PETs and Polygonal Outer Billiards: Flux, Renormalization, and Compacification

Richard Evan Schwartz *

July 4, 2024

Abstract

This article, which loosely follows lectures I gave at Luminy, discusses some ideas about polytope exchange transformations and polygonal outer billiards.

1 Introduction

This article is an elaboration on two of the four lectures I gave at C.I.R.M., Luminy, in August 2023. Two of the lectures discussed polygonal outer billiards and piecewise isometric maps. Amongst the piecewise isometric maps I mostly talked about polytope exchange transformations, or PETs. The third lecture discussed ordinary billiards, and the fourth lecture – kind of a wild card topic I threw in at the spur of the moment – discussed a connection between continued fractions and a special case of the Four Color Theorem.

This article will concentrate on the two lectures about polygonal outer billiards and PETs. Aside from considerations of length, let me give some rationale for this choice. First, I recently wrote a survey article that concentrated on ordinary billiards for the Proceedings of the 2022 ICM. See [**S1**]. Second, the mathematics on the Four Color Theorem is well exposited in my paper [**S2**]. Third, I am inspired by the popularity of one of the problem

*Supported by N.S.F. Grant DMS-2102802, a Mercator Fellowship, and a Simons Sabbatical Fellowship

sessions I gave at Luminy. This problem session concentrated on a beautiful (and well studied) two dimensional example of a piecewise isometric map. See §2.3. I thought it would be nice to talk more about that kind of mathematics.

Lastly, my involvement with outer billiards is much more extensive than my involvement with the other topics I lectured on. I spent about 7 years of my life really working hard on this topic. I thought it would be nice to give an impressionistic overview of all this work, culminating in a grand but unfiulfilled vision of how I see polygonal outer billiards.

I have a wistful feeling as I write this article. While working on outer billiards and PETs, I had the impression that I could spend the rest of my life making fruitful discoveries about it. It was very much like drinking water by putting a fire hose up to your face. I guess that many people working in dynamics, especially experimentally inspired dynamics, would say similar things. However, eventually I abandoned the topic because I wanted to explore other areas of mathematics, to see what else I could do.

This article is structured as follows. In §2 I define what I mean by a piecewise defined map, and then give a bunch of examples. PETs represent the special case when everything is defined in terms of polytopes and the maps are translations. The examples start out easy and familiar and then gradually evolve into more intricate ones. One theme in §2 is the role played by renormalization in understanding the deep structure of the dynamics of these kinds of maps. These maps do not always have a renormalization scheme associated to them, but then they do it is a powerful tool.

In §3 I will discuss polygonal outer billiards, often through examples. One general idea I want to get across is that there is an associated unbounded PET which I call *the pinwheel map*. This map is often easier to understand than outer billiards but, with respect to unbounded orbits, it contains precisely the same information. I will use the structure of the pinwheel map to explain the celebrated result of Vivaldi-Shaidenko [**VS**], Kolodziej [**K**], and Gutkin-Simanyi [**GS**] that all orbits are bounded for so-called quasi-rational polygons. (I stick to the case where there are no parallel sides.) As a corollary, one gets the familiar and very appealing result that all orbits are periodic with respect to a polygon with rational vertices.

In §4 I will concentrate on my result about unbounded orbits for irrational kites. This is the subject of both my monographs [**S3**] and [**S4**]. I hope to give at least a feel for how the results work. The way I will illustrate the result is through an associated geometric construction called the

arithmetic graph. This idea is closely related to the *Galois flux* for interval exchange transformations, and the *module construction* in [**VL**]. The final section in §4 discusses some ideas about a general theory of polygonal outer billiards. Many sections in this article illustrate these ideas separately, in simple situations, and then I will try to put them all together.

One thing I would like to say is that almost everything discussed in this article comes from experimentation. I invite the reader to download the software [1] and try it out. Using these programs and seeing the pictures is really the right way to absorb this kind of mathematics. I learned everything from writing and using the programs. If you are having trouble installing and using the software, feel free to send me email and I will personally help get it to work for you.

I would like to thank C.I.R.M. for their hospitality during my stay there in the summer of 2023. I had a great time. I would especially like to thank Nicolas Bedaride and Jayadev Athreya for inviting me to give these lectures. Some of the ideas about compactification of polygonal outer billiards systems benefitted from discussions with John Smillie. He and I had planned to work on some of this together but somehow never did. I'd like to thank John for those conversations. My work is currently supported by the U.S. National Science Foundation, by a Mercator Fellowship (from Germany) and by a Simons Sabbatical Fellowship. I thank all these institutions for their support.

[1] In a few simple cases, as for 3-IETs, my code is not publicly available because it is not a very polished program; I can send it out upon request.

2 Piecewise Isometric Maps

2.1 General Definitions

In this section we give very general definitions, and then we give concrete examples in subsequent sections. Some special cases of maps we define have been studied in quite a bit of detail and for a long time. See e.g. [**Go**] for a survey in the piecewise isometric case. Since the definition is so general, many other cases have not been explored at all.

Piecewise Defined Maps: Suppose that $X \subset \boldsymbol{R}^n$ is a subset with non-empty interior. Suppose that $\mathcal{A}$ is some collection of maps from $\boldsymbol{R}^n$ to $\boldsymbol{R}^n$. A *piecewise-$\mathcal{A}$* map of X is given by the following data:

- A union $X = X_1 \cup ... \cup X_m$ such that each X_i has non-empty interior and these interiors are pairwise disjoint. We call this decomposition an *almost partition* of X.
- A choice $f_i \in \mathcal{A}$ for all $i = 1, ..., m$ such that $f_i : X_i \to X$ is well-defined.

We then define $f : X \to X$ by the rule $f(p) = f_i(p)$ when p lies in the interior of X_i. The map is undefined on points of X which do not lie in the interior of some X_i. We have set things up in such a way that f is almost everywhere defined in X. That is, f is defined except on a set of Lebesgue measure 0.

Invertibility: We call f *invertible* if the maps f_i are all injective, and $f_i^{-1} \in \mathcal{A}$, and the sets $f_1(X_1), ..., f_m(X_m)$ form an almost partition of X. In this case, the inverse map f^{-1} is also a piecewise-$\mathcal{A}$ map. All the examples we consider will be invertible. Usually the maps we consider are invertible in this way. In the invertible case, f is defined in terms of the *first partition* $X_1, ..., X_m$ and the *second partition* $f(X_1), ..., f(X_m)$, and the maps between them.

Periodic Points and Islands: We denote the k-fold composition of f by f^k. A point $p \in X$ is *periodic* if $f^k(p) = p$ for some k. A subset $Y \subset X$ is called a *periodic island* if there is some k such that f^k is the identity on Y. Given a periodic point p, we define the *symbolic sequence* of p to be the finite sequence $i_0, ..., i_{k-1}$ such that $f^j(p) \in X_{i_j}$. So, in other words, $p \in X_{i_0}$ and $f(p) \in X_{i_1}$, etc.

First Return Map: Given $Y \subset X$ and a point $p \in Y$ we can attempt to define the *first return* $f|_Y(p) \in Y$ to be the first positive iterate $f^k(p) \in Y$. Of course, it might happen that this is not defined at all. We say that $f|_Y$ is *well-defined* if there is a piecewise-$\mathcal{A}$ map $f^* : Y \to Y$ which agrees with $f|_Y$ whenever both are defined. In particular, $f|_Y$ is defined on some almost partition of Y.

Renormalization Sets: We call Y a *renormalization set* for f if $f|_Y$ is well defined and if there is some k such that $X - Y$ has an almost partition into periodic islands and sets of the form $f^j(Y)$ for $j \leq k$. So, ignoring a set of measure 0, all points X either belong to periodic islands or else land in Y after a certain number of iterates. If we ignore the finite number of periodic islands in $X - Y$, then the dynamics of f on X essentially reduces to the dynamics of $f|_Y$ on Y. The kind of situation sometimes gives a lot of information about f, particularly if we can relate $f|_Y$ back to f in some way.

2.2 Interval Exchange Transformations

The simplest maps of the kind we are considering are *interval exchange transformations*. These are often called *IET*s for short. In this case $n = 1$ and X is a line segment and $\mathcal{A}$ is the set of translations of the line. There is an enormous literature on these, and we will only mention a few things. For information about IETs, especially in connection with rational billiards, translation surfaces, and Teichmuller dynamics, see [**FM**] and [**Z**] and the references therein.

General Case: Let S_m be the group of permutations on $\{1, ..., m\}$. In general, an interval exchange transformation is defined by three pieces of data: an integer m, a permutation $\pi \in S_m$, and positive numbers $\lambda_1, ..., \lambda_m$. Setting $\lambda = \lambda_1 + ... + \lambda_m$, our initial partition of $[0, \lambda]$ is given by intervals of lengths $\lambda_1, ..., \lambda_m$ ordered from left to right. We then reorder these intervals according to the partition π and obtain a second partition of $[0, \lambda]$. The map f in this instance carries the kth interval of the first partition to the $\pi(k)$th interval of the second partition by a translation. The space of all m-interval IETs is naturally identified with $S_m \times \mathcal{C}_m$, where $\mathcal{C}_m$ is the strictly positive orthant in $\boldsymbol{R}^m$.

A permutation π is called *irreducible* if π does not preserve any subset

$\{1, ..., k\}$ of $\{1, ..., m\}$ with $k < m$. Otherwise we call π *reducible*. Usually we ignore the components of $S_m \times \mathcal{C}_m$ corresponding to reducible permutations. The corresponding IETs can be understood by breaking them into simpler pieces.

Rotations: The simplest example of an interval exchange transformation arises when the unit interval $[0, 1]$ is divided into two intervals $[0, \lambda]$ and $[\lambda, 1]$. The interval exchange transformation switches the two intervals. In other words, $f : [0, 1] \to [0, 1]$ is defined so that $f(x) = x + (1 - \lambda)$ if $x \in (0, \lambda)$ and $f(x) = x - \lambda$ if $x \in (\lambda, 1)$. We do not define f on the endpoints of the intervals. The map f agrees with the rotation $f^*(x) = [x - \lambda]$ in $\boldsymbol{R}/\boldsymbol{Z}$ wherever f is defined. When λ is irrational, every defined orbit of f is dense. When λ is rational, every defined orbit is periodic.

Three Interval IETs: Three interval IETs are still quite special and well understood. Let me describe a nice way to visualise them. A similar idea, in a slightly different context, is considered by [**VL**].

Let $\omega = \exp(2\pi i/3)$ be the usual third root of unity. Let $\nu_k = \omega^{k-1}$. Thus ν_1, ν_2, ν_3 are the three third roots of unity. Given a point $p \in [0, \lambda]$ with a well defined orbit we first assign the integer *symbolic sequence* $i_0, i_1, i_2, ...$ such that $f^k(p)$ lie in the i_kth interval. We define

$$z_k = \nu_{i_0} + ... + \nu_{i_k}, \qquad k = 0, 1, 2, ... \tag{1}$$

Thus we define a path $z_0 \to z_2 \to z_2, ...$ in the plane which chooses its direction to turn based on the itinerary of the orbit. Call this path $\Pi(p)$.

The path $\Pi(p)$ is not quite embedded. However, here is an interesting phenomenon. We define the *drift vector*

$$d = \sum_{i=1}^{3} \lambda_i \nu_i. \tag{2}$$

We then define the modified path $\Pi_\epsilon(p)$ so that its vertices are $z_0', z_1', z_2', ...$ with

$$z_k' = z_k + \epsilon_k. \tag{3}$$

We are giving the path $\Pi(p)$ a little push in the direction d and calling the new path $\Pi_\epsilon(p)$. This is something I noticed computationally, and just for fun. I would not be surprised if the result is known, though I don't know a

reference. Let me show some pictures, when $\epsilon = .3$. In each case, the black part of the path corresponds to one period, and then the lighter grey part is the periodic continuation. The chosen point p is rather random; other points would yield similar pictures: The (inplied) infinite grey path would stay the same and the black portion would shift.

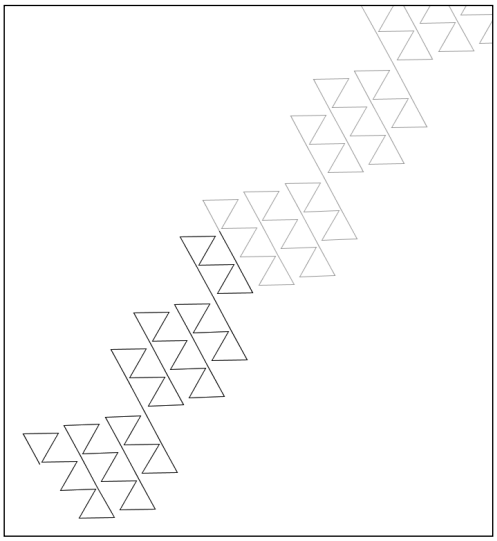

Figure 2.1: $\pi = (13)$ and $\overrightarrow{\lambda} = (24/64, 23/64, 17/64)$.

