ldots, s_n)$ be the side-code of A. The unfolding of P along a billiard orbit $\gamma \in A$ is obtained by reflecting P consecutively about $s_1, \ldots, s_n$. Hence it is determined by A; we denote it by $\tilde{P}_A$. Although $\tilde{P}_A$ is not a polygon, in general,[13] the basic concepts of polygonal billiard apply to it. In particular, the billiard orbits $\gamma \in A$ uniquely unfold into billiard segments $\tilde{\gamma}$ in $\tilde{P}$. Let $(\tilde{X}, \tilde{\Gamma})$ be the convex cell complex associated with the billiard in $\tilde{P}_A$. Then A determines an atom, $\tilde{A}$, of the side partition of $\tilde{X}$, and $r_n(A)$ is the number of faces of $\tilde{\Gamma}$ that partition $\tilde{A}$.

[12] This is a special case of a more general result in [9]. It can also be obtained directly.

[13] If P is convex, then $\tilde{P}_A \subset \mathbf{R^2}$ is a polygon. In general, $\tilde{P}_A$ is a flat surface with a boundary; the reader may think of $\tilde{P}_A$ as a *polygon in a (branched) covering of* $\mathbf{R^2}$.

If P has p corners, then the number of corners of $\tilde{P}_A$ is at most pn. Lemma 5 applies, hence $r_n(A) \leq 2p^3n^3$. Since A was arbitrary, we have $r_n \leq \mathrm{const} n^3$, and $\mathrm{const} n^{-3} s_1(n) e^{h_{\mathrm{top}} n} \leq f_\Gamma(n)$. Set $s(n) = f_\Gamma(n) e^{-h_{\mathrm{top}} n}$. ∎

4 Complexity of the outer billiard

Let (X, Γ) be a convex cell complex, and let (X, Γ, T) be a piecewise convex transformation. Suppose that X is a metric space, and the chords are geodesic segments. Assume that for any 2-cell, the restriction $T_F : \overline{F} \to X$ is an isometry. See Definition 4. Then (X, Γ, T) is a *piecewise convex isometry*. This is a special class of piecewise isometries, in general; there are many other special classes, for instance, interval exchange maps. These arise, in particular, from the billiard in rational polygons, and have been extensively studied [14].

We will investigate a particular subclass of piecewise convex isometries – the outer billiard transformations. Let $P \subset M$ be a convex geodesic p-gon, where M is a simply connected surface of constant curvature $\varkappa$. For $x \in M$ denote by $T_x : M \to M$ the geodesic symmetry about x. Let $a, b, c, \dots$ be the corners of P listed counterclockwise. If $\varkappa = 0, -1$, set $X = X(P) = M - P$. If $\varkappa = 1$ (i. e., $M = \mathbf{S^2}$), let P' be the antipodal polygon,[14] and set $X = X(P) = M - P - P'$.

For a corner, say a, of P, let $R_a \subset X$ be the geodesic ray extending the side ab in the direction of a. The set $X \subset M$ is a metric space, the chords are the geodesics in X, and the chord graph $\Gamma = R_a \cup R_b \cup \dots$ is convex. Set $\mathcal{P} = \mathcal{P}(\Gamma)$, and let $X_a, X_b, \dots$ be the closed 2-cells. See figure 6.

Definition 7 *The* outer billiard *about P is the piecewise convex isometry (X, Γ, T) such that the restrictions $T|_{X_a}, T|_{X_b}, \dots$ are the geodesic symmetries $T_a : X_a \to X, T_b : X_b \to X, \dots$. The space $X \subset M$ is the* phase space *of the outer billiard.*

We will use the notation $T : X \to X$ for the outer billiard.[15] The *complexity of the outer billiard* is the complexity of (X, Γ, T) with respect

[14] In this section we assume, as before, that P is an admissible polygon.

[15] If a danger of confusion with the inner billiard arises, we will use the superscript, $T^{\mathrm{out}} : X \to X$.