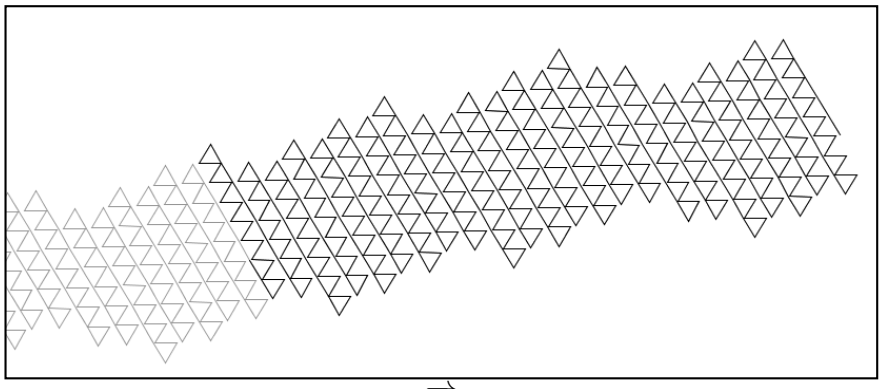

Figure 2.2: $\pi = (13)$ and $\overrightarrow{\lambda} = (157/512, 175/512, 180/512)$.

These pictures are a somewhat biased sample. I tried to pic ones which look especially interesting. The reader might enjoy writing a program that codes up this construction, as I have done. My own software for this – a fairly crude program – is available upon request.

Generic Minimality: The literature on IETs is enormous, and there are many, many structural results. Let me mention two that relate well to the rest of the material in this article. The first is *generic minimality*. If π is an irreducible permutation then almost every point in $\{\pi\} \times \mathcal{C}_n$ gives rise to an IET in which every defined orbit is dense. This is a corollary of a somewhat more precise result of Keane [**Ke**]. When π is reducible, one could break the corresponding IETs into suitable pieces and make a similar statement about each irreducible piece. So, at least generically, there are no periodic islands.

Rauzy Induction: The second property is a general renormalization scheme called *Rauzy renormalization* or *Rauzy induction*. See [**R**] for the original source; there are also many other articles which use Rauzy induction. Let f be an n-interval IET. There are three cases to consider:

- If the last interval of the first partition is shorter than the last interval of the second partition. We define $\rho(f)$ to be the first return map of f to the first $n-1$ intervals of the first partition. This union is an renormalization set.
- If the last interval of the second partition is shorter than the last interval of the first partition. We define $\rho(f)$ to be the first return map of f to the union of the first $n-1$ intervals of the second partition. This union is an renormalization set.
- Otherwise we do not define $\rho(f)$.

Rauzy proves that $\rho(f)$ is another n-interval IET. On each component of $S^n \times \mathcal{C}_n$, the map ρ is either entirely undefined or else almost everywhere defined and piecewise affine.

2.3 The Isosceles Triangle Example

Going to higher dimensions, we can take $\mathcal{A}$ to be the set of translations of $\boldsymbol{R}^n$ and insist that X and its partitions are all convex polytopes. (A *convex polytope* is the convex hull of a finite set of points in $\boldsymbol{R}^n$.) When these maps are invertible, we call them *polytope exchange transformations*, or *PET*s for short.

We can generalize this further by allowing $\mathcal{A}$ to be the set of all isometries of $\boldsymbol{R}^n$, or perhaps the set of all affine transformations of $\boldsymbol{R}^n$. The nice thing

about affine transformations is that they form a finite dimensional Lie group, and they preserve the collection of convex polytopes. So, if we deal with piecewise affine maps, ultimately we are describing how various polytopes move around in space. This is as general as we are going to get.

In this section we describe a classic example in $\boldsymbol{R}^2$. See [**Go**] for another discussion of this example. Comparing the discussion here to the general case discussed in §2.1, we switch notation somewhat so that $X = T_0$ and $(X_1, X_2) = (T_0, U)$. Let T_0 be an isosceles triangle with side lengths $1, \phi, \phi$, where $\phi = (1 + \sqrt{5})/2$ is the golden ratio. The left side of Figure 2.3 shows how one can use a regular pentagon as scaffolding to draw T_0 and also a partition of T_0 into two smaller isosceles triangles T_0 and U_1.

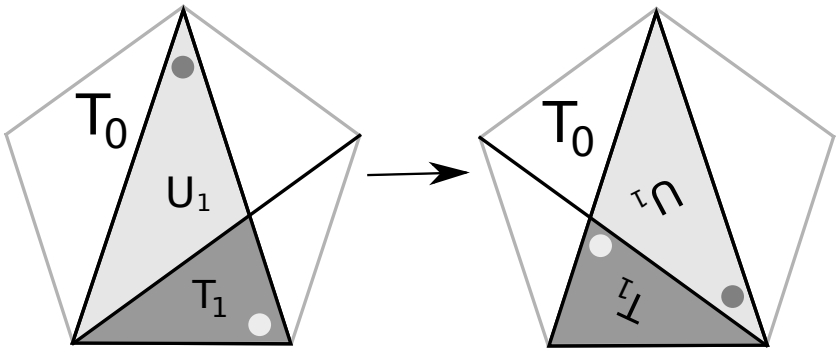

Figure 2.3: The partition of T_0 into T_1 and U_1.

The right side of Figure 2.3 shows a second partition of T_0 into two smaller isosceles triangles. The rotated labeling is deliberate. We map T_1 and U_1 to their counterparts by rotations. The rotated labels and dots indicate this. This piecewise rotation is defined everywhere on T_0 except for the interface $T_1 \cap U_1$. The resulting map f is an example of an invertible piecewise isometry.

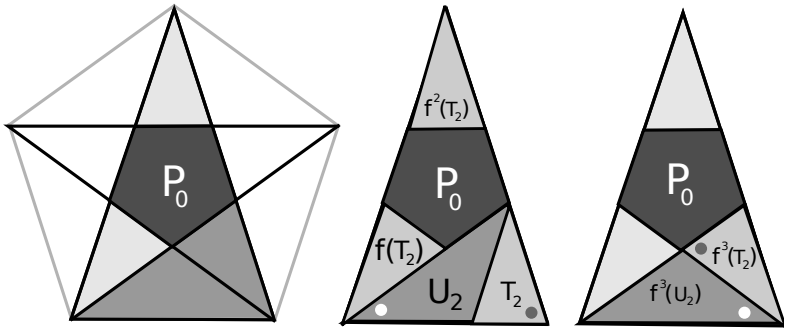

Figure 2.4: The periodic island P_0 and the first return map $f|_{T_1}$.

The left side of Figure 2.4 shows a regular pentagon P_0 which is a periodic island. The iterate f^5 acts as the identity on P_0.

Let's look at the middle and left part of Figure 2.4. The triangle T_1 breaks into two smaller pieces T_2 and U_2. The first return map $f|_{T_1}$ is well defined, and there is some additional structure. There is a similarity $\phi : T_1 \to T_0$ such that

$$f^3|_{T_1} = \phi^{-1} \circ f \circ \phi. \tag{4}$$

Furthermore, we have the almost partition.

$$U_1 - P_0 = f(T_2) \cup f^2(T_2). \tag{5}$$

All this says that T_1 is a renormalization set for f. Moreover, the action of $f|_{T_1}$ on T_1 is essentially a smaller copy of the action of f on T_0.

This structure lets us recursively analyze the periodic islands. First, if we see a certain pattern of periodic islands in T_0 then we can shrink and copy that pattern and put it inside T_1. Second, we can use f and f^2 to transport any pattern of periodic islands in T_2 to a pattern in $f(T_2)$ and $f^2(T_2)$. Figure 2.5 shows this in action.

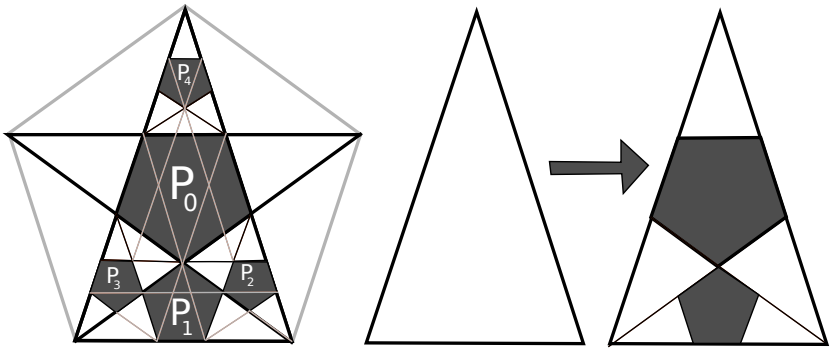

Figure 2.5: Recursively constructing periodic islands

The left side of Figure 2.5 shows some more periodic islands whose existence is implied by the two rules above. P_1 is a smaller copy of P_0, and then P_2 is a smaller copy of P_1. Finally, P_3 and P_4 are images of P_2 under f and f^2 respectively. We jave added some additional line segments to illustrate how these islands may be constructed just using a ruler. We could continue this analysis forever, producing a dense set of pentagonal periodic islands. The middle and right pictures in Figure 2.5 show a subdivision rule which breaks an isosceles triangle into two pentagons and 5 smaller isosceles triangles of the same shape. If we iterate this subdivision rule, we produce all the periodic islands for the map f.

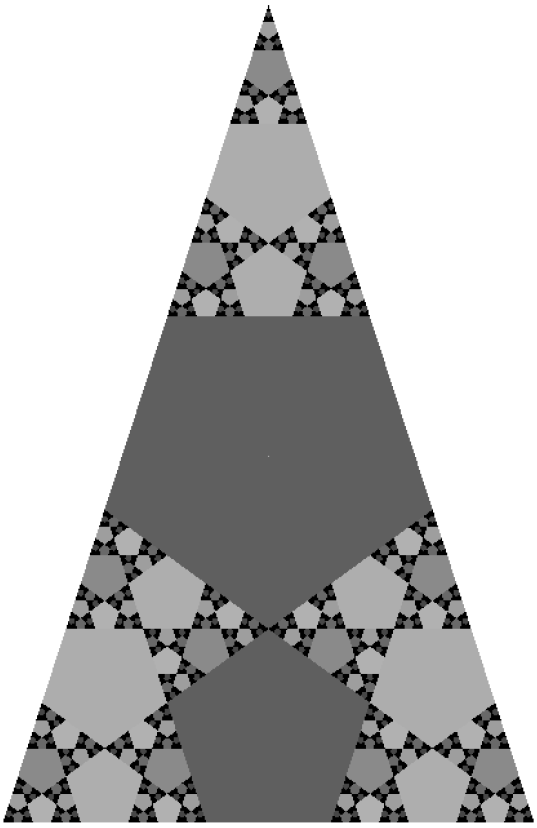

Figure 2.6: The biggest periodic islands

Figure 2.6 plots all the periodic islands having period less than 1275. The islands are shaded randomly, based on a random number that is associated to each possible period.

Visualization: There is a neat way to visualize the dynamics of this map. I don't know if this method has been studied by other authors, but it is closely related to what people call the *Galois flux* for IETs. We associate to each orbit the symbolic sequence using the numbers $\{1, 4\}$. We assign a 1 when the orbit lies in T_1 and a 4 when the orbit lies in U_1. (Doing the reverse would lead to the same kind of picture.) Call this sequence $m_0, m_1, m_2, ...$ We set $n_k = (m_0 + ... + m_k) \bmod 5$. We then let $\zeta_k = \omega^{n_k}$ where $\omega = \exp(2\pi i/5)$. The number ζ_k is the Galois conjugate of the unit complex number which computes the local rotation of the map f^k applied to the initial point. Fi-

nally, we associate to the orbit the polygon whose successive vertices are ζ_0, $\zeta_0 + \zeta_1$, $\zeta_0 + \zeta_1 + \zeta_2$, etc. Here is what a long periodic orbit looks like:

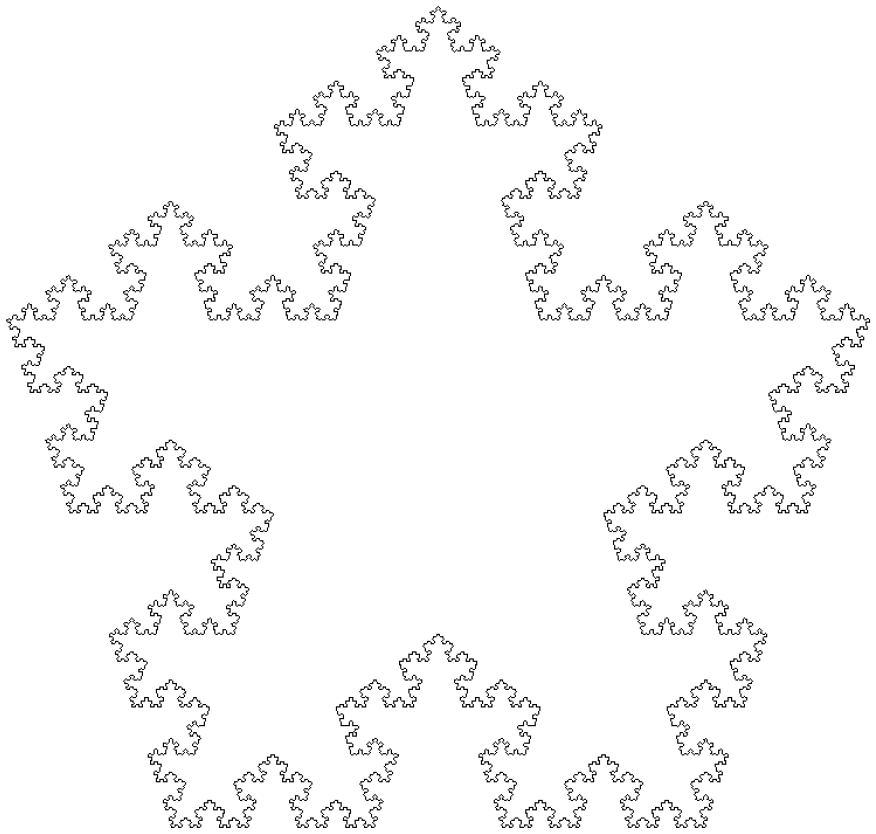

Figure 2.7: The Galois flux of a long periodic orbit.