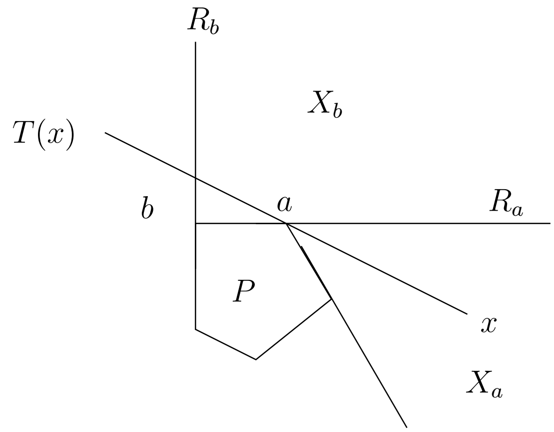

Figure 6: Definition of the outer billiard map

to the partition $\mathcal{P}$. Now we introduce the notation and terminology that will be used throughout this section. If $g(n), h(n)$ are two positive sequences, we write $g \prec h$ if there is a constant C such that for all n sufficiently large $g(n) \leq Ch(n)$. If $g \prec h$ and $h \prec g$, we write $g \sim h$; we will say that the sequences *have the same growth* or are in the *same (growth) class*. If $g \prec n^d$ then we say that g grows at most polynomially with degree d, or that g is bounded by n^d.

If G is a group with a finite set $S = \{s_1, \ldots, s_p\}$ of generators, we denote by $G_S^{(n)} \subset G$ the set of elements that can be represented by products of at most n elements of S and their inverses. The growth class of the sequence $g_S(n) = |G_S^{(n)}|$ does not depend on the choice of S [7]. If $g_S(n) \sim n^d$, then we say that the group G *grows polynomially, with degree d*. In our case, $G = G(P) \subset Iso(M)$ is the group generated by the set $S = \{T_a, T_b, \ldots\}$ of geodesic symmetries about the corners of P. We proceed to study the three cases at hand.

4.1 The euclidean case

We will obtain polynomial bounds on the complexity of outer billiard.

Theorem 6 *Let P be a convex euclidean p-gon, and let $f(\cdot)$ be the complexity of the outer billiard about P. Then $n \prec f(n) \prec n^{p+1}$.*

Proof. The edges of the graph Γ_n are parallel to the sides of P; each edge is a segment or a half-line. Assume, for simplicity of exposition, that P has no parallel sides. Then there are p directions. For each direction there are n parallel half-lines, hence their total number is pn. Since they partition X into pn components, the number of faces of Γ_n is at least pn. This yields the linear lower bound on complexity.[16] ∎

Let $G = G(P)$, and let $S = \{T_1, \dots, T_p\}$ be the natural set of generators. For the proof of the upper bound, we will need a few lemmas.

Lemma 7 *The growth of G is bounded by n^{p-1}.*

Proof. The subgroup $H \subset G$ generated by $T_1T_p, T_2T_p, \dots, T_{p-1}T_p$ is a quotient group of $\mathbf{Z}^{p-1}$; hence its growth is bounded by n^{p-1}. Since H is a normal subgroup of G of index 2, the two groups have the same growth. ∎

Let $\Gamma = \partial\mathcal{P}$, and let $\Gamma_1, \Gamma_2, \dots$ be the canonical sequence of graphs; see section 1. Let γ_n be the set of edges of $\Gamma_n \setminus \Gamma_{n-1}$.

Lemma 8 *The first difference of the sequence $|\gamma_n|$ is bounded by n^{p-1}.*

Proof. The edges of γ_{n+1} are obtained from the edges of γ_n by applying T^{-1}. Each time a singularity half-line of T^{-1} intersects an edge of γ_n, this edge splits into two, and thus contributes 1 to $|\gamma_{n+1}| - |\gamma_n|$.

Let L_n be the set of straight lines obtained by reflecting at most n times in the corners the extentions of the sides of P. By Lemma 7, $|L_n| \prec n^{p-1}$. Each of these lines intersects a singularity half-line of T^{-1} at most once, therefore the total number of intersections of the lines in L_n with the singularity half-lines of T^{-1} is bounded above by n^{p-1}. The edges of γ_n belong to the lines from L_n, therefore the total number of intersections of these edges with the singularity half-lines of T^{-1} is bounded above by n^{p-1}. (Note that the number of edges of γ_n could be bigger.) ∎

We will now estimate the number of faces of Γ_n. Denote by $|F_n|, |E_n|, |V_n|$ the number of faces, edges, vertices of the graph Γ_n respectively. By Lemma 8, growth of the second difference of the sequence $|E_n|$ is at most polynomial of

[16] We conjecture that there is a quadratic lower bound.