2.4 Subattice PETs

Here I describe a general construction I made in [**S6**] and [**S7**] for lattices. A *sublattice* in $\boldsymbol{R}^n$ is the $\boldsymbol{Z}$-span of a list of $k \leq n$ linearly independent vectors. Here k is the *rank* of the sublattice. When $k = n$, a sublattice is known as a *lattice*. A *fundamental domain* for a sublattice L is a subset $X \subset \boldsymbol{R}^n$ such that union of translates of X by L gives an almost partition of $\boldsymbol{R}^n$. For instance, the unit cube $[0,1]^n$ is a fundamental domain for $\boldsymbol{Z}^n$. One typical example we consider is that of a parallelotope fundamental domain for a lattice. Another example we consider is an infinite strip fundamental

domain for a rank one sublattice in $\boldsymbol{R}^2$. Every sublattice has infinitely many fundamental domains.

Given a sublattice L and a fundamental domain X there is a map

$$f_{X,L} : \boldsymbol{R}^n \to X$$

Given $p \in \boldsymbol{R}^n$, the point $f_{X,L}(p)$ lies in L and differs from p by a member of L. For almost all $p \in \boldsymbol{R}^n$ this map is uniquely defined. For all other points we leave the map undefined.

Suppose now that Γ is a finite bipartite graph, with black and white vertices. We say that a *decoration* of G is an assignment of a sublattice to each white vertex and a subset of $\boldsymbol{R}^n$ to each black vertex such with the following property: For each edge, the set associated to the black vertex is a fundamental domain for the sublattice associated to the white vertex.

We now consider paths in Γ that start at black vertices. Suppose that γ is a length 2 path in Γ, whose consecutive vertices are b_0, w_1, b_2. Let X_0, L_1, X_2 be the corresponding decorations. We have a canonical piecewise isometry $X_0 \to X_2$ defined as the composition

$$\phi_\gamma = \phi_{b_2,w_1,b_0} = f_{X_2,L_1} \circ \iota, \tag{6}$$

where ι is the inclusion map. More generally, if γ is a length $2n$ path with vertices $b_0, w_1, ..., b_{2n}$ then we have the composition

$$\phi_\gamma = \phi_{b_{2n},w_{2n-1},b_{2n-1}} \circ ... \circ \phi_{b_2,w_1,b_0}. \tag{7}$$

When γ is a loop, meaning that $b_{2n} = b_0$, the map ϕ_γ is a PET defined on X_0. The inverse of the map is given by $\phi_{\gamma^{-1}}$, where γ^{-1} is the same loop as γ but traced in the reverse order.

One case we specially consider is as follows:

- Γ is a 4-cycle.
- γ is a loop that traverses all 4 vertices.
- The white vertex decorations are lattices.

In this case, we call ϕ_γ a *double lattice PET*. We have two lattices and two fundamentals for them. We consider such examples in the next section.

2.5 The Octagonal PETs

Now I will specialize the discussion above to a 1-parameter family of 2-dimensional double lattice PETs. These are the main objects in [**S7**].

Given $s \in \boldsymbol{R}$, the two fundamental domains are

$$X_0 = \langle (2s, 2s), (2, 0) \rangle, \qquad X_2 = \langle (-2s, 2s), (0, 2) \rangle. \tag{8}$$

Here $\langle v_1, v_2 \rangle$ denotes the parallelogram with vertices $\pm v_1/2 \pm v_2/2$. These paralleograms are 90-degree rotations of each other. The two lattices are

$$L_1 = \{(2s, 2s), (0, 2)\} \qquad L_2 = \{(-2s, 2s), (-2, 0)\} \tag{9}$$

Here $\{v_1, v_2\}$ denotes the $\boldsymbol{Z}$-span of v_1 and v_2. These two lattices are also 90-degree rotations of each other. Each parallelogram is a fundamental domain for each lattice, and so the map f_s, defined by the loop b_0, w_1, b_2, w_3 is a double lattice PET.

It turns out that f_s has a dense set of periodic islands for any $s \in \boldsymbol{R}$. These periodic islands are either right-angled isosceles triangles or semi-regular octagons. (A square counts as a semi-regular octagon.) When s is irrational, there are no triangles. For a generic choice of s, once sees a dense set of shapes of semi-regular octagons amongst the set of periodic islands. All of this is proved in [**S7**].

Here is the picture for $s = \sqrt{2}/2$. All the periodic islands are regular octagons. This picture is reminiscent of Figure 2.6, and there is a similar renormalization scheme, discussed below, that is responsible for this structure of periodic islands.

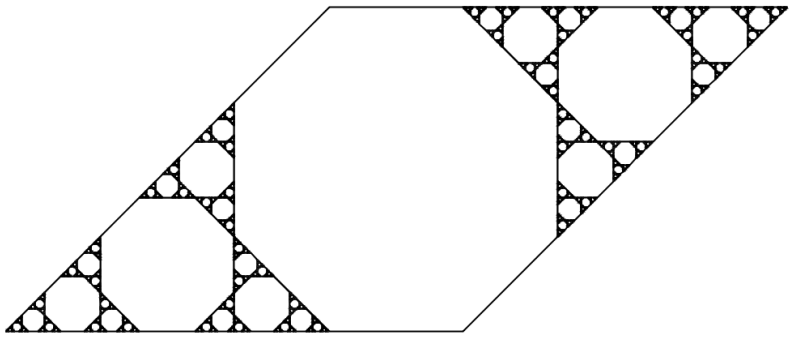

Figure 2.8: The periodic islands for $s = \sqrt{2}/2$.

Here is a picture for $s = \sqrt{3}/2 - 1/2$. This time all the periodic islands are squares. I have shaded them in to indicate the two different types. The interface between the squares is a union of two continuous embedded arcs.

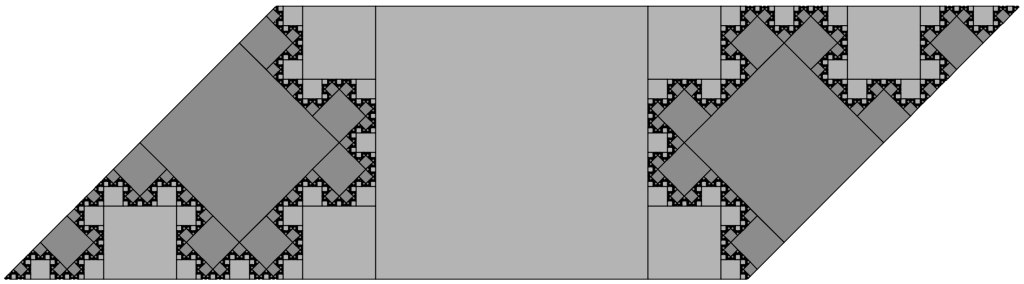

Figure 2.9: The periodic islands for $s = \sqrt{3}/2 - 1/2$.

If we think of this picture as living on a torus rather than on a parallelogram, then the two arcs piece together to make a continuous embedded loop. The restriction of f_s to this loop is conjugate to an irrational rotation (wherever it is defined). A renormalization scheme is also responsible for all this structure.

Now we explain how the renormalization scheme works in general. We define $R : (0,1) \to [0,1)$ as follows:

- $R(s) = 1 - s$ if $s > 1/2$.
- $R(s) = 1/(2s) - \text{floor}(1/(2s)$ if $s < 1/2$.

Recall that $f_s : X_{0,s} \to X_{0,s}$ is our PET. We let $Y_{0,s} = X_{0,s} - X_{2,s}$. In Figures 2.8 and 2.9, the set $Y_{0,s}$ is the complement of the central tile in the picture. This set has two components in all cases. The intersection $X_{0,s} \cap X_{2,s}$ is always a semi-regular octagon and a periodic island. So, we omit this periodic island to get $Y_{0,s}$.

Let $t = R(s)$. The main result of [**S7**] is that there is a renormalization set $Z_s \subset Y_s$ together with a piecewise similarity $\phi_s : Z_s \to Y_t$ which conjugates the map $f_s|_{Z_s}$ to the map $f_t|_{Y_t}$. The map ϕ_s is a similarity on each of the two symmetrically placed components of Z_s. We illustrate this result with the same pictures that we use in [**S7**]. We are not showing the dynamics, but rather the collection of periodic islands. Comparing the pictures in Figures 2.10a and 2.10b, and then again in Figures 2.11a and 2.11b, one can get a sense of what the main theorem says. The details are worked out in great detail in [**S7**].

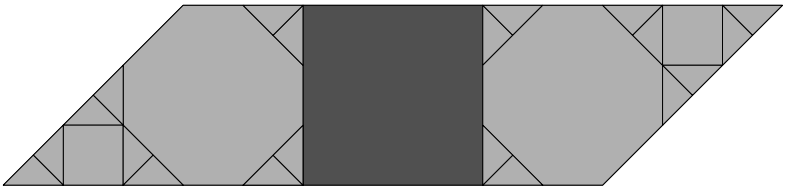

Figure 2.10a: Y_t lightly shaded for $t = 3/10 = R(5/13)$.

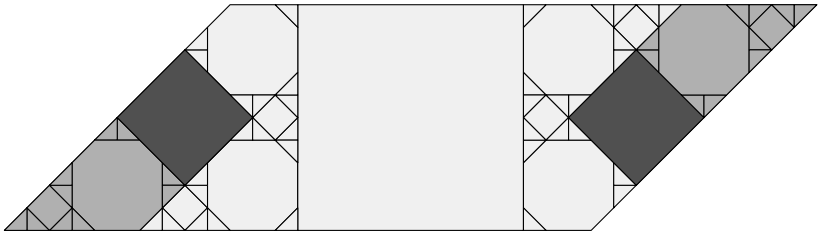

Figure 2.10b: Z_s lightly shaded $s = 5/13$.

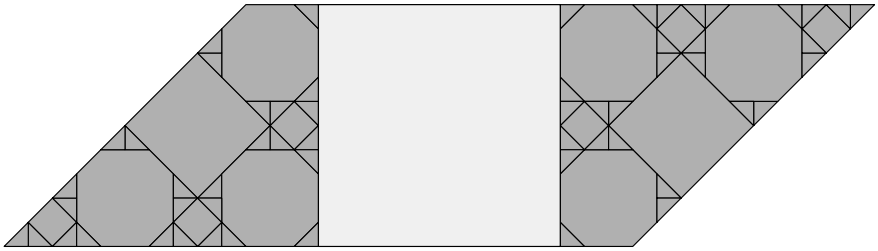

Figure 2.11a: Y_t lightly shaded for $t = R(8/13) = 5/13$.

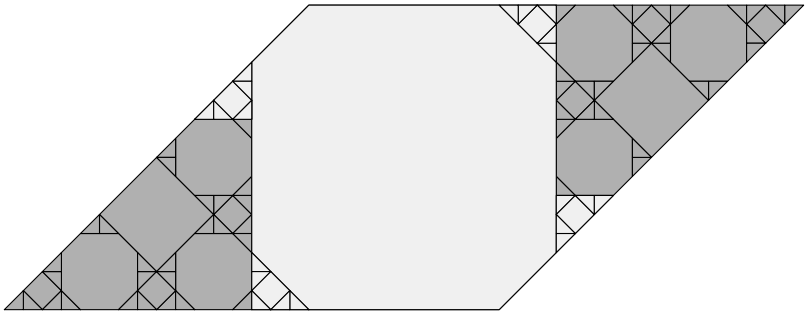

Figure 2.11b: Z_s lightly shaded for $s = 8/13$.

The self-similar nature of the periodic islands for $s_0 = \sqrt{2}/2$ comes from the fact that $R^2(s_0) = s_0$. The self-similar nature of the periodic islands for $s_1 = \sqrt{3}/2 - 1/2$ comes from the fact that $R(s_1) = s_1$. In general, any quadratic irrational parameter is pre-periodic under R and exhibits the same kind of self-similar behavior, albeit more complicated. The typical orbit of s under R is dense in $[0,1]$. This accounts for our statement that for generic choice of s, one seems a dense set of shapes of semi-regular octagon amongst the periodic islands.

The most significant property of R is that the orbit of any rational number under R is finite and ends at $1/2$. The map R has a simplifying effect on rationals. For instance, $3/10 = R(5/13)$ is simpler than $5/13$. The whole orbit of $5/13$ is

$$5/13 \to 3/10 \to 2/3 \to 1/3 \to 1/2.$$

The simplification property allows one to inductively understand the structure of f_s when s is rational. Then, taking limits, one can understand many things about f_s for any value of s.