degree $p-1$, hence $|E_n| \prec n^{p+1}$. The edges of Γ_n are parallel to the sides of P, thus may have at most p possible directions. Therefore, each face of Γ_n is at most a $2p$-gon, and the valence of each vertex of Γ_n is at most $2p$. Thus, $|E_n| \leq p|F_n|$, $|E_n| \leq p|V_n|$. Euler's formula $|V_n| - |E_n| + |F_n| = 0$ implies $p|F_n| \leq (p-1)|E_n|$, hence $|F_n| \prec |E_n|$. ∎

We have assumed that the abelian group generated by the sides of P has maximal rank, $p-1$. Although generically this is the case, the rank may drop. Our argument proves, in fact, the statement below.

Corollary 2 *Let P be a convex euclidean p-gon, and let $r \leq p-1$ be the rank of the abelian group generated by translations in the sides of P. Then the complexity of the outer billiard about P is bounded by n^{r+2}.*

A polygon is *rational* if the rank above is 2. Rational polygons are dense in the space of all polygons. We will study complexity of the outer billiard about a rational polygon.

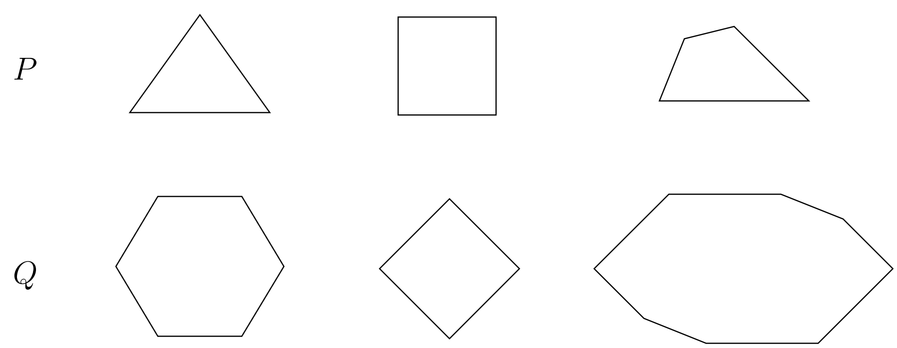

Figure 7: Outer billiard; examples of polygons P and Q

We regard the plane as a vector space, with the center in the interior of the convex p-gon P. A well known construction [16] associates with P a homothetic family of centrally symmetric convex polygons with at most (resp. exactly) $2p$ sides (resp. if P is a generic p-gon). Let Q be a particular polygon in this family. Each of its sides is parallel to a diagonal of P. See figure 7. We endow the plane with a Minkowski norm such that Q is the unit disc. The vector norm $|\cdot|$, radius, etc, will be understood with respect to it. We set $Q(r) = r \cdot Q$.

The polygon Q determines the geometry of orbits of T^2 "at infinity" [16]. We will elaborate. Let x be a point in the plane which is sufficiently far from the origin. Let Q_x be the circle centered at the origin and passing through x. Let $a \subset Q_x$ be the side containing x, and let d be the corresponding diagonal of P (parallel to a). Then T^2 translates x along a by $2|d|$; this continues until the orbit of x overshoots a. Let $y = T^{2m}$ be the corresponding point. Then the same recipe is applied to y, etc. See figure 8.

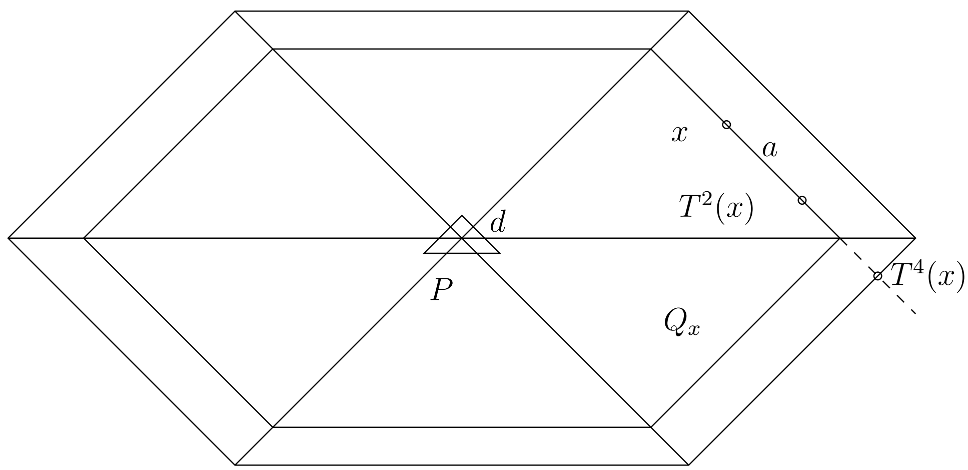

Figure 8: Second iteration of the outer billiard map "at infinity"

Let a be an arbitrary side of Q, and let d be the corresponding diagonal of P. The polygon P is *quasirational* if, up to a common factor, the p numbers $r_a = |a|/|d|$ are rational.