2.6 PETs Constructed from Strips

Suppose now we have a bipartite graph which is a $2n$-cycle. Suppose that to the black vertices we have associated infinite strips

$$\Sigma_0, \Sigma_2, , ..., \Sigma_{2n-2} \subset \boldsymbol{R}^n.$$

We insist that cyclically consecutive strips are not parallel.

For each odd index k there are precisely two rank-1 sublattices having both strips Σ_{k-1} and Σ_{k+1} as fundamental domains. Both choices have the form $\boldsymbol{Z}v_k$ where v_k is a vector which is the difference between a pair of opposite vertices of the parallelogram $\Pi_k = \Sigma_{k-1} \cap \Sigma_{k+1}$.

We suppose that the kth white vertex is decorated by one of these sublattices. So, up to 2^k choices, our strips define for us a sublattice PET.

Quasi-Rationality: There is an equivalence amongst these sublattice PETs. If A is any affine transformation of the plane, then the strips

$$A(\Sigma_0), ..., A(\Sigma_{2n-2})$$

define an affinely conjugate PET provided that we make the correct choices for the sublattices. We say that the PET is *quasi-rational* if A may be chosen so that all the parallelograms considered above have integer area. This

notion is closely related to the notion of quasi-rationality in polygonal outer billiards. We will discuss this in the next chapter.

Parallelogram Rotations: There is an alternate way to view these kinds of PETs. Each planar parallelogram has a unique order 4 affine involution that rotates clockwise. For instance, the square $[-1,1]^2$ the map is the restriction of order 4 clockwise rotation about the origin. This special case is affinely equivalent to the general case. We call this map the *fundamental rotation* of the parallelogram, even though it is only a rotation in the affine sense.

Say that a *parallelogram tiling* of an infinite strip Σ is an almost partition of Σ into parallelograms which are all translates of each other. Say that a *parallelogram map* of Σ is the piecewise affine map obtained by applying the fundamental rotation to each parallelogram separately. It is a fun exercise to show that PETs we have constructed here are compositions of n parallelogram maps of the strip Σ_0.

In the quasi-rational case, the parallelogram widths are all commensurable, and we can scale so that they are integers. In this case, the union of all the parallelogram tilings (superimposed on top of each other) is periodic modulo a certain translation. This fact is the key to understanding quasi-rational outer billiards.

Quarter-Turn Compositions: There is yet another interpretation in which one alternates parallelogram maps based on rectangle tilings with shears of the strip. I call these *quarter turn compositions*. My paper [**S6**] takes this point of view and makes a detailed study.

For these strip maps, the first case of interest in all this is when $n = 3$. In this case the space of affinely inequivalent PETs is $3 + 3 + 3 - 6 = 3$ dimensional. The point here is that the affine group is 6-dimensional and the space of directed strips in $\boldsymbol{R}^2$ is 3-dimensional. I have studied this 3-dimensional space to some extent, though not in enough detail to report on any interesting results. This seems like an attractive object for further study.

3 Polygonal Outer Billiards

3.1 Basic Definitions

Outer billiards goes back at least to B. Neumann's short article [**N**]. J. Moser considered outer billiards as a toy model for planetery motion [**M1**], [**M2**]. I will concentrate on the case of polygons.

Polygonal outer billiards is a piecewise isometric map of $\boldsymbol{R}^2$ that is based on a convex polygon $P \subset \boldsymbol{R}^2$. Given a point $p_0 \in \boldsymbol{R}^2 - P$ one defines $p_1 = f(p_0)$ to be the point such that the line segment $\overline{p_0 p_1}$ is tangent to P uniquely at its midpoint, and a person walking from p_0 to p_1 would see P on the right. Figure 3.1 shows the construction.

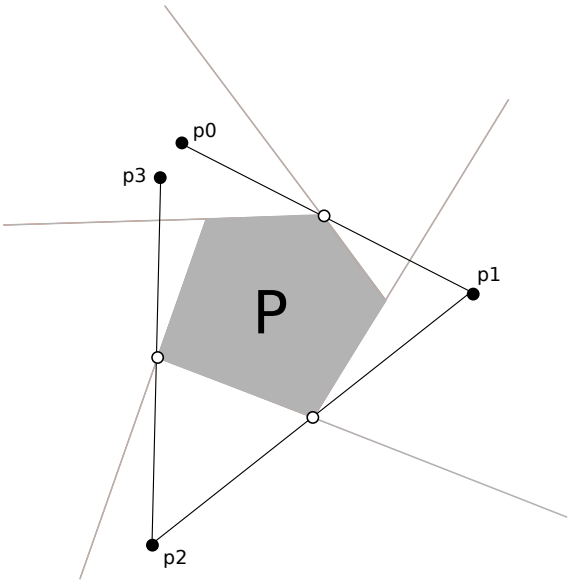

Figure 3.1: Definition of outer billiards

The points $p_1, p_2, p_3, ...$ are the forward iterates of p_0 under the outer billiards map. The map is defined in the complenent of the grey rays which emanate from the edges of P like a pinwheel. The inverse map is defined by a similar construction, with the words *left* and *right* swapped. For any n, the first n iterates of f, both forwards and backwards, are defined in the complement of a finite union of lines. In particular, the full orbit of the map is defined in the complement of a countable union of lines.

It is sometimes nicer to consider the square map $f^2 = f \circ f$ instead. This map is a piecewise translation.

3.2 Some Examples

Triangles and Parallelograms: Outer billiards is an affinely natural system. If A is an affine transformation, then A conjugates the outer billiards map with respect to P to the one with respect to $A(P)$. Thus, for instrance, one can understand outer billiards with respect to any triangle by understanding the equilateral triangle case. Likewise, one can understand outer billiards with respect to any parallelogram by understanding the square case. For the case of equilateral triangles, the initial triangle extends to the usual tiling of $\boldsymbol{R}^2$ by equilateral triangles. Each of these tiles is a periodic island. The square case has a similar description in terms of the square tiling.

The Regular Pentagon: Here is what outer billiards looks like on the regular pentagon.

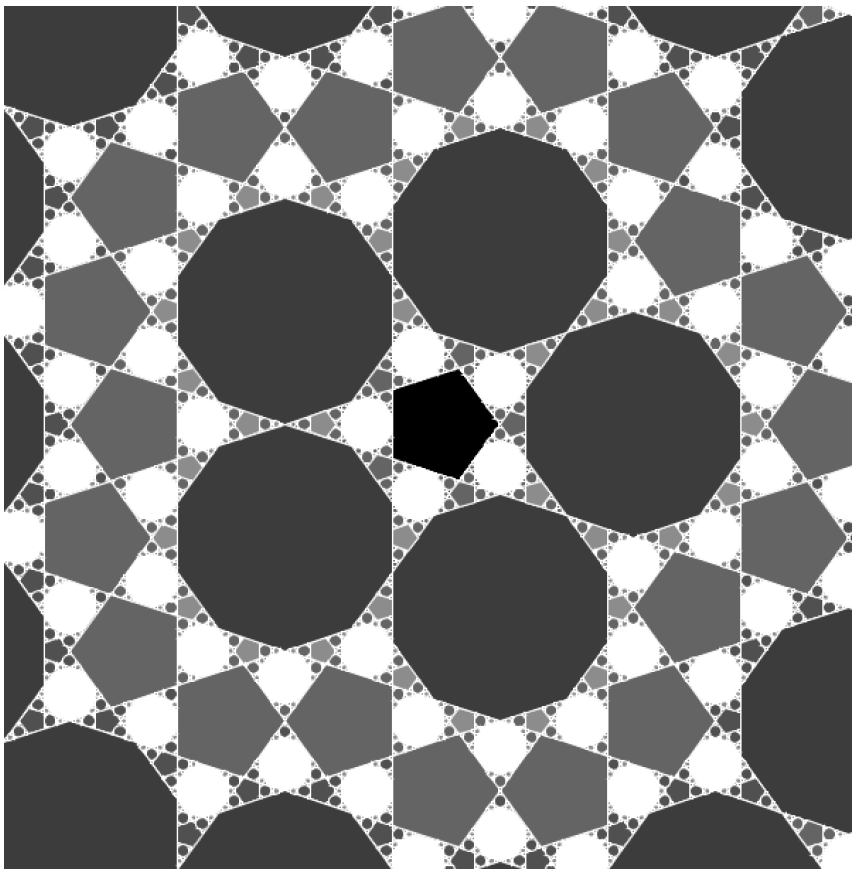

Figure 3.2: Outer billiards on the regular pentagon

The structure in this case is first analyzed in [**T2**]. There is a dense set of periodic islands, either regular pentagons or regular decagons. There is an infinite union of "pentagon necklaces" made from pentagons isometric to the central one. These necklaces alternate with an infinite union of "decagon necklaces". Figure 3.2 just shows the first of each necklace, and about half of the second decagon necklace.

Between these necklaces there is a self-similar pattern that is reminiscent to the isosceles triangle example considered in §2.3. S. Tabachnikov establishes this self-similar property using a similar kind of renormalization scheme. The picture between consecutive decagon necklaces is essentially independent of the position. To say this somewhat more precisely, we can consider the intersection of the picture with a horizontal strip. Once we go sufficiently far out to the right, the picture becomes periodic. Figure 3.3 illustrates what we mean.

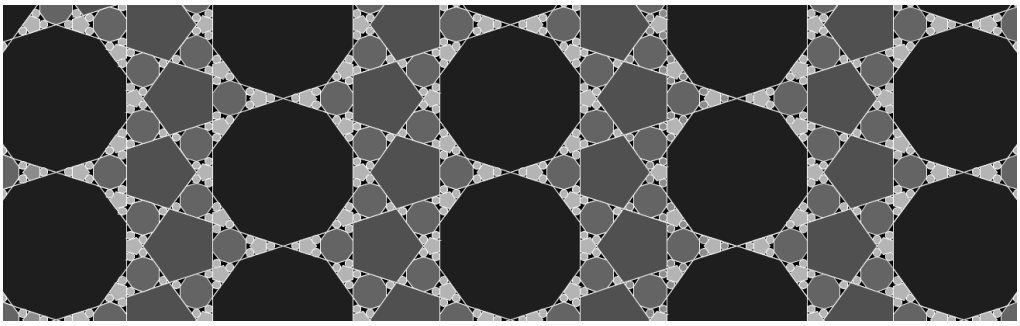

Figure 3.3: Outer billiards (again) on the regular pentagon

The Regular Octagon: Figure 3.4 shows some of the periodic islands for outer billiards on the regular octagon. Once again we have an infinite sequence of octagon necklaces with a fractal pattern in between them. We are showing the innermost later. Notice that the pattern inside the highlighted parallelogram exactly matches the one in Figure 2.8. Considered as a semi-regular octagon, the regular octagon has parameter $s = \sqrt{2}/2$. The parameter is such that the highlighted rectangle has the same parameter as the fundamental domain $X_{0,s}$.

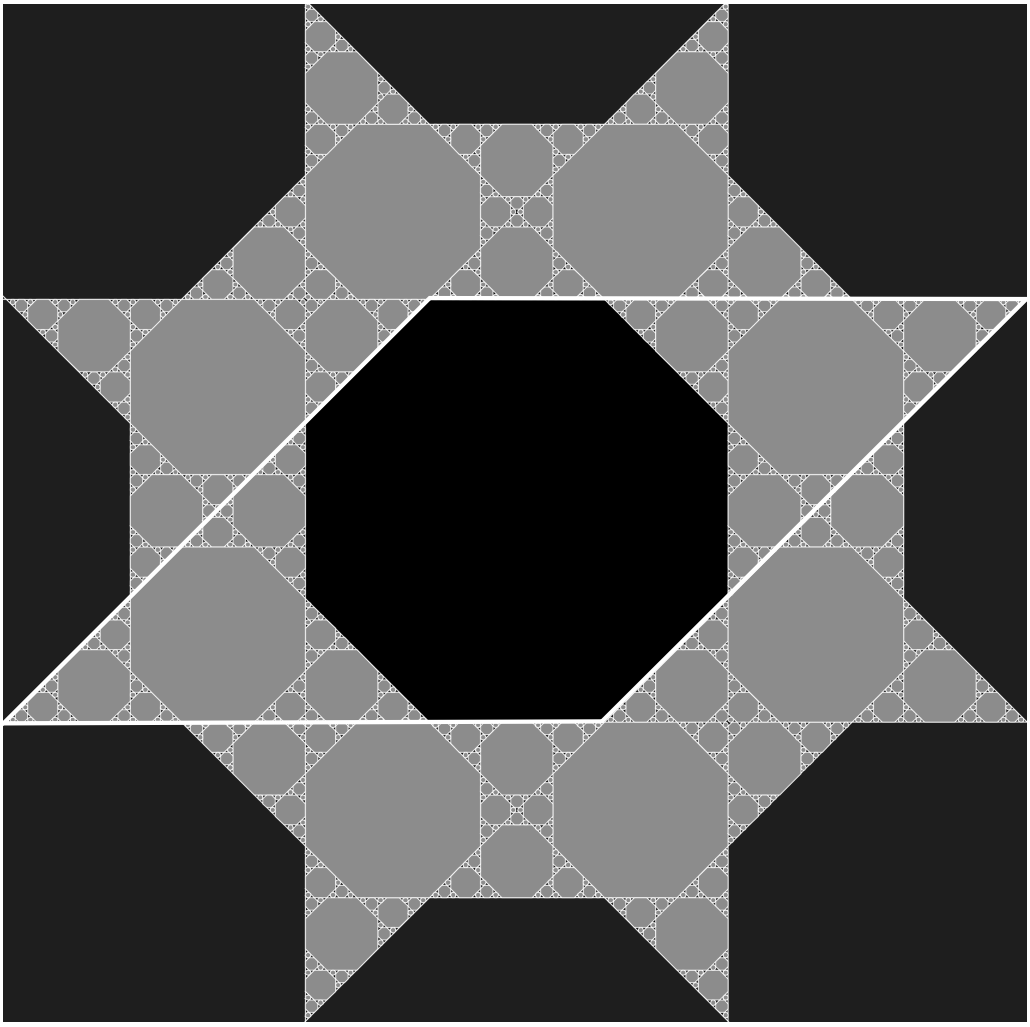

Figure 3.4: Outer billiards on the regular octagon

Semi-Regular Octagons: Figure 3.5 shows outer billiards with respect to the the semi-regular octagon with pattern $s = 3/2 - \sqrt{3}/2$. Again, we are just showing the inner layer. The fractal pattern that lives between the octagon necklaces matches the one for the octagonal pet with the same parameter. The given parameter s is such that $R(s)$ is the parameter for the octagonal PET shown in Figure 2.9. You can see the close resemblance. Were we to plot the picture for the pet f_s with $s = 3/2 - \sqrt{3}/2$ we would get exactly the picture inside the white parallelogram.

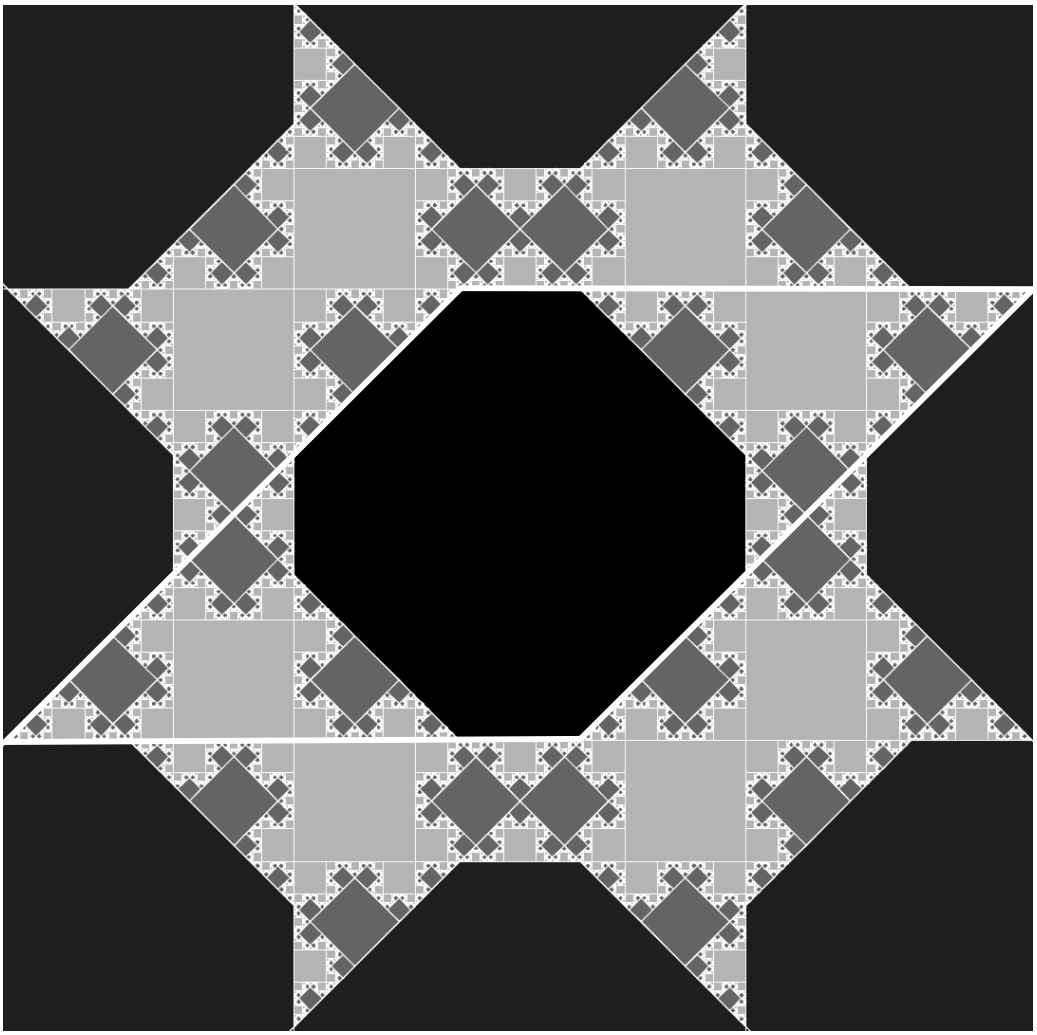

Figure 3.5: Outer billiards on the regular octagon

Let $O(s)$ be the semi-regular octagon with parameter $s \in (1/2, 1)$. We proved in [**S7**] that $\boldsymbol{R}^2 - O(s)$ is tiled by congruent parallelograms, each isometric to $X_{0,s}$, such that the union of periodic islands in each parallelogram is isometric to the corresponding union for the octagonal PET $f_s : X_{0,s} \to X_{0,s}$. Technically, Figures 3.4 and 3.5 are illustrating the phenomenon for a central parallelogram that is not part of our parallelogram tiling, but one could also formulate a version of the result that works for this central parallelogram.

The dynamics of outer billiards is not quite the same as the dynamics in the corresponding octagonal PET. The two systems are "commensurable" in

a sense made precise in [**S7**]. In any case, the octagonal PETs control the structure of the periodic islands for outer billiards in semi-regular octagons.

There is an important point to note. The octagon $O(s)$ is only defined when $s \in (1/2, 1)$ and the octagonal PETs are defined when $s \in (0, 1)$. Our renormalization scheme for octagonal PETs works for the whole interval $(0, 1)$. One could consider the first return map to $(1/2, 1)$ but this is a bit awkward.

What I am trying to say is that to notice that outer billiards on semi-regular octagons has a renormalization scheme, one has to enlarge the class of systems one needs to consider. Perhaps this phenomenon occurs more generally in polygonal outer billiards. A natural generalization of the family of semi-regular octagons is the family of all polygons which have fixed interior angles. So, all members of the family are obtained by "rolling" the sides parallel to themselves. I have not explored this, however.

Regular Polygons: We have already discussed outer billiards on regular n-gons for $n = 3, 4, 5, 8$. The case $n = 6$ resembles the cases $n = 3, 4$ in that the periodic islands make a global and regular tiling. It is a good exercise to work out the picture. The cases $n = 10$ and $n = 12$ are similar to the cases $n = 5, 8$. Bedaride and Cassaigne [**BC**] study the symbolic dynamics of these maps in detail. I believe [2] that much progress has been made on outer billiards for the regular 7-gon, though this case is considerably more intricate.

One recent general result is that outer billiards on a regular n-gon has periodic orbits provided that $N > 4$ and either N is odd or $N/2$ is an odd integer. See [**KRTZ**]. Beyond this result, and the ones mentioned above, not much is known about outer billiards on regular n-gons. Gordon Hughes has spent many years producing beautiful and intricate numerical studies of these. See [**H**].

3.3 The Pinwheel Map

In §2.6 we discussed a certain kind of sublattice PET which we called a strip map. Here we relate that construction to outer billiards. The results mentioned here are proved in [**S8**]. As in [**S8**] we work with polygons having no

[2] I seem to remember that Alexei Kanel-Belov was one of the authors involved, but I couldn't find a preprint.

parallel sides. I have not carefully thought through the construction when P has some parallel sides, but I know how to do it for the regular octagon. See [**S9**].

The Strips: Each edge e of P defines an infinite strip X. One boundary component of X is the line extending e. The centerline of X contains the (unique) point of P farthest from e. Figure 3.6 shows the construction.

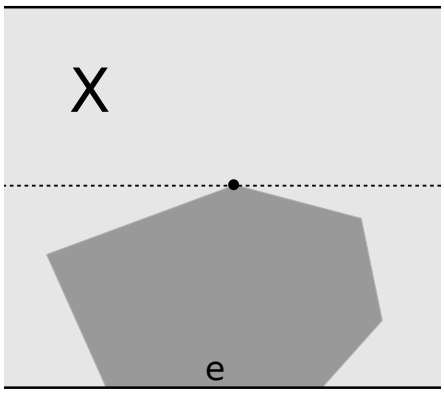

Figure 3.6: The strip X defined by (P, e).

The n-gon P thus determines n-strips. A very large circle centered at the origin intersects these strips in a certain order, and the order stabilizes once the circle is sufficiently large. We order the strips according to their counter-clockwise intersection with such large circles. Put another way, we order the strips according to their slopes in $\boldsymbol{R} \cup \infty$. We label these strips $X_0, X_2, ...$ See Figure 3.7 below.

The Sublattices: For each consecutive pair of strips (X_{k-1}, X_{k+1}) there is a unique choice $L_k = \boldsymbol{Z} v_k$ of sublattice such that

1. $X_{k\pm 1}$ is a fundamental domain for L_k.
2. Lines parallel to v_k very far from the origin intersect X_{k-1} and X_{k+1} consecutively.

Figure 3.7 shows the correct choice for an example. Here are take $k = 1$. We draw the a segment for v_k instead of a vector because $\pm v_k$ works here, independent of sign. We are not drawing the polygon P here; in fact we just made up this artificial example by putting down 5 strips

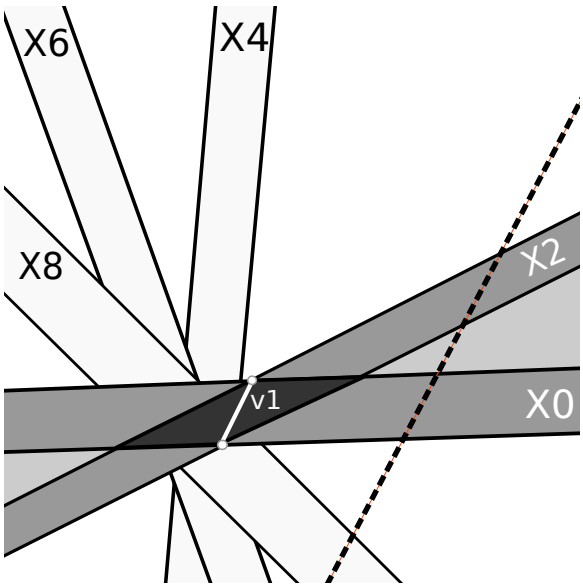

Figure 3.7: Defining the lattice L_1 in terms of the strips.

We introduce the "double wedge" W_k as the union Σ_{k-1} and Σ_{k+1} and the two infinite cones between these strips. This set is highlighted in Figure 3.7. (It is made from 4 different shades of grey.) Let f^2 be the square of the outer billiards map f. The basic geometric principle is that if $p \in W_k - X_{k+1}$, and p is sufficiently far from the origin, then $f^2(p) = p \pm v_k$, where the sign is chosen so that $p \pm v_k$ is one step closer to X_{k+1}.

Shallow and Deep Equivalence: The discussion above has the following consequence. If $p \in X_0$ is very far from the origin then f^2 moves p along a line parallel to v_1 until the point lands in X_2, then f^2 moves p along a line parallel to v_3 until the point hits X_4, etc. When p is close to the origin, the relationship between the sublattice PET and f^2 is more subtle, and we will discuss this below. Let ψ denote the second return map of f^2 to X_0. The reason we consider the second return map is that we want to consider orbits which circulate all the way around P and not just halfway around. Let ψ^* denote the square of the PET defined by our decoration of the bipartite $2n$-cycle. We take the square here to match the behavior of ψ. We call ψ^* the *pinwheel map*.

The discussion above justifies the result that $\psi = \psi^*$ outside a compact subset of $\boldsymbol{R}^2$. We call this fact *shallow equivalence*, because it is not very hard to prove. The discussion above practically constitutes a proof.

We now mention the main result of [**S8**], which establishes a much deeper equivalence between the two maps.

Theorem 3.1 (Deep Equivalence) *There exists a large disk $\Delta \subset \boldsymbol{R}^2$, centered at the origin, with the following two properties:*

- *Let $p \in \boldsymbol{R}^2$ be any point for which the first n iterates of ψ are defined, and both p and $\psi^n(p)$ are outside K. Then there is some $n^* > 0$ such that the first n^* iterates of ψ^* are defined on p and $(\psi^*)^{n^*}(p) = \psi(p)$.*
- *Let $p \in \boldsymbol{R}^2$ be any point for which the first n iterates of ψ^* are defined, and both p and $(\psi^*)^{n^*}(p)$ are outside K. Then there is some $n > 0$ such that the first n^* iterates of ψ are defined on p and $(\psi^*)^{n^*}(p) = \psi(p)$.*

The analogous result holds for $n < 0$. We have stated the case $n > 0$ just for ease of exposition.