Theorem 7 *Let P be a rational polygon, and let $f(\cdot)$ be the complexity of the outer billiard about P. Then $f(n) \sim n^2$.*

Proof. Every rational polygon is quasirational. By a construction of R. Kolodziej [13],[17] there is a nested sequence of T-invariant, polygonal, simply connected domains $\cdots \subset U_i \subset U_{i+1} \subset \cdots$ exhausting the plane. By [13], there exists a constant $C = C(P) > 0$ such that the *Kolodziej domains* satisfy $Q(Ci) \subset U_i \subset Q(C(i+1))$.

[17] Using it, Kolodziej proved that the outer billiard orbits for quasirational polygons are bounded [13]. For general polygons this question is open [16].

Lemma 9 *Let P be an arbitrary convex polygon, and let $f(\cdot)$ be the complexity of the outer billiard about P. There exists $C_1 > 0$ such that the contribution to $f(n)$ of the exterior of the disc of radius $C_1 n$ grows linearly.*

Proof. We will use the preceding notation and terminology. Let $C_1 > 2/r_a$ for all sides of P.

Consider the T^2-orbit of length n of an arbitrary point x outside of $Q(C_1 n)$. It follows a side, a, of Q for $k \leq n$ iterations, then it "jumps" to the adjacent side, a', and follows it for $n-k$ iterations. Counting the possibilities (and assuming that Q is a $2p$-gon, which is the generic sitiation) we obtain $2p(n+1)$ types of T^2-orbits of length n. But different types mean different contributions to $f(2n)$, and vice versa. ∎

Since P is a rational polygon, the group $G \subset Iso(\mathbf{R^2})$ is discrete. The graphs Γ_n are obtained from a finite collection of half-lines by G-action, hence $\Gamma_\infty = \cup_{n\geq 1}\Gamma_n$ belongs to a discrete collection of lines. Therefore Γ_∞ is a graph, and the sequence $\Gamma_1 \subset \cdots \subset \Gamma_n \subset \dots$ stabilizes on compacta. Moreover, there is a finite collection of convex polygons, such that every face of Γ_∞ is congruent to a polygon in this collection. Hence the areas of the faces of Γ_∞ are bounded away from zero and infinity.

Note that the constant C_1 in Lemma 9 can be chosen arbitrarily large. We choose it so that $\frac{C_1}{C} = \tau \in \mathbf{N}$. Then for all n sufficiently large

$$Q(C_1 n) \subset U_{\tau n} \subset Q(C_1 n + C). \tag{24}$$

By Lemma 9, up to a linear term, $f(n)$ is the number of faces of Γ_n intersecting $Q(C_1 n)$. By the left inclusion in (24), this is less than or equal to the number of faces of Γ_∞ in $U_{\tau n}$. By preceding remarks, there is $C_2 > 0$ such that that number is bounded by $C_2 \text{area}(U_{\tau n})$. By the right inclusion in (24), $\text{area}(U_{\tau n})$ is quadratic in n. We have obtained the bound $f(n) \prec n^2$.

Now for the lower bound. All regular points in X are periodic [13]. A face $F \subset X$ of Γ_k is *stable* if F is a face of Γ_∞. Let $V_n \subset X$ be the set of points with period at most n. Each connected component of V_n is an open, stable face of Γ_n. By remarks above, the number of connected components of V_n has the same growth as the area of V_n, thus $\text{area}(V_n) \prec f(n)$. By Proposition 6 below, $\text{area}(V_n) \sim n^2$. ∎

The following proposition is used in the proof of Theorem 7. It is also of independent interest. If g, h are positive functions on $Y \subset \mathbf{R^2}$, the notation $g \prec h$ means that $g(x)/h(x)$ is bounded as $|x| \to \infty$. The notation $g \sim h$ means that $g \prec h, h \prec g$.

Proposition 6 *Let P be a convex polygon and let $X_{\text{per}} \subset X$ be the set of periodic points of the outer billiard. For $x \in X_{\text{per}}$ let $p(x)$ be the period.*
1. We have $|x| \prec p(x)$.
2. Let P be a rational polygon. Then for all regular points $p(x) \sim |x|$. Let V_n be the set of points such that $p(x) \leq n$. Then area$(V_n) \sim n^2$.

Proof. We assume, without loss of generality, that $p(x) = 2m$. Let Q_x be the circle through point x. The sequence $x, T^2(x), \ldots$ roughly follows Q_x. To come back to x, the sequence has to go around Q_x at least once. Let δ be the "largest step" of T^2. Then we need at least perimeter$(Q_x)/\delta$ steps to return. Since perimeter$(Q_x) \sim |x|$, the first claim follows.

Let now P be a rational (hence quasirational) polygon, and let U_k, $k \geq 1$, be the Kolodziej domains. Let $k = k(x)$ be such that $x \in U_k \setminus U_{k-1}$. The relations $Q(Ck) \subset U_k \subset Q(C(k+1))$ imply that the function $k(x)$ satisfies $k(x) \sim |x|$. By inclusion $U_k \setminus U_{k-1} \subset Q(C(k+1)) \setminus Q(C(k-1))$, we have area$(U_k \setminus U_{k-1}) \sim |x|$. The point x belongs to a unique face, $F = F(x)$, of Γ_∞, hence $p(x)$area$(F) \leq$ area$(U_k \setminus U_{k-1})$. By preceding remarks, $p(x) \prec$ area$(U_k \setminus U_{k-1})$, implying $p(x) \prec |x|$, and hence the equivalence $p(x) \sim |x|$.

By this relation, there are constants $C_3, C_4 > 0$ such that, for n sufficiently large, $Q(C_3 n) \subset V_n \subset Q(C_4 n)$, proving the last claim. ■

4.2 The elliptic and the hyperbolic cases

We will first consider the elliptic case.

Theorem 8 *Let $P \subset S^2$ be a convex spherical polygon. The complexity of the outer billiard about P grows subexponentially.*

Proof. For $x \in S^2$ let $l = x^*$ be the appropriately oriented great circle centered at x. This diffeomorphism $S^2 \to Geo(S^2)$ is the *spherical duality*, and we denote by $x = l^*$ the inverse diffeomorphism.

Let $a, b, \ldots$ be the corners of P, and let P^* be the convex polygon bounded by the geodesics $a^*, b^*, \ldots$. The correspondence $P \mapsto P^*$ is an automorphism of the space of convex spherical polygons. The proof of the following lemma is contained in [16].

Lemma 10 *Let $P, P^* \subset S^2$ be as above. Let X_o, X_b be the phase spaces of the outer billiard about P, inner billiard in P^*; let $T^{out} : X_o \to X_o$, $T_{bil} : X_b \to X_b$ be the respective maps.*

The spherical duality induces a diffeomorphism $X_o \to X_b$; it conjugates $T^{out} : X_o \to X_o$ and $T_{bil} : X_b \to X_b$; it induces an isomorphism of the codings.

Figure 9 illustrates Lemma 10. Let $f_o(n)$ (resp. $f_b(n)$) be the corner complexity of the outer billiard about P (resp. the side complexity of the billiard in P^*). By Lemma 10, $f_o(n) = f_b(n)$. The claim now follows from Theorem 4. ■