Theorem 3.1 says that near K the action of outer billiards and the pinwheel map, though perhaps quite different from each other, is "removable". The two different dynamical systems might do very different things to a point once it wanders into K, but the point emerges from K in the same way for both systems. I like to describe this theorem intuitively as saying that "what happens in Las Vegas stays in Las Vegas". The unequal behavior of the two systems is entirely trapped in K, and it has no impact on what happens outside of K.

Theorem 3.1 immediately implies, for instance, that ψ has an unbounded orbit if and only if ψ^* has an unbounded orbit.

3.4 Quasi-Periodic Polygons

We call the polygon P *quasi-rational* is the associated sublattice PET is quasi-rational. Let us unpack this definition. The polygon P defines the strips $X_0, X_2, ...$ as above. We have the parallelograms $\Pi_1, \Pi_3, ...$ where

$$\Pi_k = X_{k-1} \cap X_{k+1}.$$

The polygon P is quasi-rational if it may be scaled so that all these parallelograms have integer areas. Semi-regular and regular polygons are all quasi-rational, and so are polygons whose vertices have rational coordinates.

One well known result in polygonal outer billiards is that all orbits are bounded for quasi-rational polygons. This result is proved by three teams of

authors. See [**VS**], [**K**], and [**GS**]. Here I will sketch a proof. In some sense all the proofs are the same. The proof here only uses shallow equivalence, so it is quite elementary.

We keep the notation from the previous section. We think of our sublattice PET as a product of parallelogram rotations, as described in §2.6. As we mentioned in §2.6, the superposition of all the parallelogram tilings of X_0 is periodic in the quasi-rational case. The intersection of all the strips is precisely the polygon P. This means that around the origin there is a certain collection of parallelograms, one per tiling, whose intersection is P. Call this the big intersection "the sweet spot".

Corresponding to the sweet spot, there is a copy of P on which ψ acts as the identity. But then there are infinite many translated copies of the sweet spot and this infinitely many translated copies of P on which ψ^* acts as the identity. Since $\psi = \psi^*$ outside a compact set, there are infinitely many copies of P on which ψ acts as the identity. Figure 3.8 shows what this looks like.

Figure 3.8: Periodic array of copies of P.

By construction P has a unique vertex on the centerline of the strip X_0. Let Q be the result of reflecting Q in this vertex. Let P' be any copy of P in our periodic array that is sufficiently far from the origin. By construction, $f(P') = Q'$, where Q' is the corresponding copy of Q. Here f is the outer billiards map. Figure 3.9 shows $f^{2k}(P)$ for $k = 0, 1, 2, 3$ and $f^{2k}Q)$ for $k = 0, 1, 2$. By construction, the consecutive images share a vertex in common.

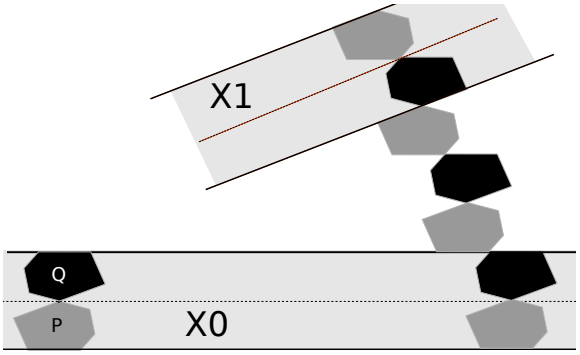

Figure 3.9: The formation of a necklace

We have depicted the case when $f^4(Q) \in X_2$. In general there will be some first $k > 0$ when either

- $f^{2k}(P) \in X_2$ and all previous polygons lie in $W_1 - X_2$.
- $f^{2k}(Q)] \in X_2$ and all previously defined polygons lie in $W_1 - X_2$.

The choice of P or Q depends on whether or not the edges defining X_0 and X_2 are adjacent or not. We are showing the non-adjacent case. In either case, we now consider the images between X_2 and X_4. The significant thing is that P sits inside X_2 just as it sits inside X_0. Repeating all the same steps, we see that the orbit of P' continues into W_3, starting and stopping in X_2 and X_4.

Continuing all the way around, we produce a necklace orbit, one consisting of cyclically tangent isometric copies of P. Compare Figure 3.2. But we can make this construction for any P' sufficiently far from the origin. This gives us an infinite sequence of necklace orbits. All other orbits are trapped between the necklaces and hence bounded. Compare Figure 3.3. That's it.

The Rational Case: When P has rational vertices, we can scale so that all the vertices have integer coordinates. The map f^2 is then a piecewise translation involving a finite number of integer vectors. In particular, if we start with any point $p \in \boldsymbol{R}^2 - P$ with a well-defined orbit, the orbit of p is bounded and consists of points which differ from p by integer vectors. Hence, this orbit must be finite. Hence, all orbits are periodic in this case. This is an especially appealing situation, because every rational polygon gives rise to a dynamically invariant tiling of $\boldsymbol{R}^2$ by periodic islands.

Culter's Theorem: C. Culter proved [3] hat outer billiards has a periodic orbit outside every compact subset of $\boldsymbol{R}^2$ with respect to any polygon. Here is a sketch of this result. (I have never read Culter's proof.) When the sublattice PET is quasi-rational, the necklace orbits furnish infinitely many periodic points. In the general case, we see an infinite sequence of polygons in X_0 which converge to P up to translations of X_0. Each such polygon is the intersection of the parallelograms in the tiling that contain it. For this reason ψ^* is the identity on such polygons. But then, by the shallow equivalence, the same is true for ψ far from the origin. This gives periodic orbits outside any compact set.

[3] I don't know if Culter ever wrote a paper about his result, but see [**T1**] for an account

3.5 The Arithmetic Graph

We continue with the notation from the previous section. In particular, recall that $\psi : X_0 \to X_0$ is the square of the PET map. Let $p_0 \in X_0$ be some point with a well-defined orbit. Referring to Equation 6, define

$$p_{2k+2} = \phi_{b_{2k+1} w_{2k+1} b_{2k}}(p_{2k}) \in X_{2k+2}.$$

We mean to define this for $k = 1, ..., 2n$ and we take the indices cyclically mod $2k$. Far from the origin, the points $p_0, p_2, p_4, ..$ are just the successive points of the f^2-orbit of p which lie in the strips.

There is a non-negative integer N_k such that $p_{2k+2} - p_{2k}$ is N_k times one of the two generators of the sublattice L_{2k+1}. Far from the origin, N_k simply tells the number of times we have to apply f^2 to get from p_{2k} to p_{2k+2}. We define the *spectrum* of p to be the length-n vector

$$S(p) = (N_0 - N_{2n}, N_2 - N_{2n+2}, ..., N_{2n-2} - N_{4n-2}) \in \boldsymbol{Z}^n. \tag{10}$$

Typicallly all the entries of S_p lie in $\{-1, 0, 1\}$, and for all I know this always happens.

Supposing that we have chosen a linear projection $\Lambda : \boldsymbol{R}^n \to \boldsymbol{C}$ we define the Λ-spectrum of p to be $\Lambda \circ S(p)$. One choice of projection is given by

$$\Lambda_k(S) = \sum_{k=1}^{n} S_k \omega^k, \qquad \omega = \exp(2\pi k i/n). \tag{11}$$

We call the corresponding spectrum $\Lambda_k \circ S$ the *k-cyclotomic spectrum*.

Assuming we have fixed Λ, we can associate to a periodic point p a periodic path $\Lambda(p) \subset \boldsymbol{C}$. Letting $\{p^k\}$ be the ψ^*-orbit of $p = p^0$, we form the path whose vertices are

$$\Lambda \circ S(p^0), \qquad \Lambda \circ S(p^0) + \Lambda \circ S(p^1), \qquad \Lambda \circ S(p^0) + \Lambda \circ S(p^1) + \Lambda \circ S(p^2), ...$$

When we use the k-cyclotomic spectrum, we call $\Lambda(p)$ the *k-cyclotomic graph* of p. The cyclotomic case is adapted to the case of regular polygons.

We now return to outer billiards on the regular pentagon. Recall that this system has two kinds of periodic islands: pentagonal and decagonal. Figure 3.11 shows the 2-cyclotomic graph for a pentagonal periodic island on the left and for a decagonal periodic island on the right. One should compare this with Figure 2.6.

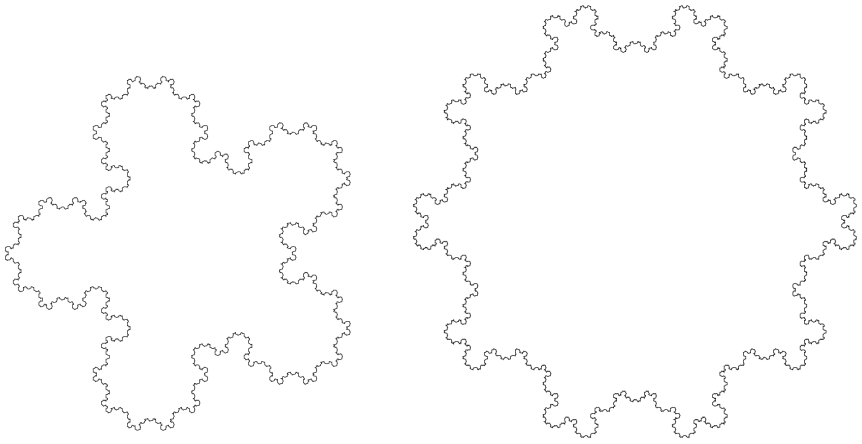

Figure 3.10: The 2-cyclotomic graph in the regular pentagon case

The 3-cyclotomic graph looks the same. The 1-cyclotomic graph and the 4-cyclotomic graph are not interesting; they essentially just reproduce the orbit.

Our computer program makes sense of the pinwheel map for regular octagons. For the regular octagon the 2-cyclotomic graph and the 3-cyclotomic graph are different. Figure 3.11 shows what the 3-cyclotomic graph looks like. This is quite similar to what happens for the regular pentagon.

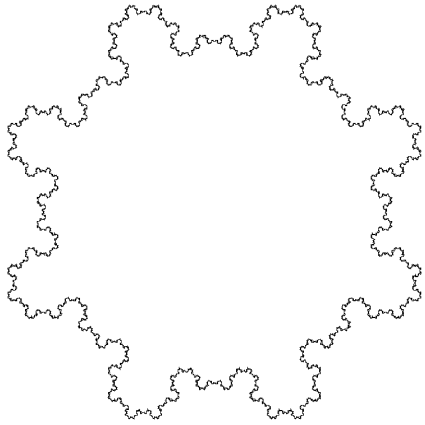

Figure 3.11: The 3-cyclotomic graph in the regular octagon case

Figure 3.12 shows the two typical pictures of the 2-cyclotomic graph for long periodic orbits. The first case is a closed polygon, though not embedded.

The rescaled limit of this thing is an object somewhat reminiscent of the classic square Sierpinski carpet. The second case is an open polygonal path, and we are showing several periods of it.

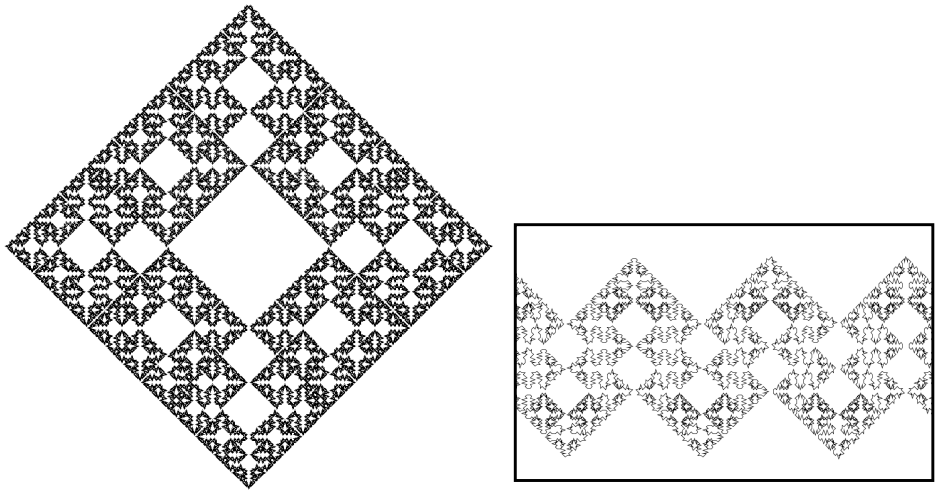

Figure 3.12: The 2-cyclotomic graph in the regular octagon case

I wrote a long article [**S9**] exploring the arithmetic graph in the regular octagon case, though I never published it. My program also draws the cyclotomic graphs with respect to other regular polygons. The interested reader can play around with the software (or write their own) and see other neat pictures.