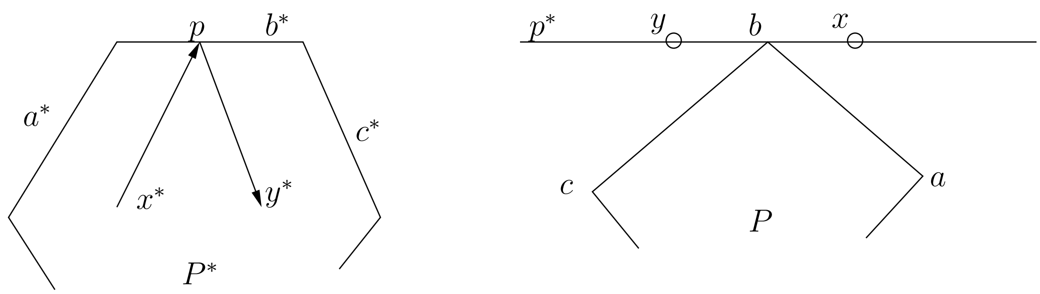

Figure 9: Duality between inner and outer billiards

Let $P \subset \mathbf{H}^2$ be a p-gon, and let $X = \mathbf{H}^2 \setminus P$. The outer billiard map[18] $T : X \to X$ extends to a homeomorphism of the circle at infinity, $\tau : S \to S$. Its rotation number satisfies $\rho(P) \geq 1/p$ [8]. The polygon P is *large* if $\rho(P) = 1/p$ and τ has a hyperbolic p-periodic orbit. See figure 10. The set of large polygons is open in the natural topology [8].

Theorem 9 *Let $P \subset \mathbf{H}^2$ be an arbitrary convex polygon, and let $f(\cdot)$ be the complexity of the outer billiard, T. Then $n \prec f(n)$. If P is a large polygon, then $f(n) \sim n$.*

[18] We refer to [8] for the background.

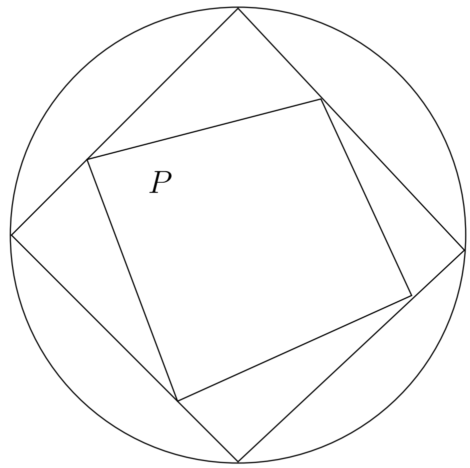

Figure 10: A large quadrilateral

Proof. The bound $n \prec f(n)$ fails iff the sequence Γ_k, $k \geq 1$, stabilizes. Assume this to be the case, and let $\Gamma_m = \Gamma_{m+1} = \cdots = \Gamma_\infty$. The outer billiard map preserves Γ_∞; its restriction to a closed face of Γ_∞ is a diffeomorphism onto another one. Since Γ_∞ is a finite graph, we find $n \in \mathbf{N}$ such that every face of Γ_∞ is invariant under T^n.

Let F be a closed face of Γ_∞. Then $\partial F \cap S$ is either empty, or a vertex, or an edge of F. We will study the latter. Let $v_1, \ldots, v_N \in S$ be the consecutive endpoints of these edges, let $e_i \subset S$ (resp. $\alpha_i \subset \mathbf{H}^2$) be the circular arc (resp. the geodesic) with endpoints v_i, v_{i+1} (we set $N+1=1$), and let F_i be the corresponding face of Γ_∞. The restriction $T^n|_{F_i}$ is induced by an isometry, $g_i \in Iso(\mathbf{H}^2)$. The elements $g_1, \ldots, g_N$ are all equal to the identity iff $\tau^N = 1$.

Lemma 11 *The map $\tau : S \to S$ is not periodic.*

Proof. Let z be a corner of P. For close points $x_1, y_1 \in S$ let $x_2, y_2 \in S$ be their reflections about z. Let $\lambda_1 = |x_2 z|/|x_1 z|$ and let $2\alpha_i$ be the angular measure of the arc $x_i y_i$, $i = 1, 2$. See figure 11. The triangles $x_1 z y_1$ and $x_2 z y_2$ are similar, therefore

$$\sin \alpha_2 = \lambda_1 \sin \alpha_1. \tag{25}$$

Let $x_1, \ldots, x_N$ be a periodic trajectory of the map τ consisting of smooth points, and let $\lambda_1 \ldots, \lambda_N$ be the respective ratios. Set $\Lambda = \Pi_{i=1}^N \lambda_i$. Let