4 Unbounded Orbits for Outer Billiards

4.1 Some Results and Questions

The Moser-Neumann problem [**N**], [**M1**], [**M2**] asks about the existence of unbounded orbits for outer billiards. I have already mentioned the result [**VS**], [**K**], [**GS**] that all orbits are bounded for quasi-rational polygons. Another result along these lines, due to Dan Genin [**Ge**], is that all orbits are bounded for trapezoids. Trapezoids fall outside the quasi-rational framework. See the introduction of [**S3**] for a survey of other boundedness results for non-polygonal shapes.

I answered the Moser-Neumann question in 2007 by proving that outer billiards on the Penrose kite has an unbounded orbit. See [**S5**]. The Penrose kite is the convex quadrilateral that arises in the famous Penrose kites-and-darts tiling. I showed, in particular, that the point p in Figure 4.1 has a well defined and unbounded orbit when outer billiards is defined with respect to the kite shown in Figure 4.1. The auxiliary lines in the picture are present to show the construction of the kite and the special point.

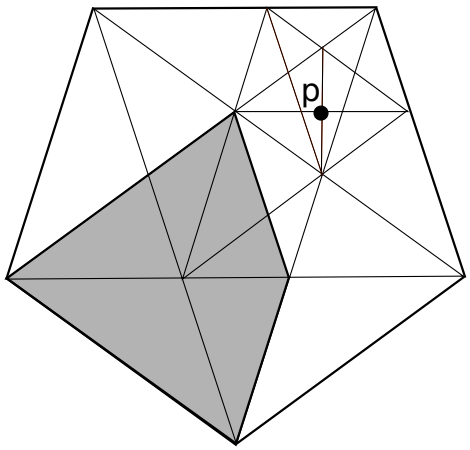

Figure 4.1: The Penrose kite and a point with an unbounded orbit.

My initial proof of this result exploited a quasi-self-similarity property of the arithmetic graph associated to this orbit. See Figure 4.X below. Initially I derived this by hand, in some sense. Later on, I found a renormalization scheme which explained this. I will discuss all this below.

In [**S3**] I proved a more general result, concerning the kites $K(A)$ having vertices $(-1,0)$ and $(0,\pm 1)$ and $(A,0)$ with $A \in (0,1)$. The Penrose kite is

affinely equivalent to $K(\sqrt{5}-2)$. Note that $K(A)$ is a rational polygon when $A \in \boldsymbol{Q}$ and not quasi-rational when $A \in \boldsymbol{R}-\boldsymbol{Q}$. My main theorem in [**S3**] is that outer billiards on $K(A)$ has unbounded orbits whenever $A \in (0,1)-\boldsymbol{Q}$. The orbits I found all lie on the set $\Sigma = \boldsymbol{R} \times \boldsymbol{Z}_{\rm odd}$, where $\boldsymbol{Z}_{\rm odd}$ is the set of odd integers.

I discovered quite a bit about these orbits. For instance

- The orbits are *erratic* in the sense that they enter every neighborhood of the points $(0, \pm 1)$ and also exit every compact subset of $\boldsymbol{R}$. So, they go in and out, in and out, repeatedly. Most orbits are erratic in both the forwards and backwards direction; a few are erratic in just one direction.

- The orbit intersects the set $(0,2) \times \{1\}$ in a set which equals a Cantor set up to deleting a countable number of points. The first return to this near Cantor set is conjugate in an explicit way to the "plus one map" in a profinite group.

I will illustrate some of this structure in the next section.

Let me also mention that the orbit corresponding to the point in Figure 4.1 is very special. It is erratic in the forwards direction and not in the backwards direction. In the backwards direction the orbit intersects every compact set in finitely many points.

Shortly after [**S3**], D. Dolgopyat and B Fayyad [**DF**] proved that outer billiards has unbounded orbits when defined relative to a semi-disk, the set you get when you cut a disk in half. They also proved this result for the sets you get when you nearly cut the disk in half. Their methods are completely different, and their unbounded orbits march straight out to infinity.

These are the only known unboundedness results. One natural question is:

Conjecture 4.1 *Outer billiards has unbounded orbits with respect to almost every polygon.*

Here is a more precise conjecture:

Conjecture 4.2 *Suppose that P is a convex polygon with no parallel sides. If P is not quasi-rational then outer billiards has unbounded orbits with respect to P.*

This second conjecture is so strong that I am not completely certain it is true. One should probably just take it as a question.

4.2 The Kite Result Illustrated

In this section I am going to use the arithmetic graph to give some ideas behind my result about unbounded orbits for outer billiards on irrational kites. In discussing the ideas, I will depart somewhat from the treatment in [**S3**] in order to better relate to the ideas we have presented above. My idea is to present things here in a way that might generalize more naturally than the special treatment given in [**S3**].

The first idea in our analysis is to understand what happens for $K(A)$ when $A = p/q$ is rational. In our proof we work with pq odd but here we will show pictures for pq even. The pictures are nicer. We let ψ_A denote the pinwheel map associated to $K(A)$, and we let $\Lambda(S_0, S_1, S_2, S_3) = (-S_0, S_2)$. We define the arithmetic graph relative to this choice of Λ. The set Σ is invariant under the outer billiards map, and so it makes sense to look for unbounded orbits on this set. We label so that X_0 is the strip shown in Figure 4.2.

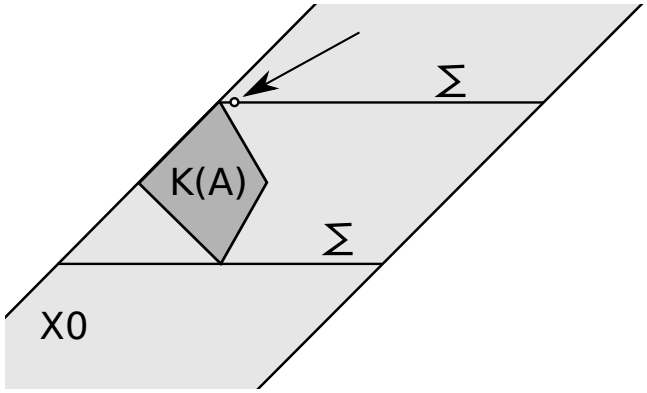

Figure 4.2: $\Sigma \cap X_0$, and the fundamental start.

Given the deep equivalence between outer billiards and pinwheel map – a result which is quite easy to prove for kites – it suffices to prove that $\psi : X_0 \cap \Sigma \to X_0 \cap \Sigma$ has unbounded orbits. This map is an infinite interval exchange transformation. When $A = p/q$, all the intervals in the exchange have length which is an integer multiple of $2/q$. For this reason, all points in the set

$$(0, 2/q) \times \{1\} \tag{12}$$

have the same combinatorial orbit. We call the point $(1/q, 1)$ the *fundamental start* and its pinwheel orbit the *fundamental orbit*. We will show pictures of the arithmetic graph of the fundamental orbit.

We first mention a few general features of these paths. The arithmetic graphs are all *lattice polygonal paths*. That is, their vertices are integer points. I proved in [**S3**] that the arithmetic graphs are embedded, and that consecutive vertices are always adjacent in $\boldsymbol{Z}^2$. That is, each coordinate of a given point differs by at most 1 from the corresponding coordinate of the adjacent points. Given the integer nature of the path, a fractal-looking example much be really huge. The point is that such a monster is meant to be scaled so that its vertices are integral.

Define the *baseline* to be the line of slope $-A$ through the origin. The distance between a point on the arithmetic graph and the baseline is comparable to the distance from the corresponding point in the orbit to the origin. Thus, if the arithmetic graph is very tall, then some points in the orbit are very far from the origin. We are interested in finding points having an arithmetic graph that rises unboundedly far above the baseline.

At the same time as this, the points of the arithmetic graph that are within 1 unit of the baseline correspond to the intersection of the orbit with the special interval $(0,2) \times \{1\}$ mentioned above. This is the top interval labeled Σ in Figure 4.2. So, the fine scale structure of the orbit in this interval can be read off the part of the arithmetic graph near the baseline.

Now we show some pictures. Again, when $A = \sqrt{5} - 2$ the kite $K(A)$ is affinely equivalent to the Penrose kite. The first 4 continued fraction approximations to A are $1/4$ and $4/17$ and $17/72$ and $72/305$. The numbers $1, 4, 17, 72, 305$ satisfy the recurrence relation $a_{n+2} = 4a_n + 1$, and the pattern continues. Figure 4.3 below shows the arithmetic graph of the fundamental orbit for these first 4 fractions.

It looks like these pictures are converging to a fractal. When they are scaled so that the edges have about unit length, their height tends to ∞. This strongly suggests that there are unbounded orbits. However, we cannot take the limit of the fundamental starting point, because this point converges to the kite vertex as the approximating rational converges to $\sqrt{5} - 2$. We need to take a limit in a careful way.

It looks like these pictures are converging to a fractal. When they are scaled so to have integral vertices, their height above the baseline tends to ∞. This strongly suggests that there are unbounded orbits. However, we cannot take the limit of the fundamental starting point, because this point converges to the kite vertex as the approximating rational converges to $\sqrt{5} - 2$. We need to take a limit in a careful way.

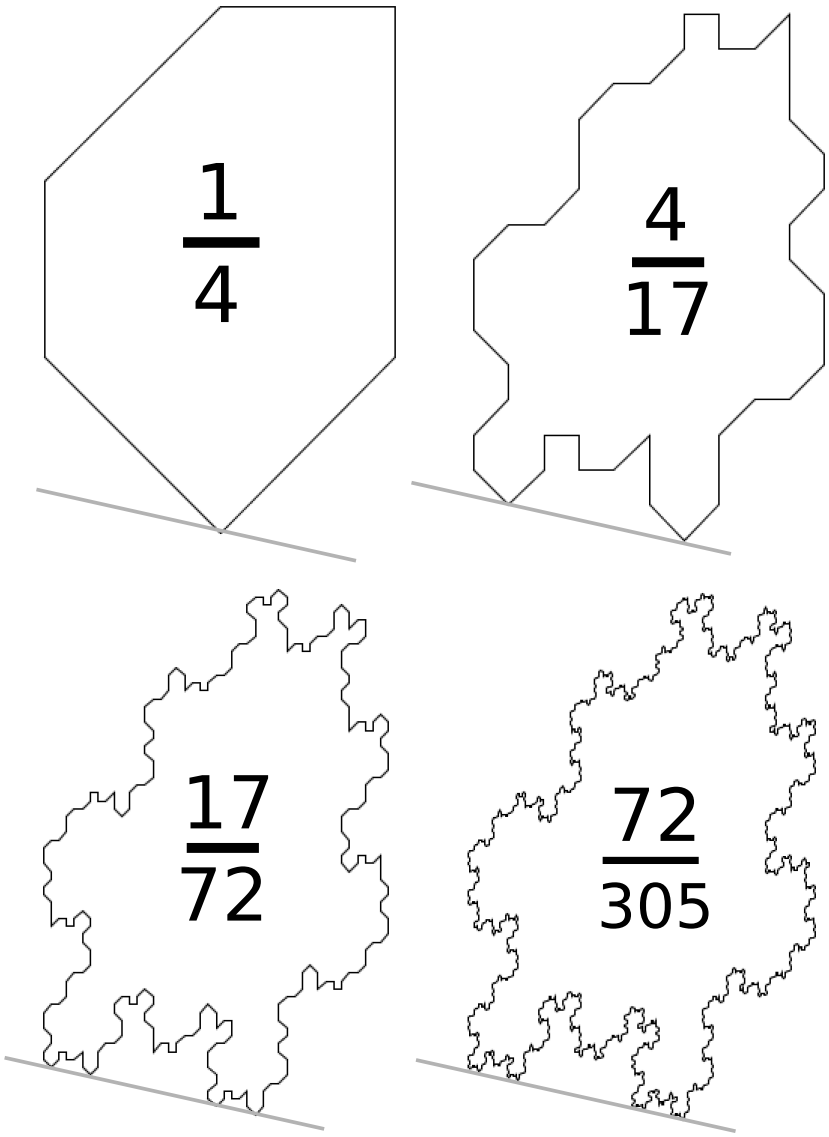

Figure 4.3: The arithmetic graph for 1/4, 4/17, 17/72, and 72/305.

There is something else to notice about these pictures. The bottom parts of the picture seem to be touching – or, rather, nearly rouching – the baseline in a Cantor set. This corresponds to the fact that the fundamental orbit is becoming dense in a Cantor set in $(0, 2) \times \{1\}$ as the continued fraction approximant converges to $\sqrt{5} - 2$. Many points on this Cantor set have well-defined orbits, and I show that these points have unbounded orbits.

Here is another especially pretty one. The rational 408/985 is a continued fraction approximant of $\sqrt{2}-1$. One can see the same kind of Cantor set formation near the baseline.