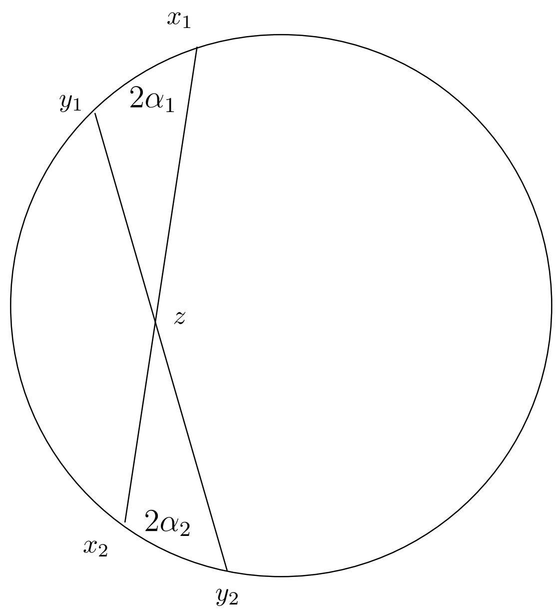

Figure 11: Computing the distortion of the map τ

y_1 be a point sufficiently close to x_1, and let $y_1, \ldots, y_N$ be its τ-orbit; we assume that both orbits reflect in the same corners of P. It follows from equation (25) that $y_1, \ldots, y_N$ is a periodic trajectory iff $\Lambda = 1$. In particular, if τ has a periodic interval, then $\Lambda = 1$ there.

Let now x_1 cross counter-clockwise a singularity half-line of T. In the notation of figure 12, $\lambda_1 = (b+c)/a$ (resp. $\lambda_1 = c/(a+b)$) right before (resp. after) this. By $(b+c)/a > c/(a+b)$, the equality $\Lambda = 1$ before a singularity half-line implies that $\Lambda < 1$ immediately after it (if, simultaneously, another x_i crosses a singularity half-line, Λ will decrease as well). ■

By Lemma 11, we can assume without loss of generality that $g_1 \neq 1$. Then g_1 is a (hyperbolic) parallel translation with the axis α_1, and F_1 is the domain bounded by α_1 and e_1. We will say that F_1 is a *lunar face* of Γ_∞. The union of lunar faces of Γ_∞ is invariant under T. Therefore for any $k > 0$ there is $l = l(k)$ such that $T^{-k}(\alpha_1) = \alpha_l$. A geodesic α_i, $1 \leq i \leq N$, cannot contain a side of P. If it does, then F_i contains a singular line of T in its interior, contrary to the definition of F_i. See figure 12, where $x_1 x_2$

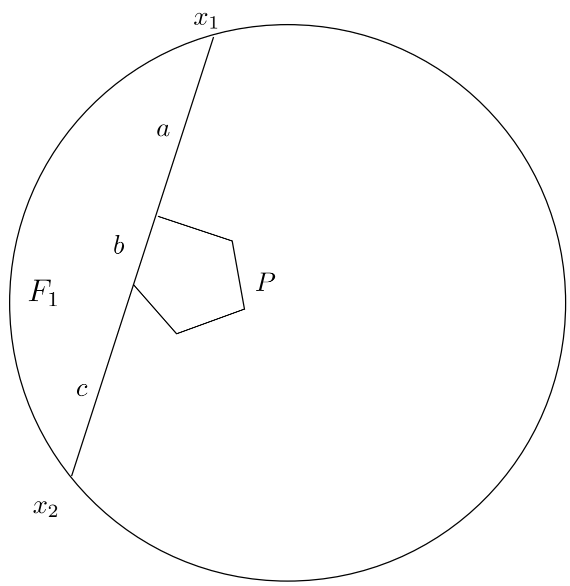

Figure 12: Destruction of a periodic orbit of τ

represents now the geodesic α_1. Thus, α_1 is not an edge of Γ_m for any m. This contradiction proves our first claim.

Let now P be a large p-gon. Then Γ_n is a disjoint union of p binary trees [8] (see figure 13), hence $|\Gamma_n|$ grows linearly. ∎

Remark 3 The function $f(\cdot)$ is bounded below by the complexity of the induced map $\tau : S \to S$ with respect to the natural partition. However, the latter may be finite. See figure 14.

References

[1] L. Abramov, *On the entropy of a flow*, Dokl. Akad. Nauk SSSR **128** (1959), 873 – 875.

[2] M. Berger, *Geometry*, Springer-Verlag, Berlin-Heidelberg-New York, 1987.

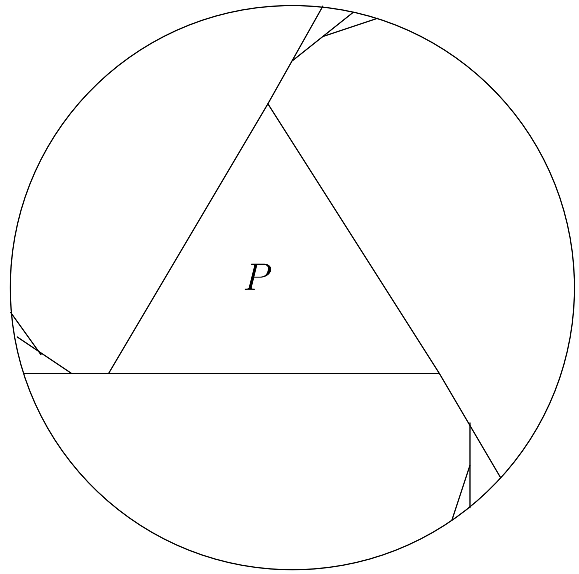

Figure 13: The graph Γ_2 for a large triangle

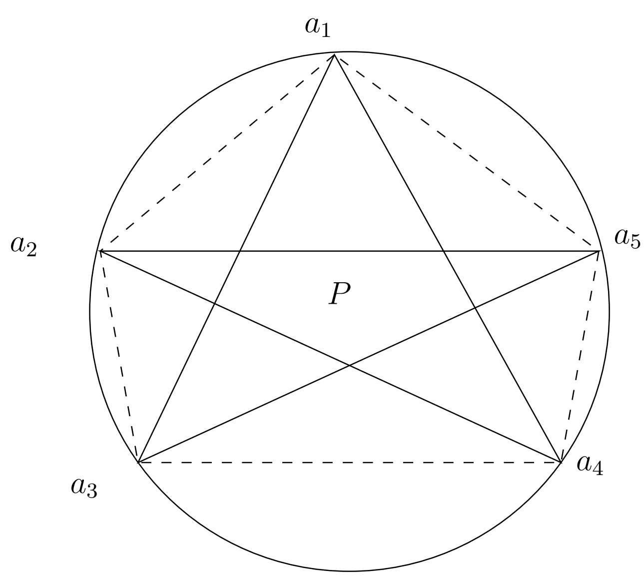

Figure 14: Finite complexity of the outer billiard map at infinity

[3] H. Busemann, B. Phadke, *Spaces with distinguished geodesics*, Marcel Dekker, New York, 1987.

[4] J. Cassaigne, *Complexité et facteurs speciaux*, Bull. Belg. Math. Soc. **4** (1997), 67 – 88.

[5] J. Cassaigne, P. Hubert, S. Troubetzkoy, *Complexity and growth for polygonal billiards*, Ann. Inst. Fourier **52** (2002), 835 – 847.

[6] H. S. M. Coxeter, *Regular Polytopes*, Pitman Publ. Co., New York, 1948.

[7] P. de la Harpe, *Topics in geometric group theory*, Univ. Chicago Press, Chicago, 2000.

[8] F. Dogru and S. Tabachnikov, *Dual billiards*, Math. Intelligencer **27** (2005), no. 4, 18 – 25.

[9] B. Gutkin, E. Gutkin, U. Smilansky, *Hyperbolic billiards on surfaces of constant curvature*, Comm. Math. Phys. **208** (1999), 65 – 90.

[10] E. Gutkin, *Billiard dynamics: a survey with the emphasis on open problems*, Reg. & Chaotic Dyn. **8** (2003), 1 – 13.

[11] E. Gutkin, N. Haydn, *Topological entropy of polygon exchange transformations and polygonal billiards*, Erg. Theory & Dynam. Syst. **17** (1997), 849 – 867.

[12] E. Gutkin and S. Tabachnikov, *Complexity of piecewise convex transformations in two dimensions, with applications to polygonal billiards*, math.DS/0412335, MathArxive (2004).

[13] R. Kolodziej, *The antibilliard outside a polygon*, Bull. Polish Acad. Sci. Math. **37** (1989), 163 – 168.

[14] H. Masur, S. Tabachnikov, *Rational billiards and flat structures*, in Handbook of Dynamical Systems, Elsevier, Amsterdam (2002), 1015 – 1090.

[15] L. Santalo, *Integral geometry and geometric probability*, Cambridge University Press, Cambridge, 2004.

[16] S. Tabachnikov, *Billiards*, Soc. Math. de France, Paris, 1995.