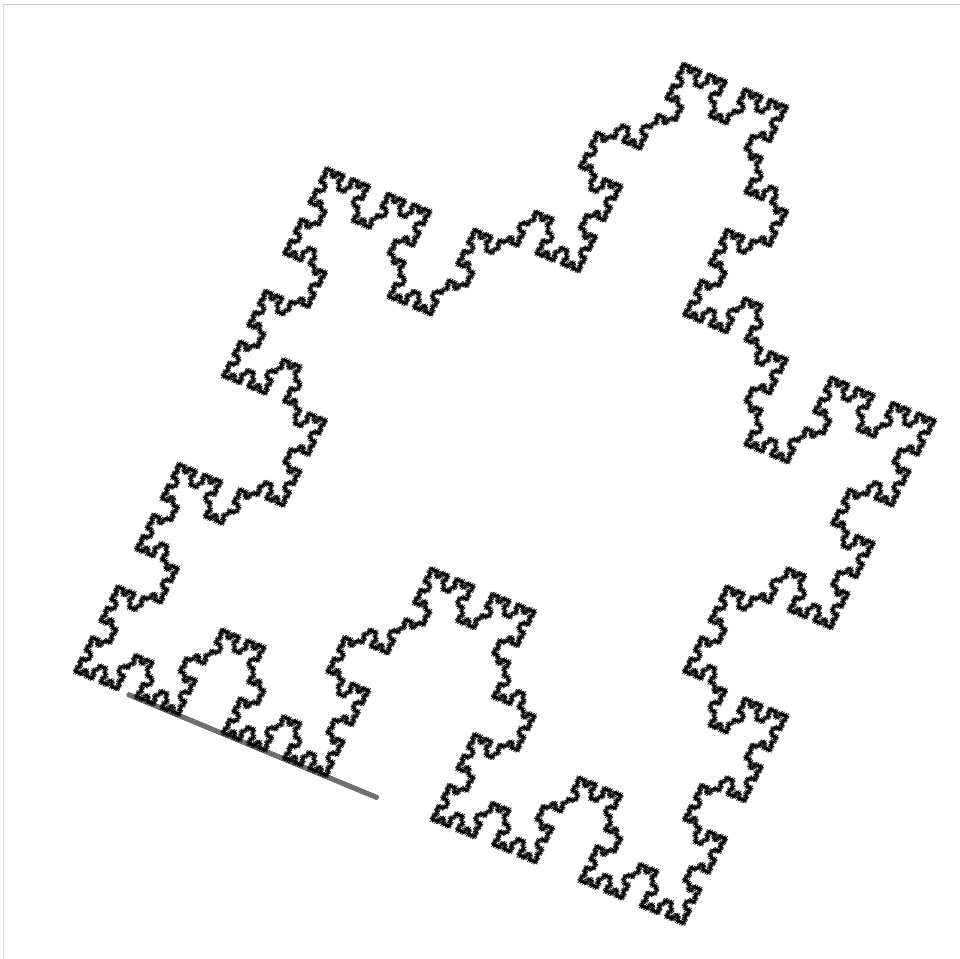

Figure 4.4: The arithmetic graph for 408/985.

The numbers $\sqrt{5}-2$ and $\sqrt{2}-1$ are fixed points of a certain renormalization operator that is similar to the one R described in §2.5. For a general choice of irrational A we also observe this Cantor set phenomenon, but the Cantor set is not as regular.

Let me illustrate the general phenomenon with one example. Two rationals p_1/q_1 and p_2/q_2 are *Farey adjacent* if $|p_1q_2 - p_2q_1| = 1$. The 4 rationals involved in the next picture are consecutively farey adjacent. We plot the

corresponding 4 arithmetic graphs at the same scale. One can see that the consecutive polygons are nested inside of each other. Each one is somehow part of the next one.

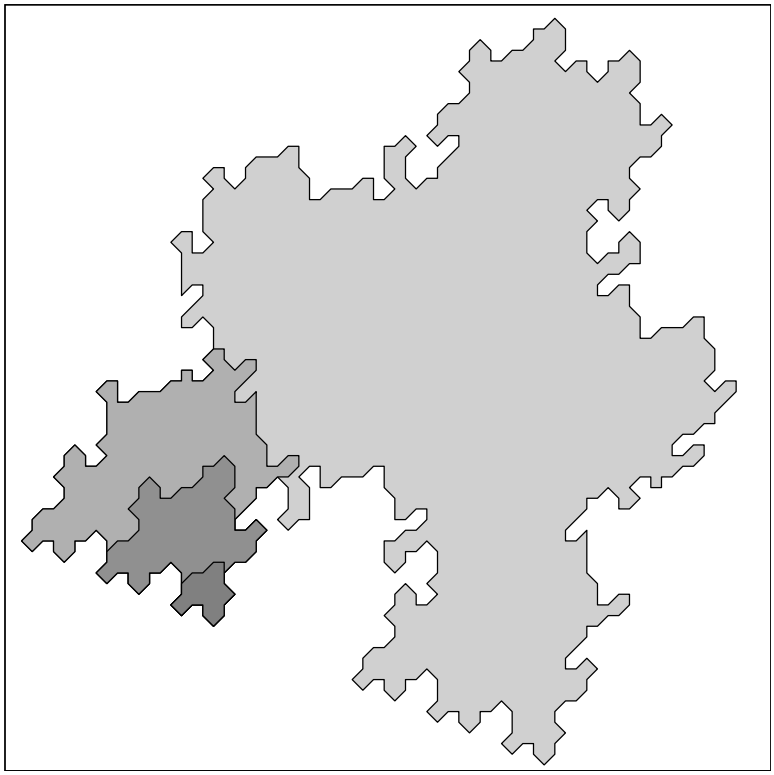

Figure 4.4: The arithmetic graph for 2/5 and 5/12 and 8/19 and 21/50.

This general nesting phenomenon is responsible for the Cantor set formation with respect to any irrational A. (I formulate this somewhat differently in [**S3**].) I want to emphasize that this is really just an impressionistic account of the unbounded orbits result. See [**S3**] for details.

One thing I should say is that I did not actually establish the strict nesting property you see in Figure 4.4. Even though I did not prove it, it seems that this strict nesting phenomenon always happens. Rather, I considered the graphs corresponding to rationals of the form p/q with pq odd, and I proved that the larger graph copies the bottom part of the smaller graph. What I am trying to say is that you see some extra-beautiful phenomena in the pictures which I did not actually prove. I just proved enough to get the main results.

4.3 Compactification

Let $\psi^* : X_0 \to X_0$ be a pinwheel map. It turns out that there is always a higher dimensional double lattice PET $\widehat{\psi} : \widehat{X}_0 \to \widehat{X}_0$ and an injective semi-conjugacy $\Theta : X_0 \to \widehat{X}_0$. This means that

$$\widehat{\psi} \circ \Theta = \Theta \circ \psi^*. \tag{13}$$

The map Θ is injective if and only if ψ^* is not quasi-rational. The dimension of the closure image is the $\boldsymbol{Q}$-rank of the list $A_1, ..., A_n$ of areas of parallelograms associated to ψ^*. So, in the quasi-rational case, this is 1 and otherwise it is greater than 1.

Generically, the $\boldsymbol{Q}$-rank is n, the number of parallelograms, and the map Θ has dense image. I proved all this in a long and unpublished preprint [**S6**]. So, the punchline here is that generically the pinwheel map, which is essentially the same as outer billiards, has a higher dimensional compactification which is a double lattice PET. In [**S6**] you can find details about how to see the arithmetic graph directly in terms of the dynamics of the double lattice PET.

All this specializes to the case of kites. In [**S3**] and [**S4**] I studied the kite case of the compactification in great detail, but only for the restriction of ψ^* to $X \cap \Sigma_0$. That is, I studied the compactification of an infinite IET. (I wrote [**S3**] first, and this gave me the idea for the generalization in [**S6**].)

In the case studied in [**S3**] and [**S4**], namely the compactification of a 1-dimensional infinite IET, the compactification is 2-dimensional when the parameter is a quadratic irrational, and 3-dimensional when the parameter is neither rational nor quadratic irrational. As we vary A, these various systems all line up and appear as slices of a single piecewise integral affine PET. I call this the *master PET*. The domain is an integer convex polytope – meaning that the vertices have integer coordinates – and the two partitions are made from smaller convex integer polytopes. I call this compactification result the *Master Picture Theorem* in [**S3**].

The large scale structure of the arithmetic graph is a consequence of the properties of this master PET. For instance, the reason that the bottom parts of the arithmetic graphes agree for Farey related rationals is that we compute the two fluxes by looking at two slices of the master PET. These slices are very close in comparison to the size of the interval, namely $2/q$, involved in the fundamental orbit. Again, this is just a brief summary of a huge amount of work.

4.4 Compactification and Renormalization

One further project I never completed was to consider the pinwheel map compactifications for all of X_0 and not just for $X_0 \cap \Sigma$. I did, however consider this in one case, for the Penrose kite. In [**S10**] I worked out the 3-dimensional compactification for the pinwheel map associated to the Penrose kite. This is a 3-dimensional double lattice PET, though I did not explicitly write it this way in [**S10**].

After extensive experimentation, I discovered that this 3-dimensional PET has a renormalization scheme much like the octagonal PET at the parameter $s = \sqrt{2}/2$. This allowed me to get an understanding of all the outer billiards orbits on the Penrose kite. Here are some sample results, all taken from [**S10**].

1. Every orbit is either periodic or unbounded in both directions.
2. The union of the unbounded orbits has Hausdorff dimension 1.
3. The set of horizontal lines containing unbounded orbis lies on a certain explicitly defined Cantor set of Hausdorff dimemsion $\log(3)/\log(\phi^3)$.
4. Suppose that $y = m + n\phi$ where ϕ is the golden ratio and $m, n \in \boldsymbol{Z}$. Then the line $R \times \{y\}$ has unbounded orbits if and only if m is odd and n is even.

This 150 page paper has many other results about outer billiards on the Penrose kite. Part of why I eventually abandoned polygonal outer billiards is that I could see that I would just end up writing forever and never do anything else. I didn't think that there would be a large enough audience for so much detail.

Let me close by showing one more picture. Figure 4.5 shows the arithmetic graph for the orbit of the point p shown in Figure 4.1. The graph itself is shown in black and then the grey path is a rescaled version. The rescaling factor is ϕ^3. The renormalization scheme I found implies that the corresponding arithmetic graphes are quasi-self similar. That is, the arithmetic graph for one of these unbounded orbits lies in a bounded tubular neighborhood of a dilated copy of the same arithmetic graph. This dilation structure also implies that the unbounded orbit itself is locally self-similar.

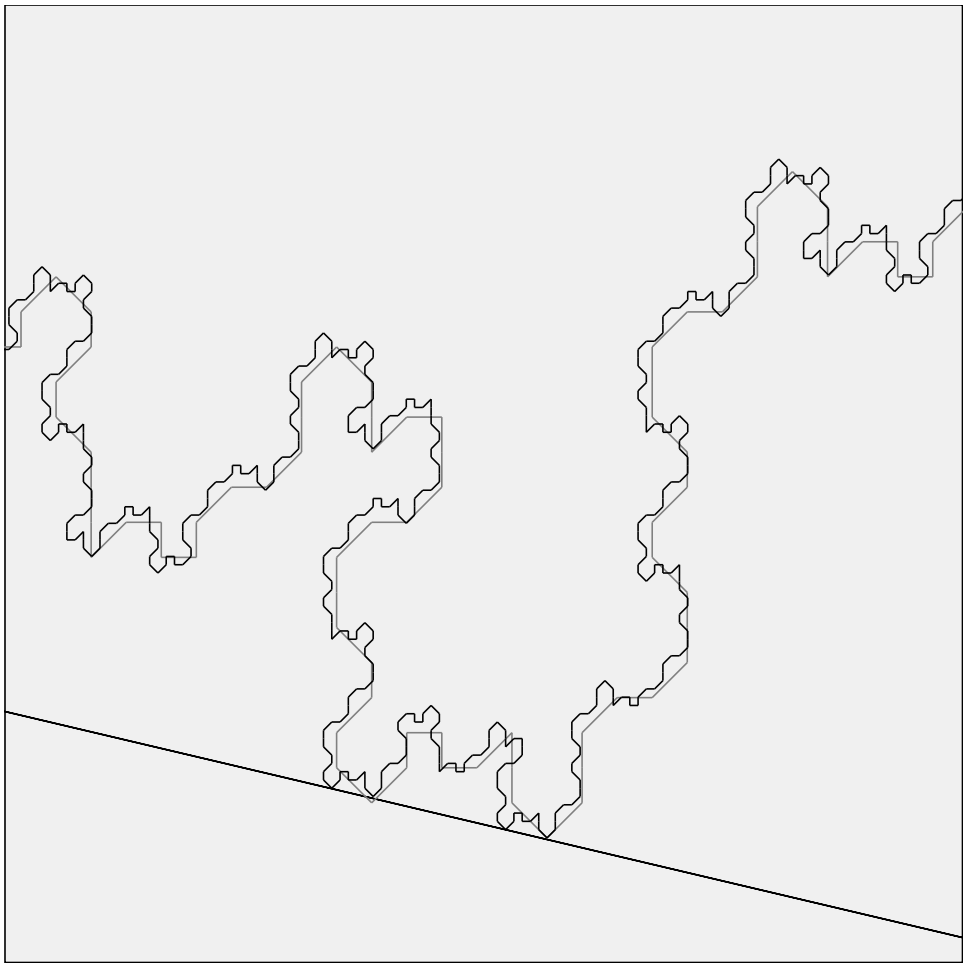

Figure 4.5: Quasi-self-similarity for the arithmetic graph.

One of my dreams is that the Master PET discussed in the previous section, or the larger 5-dimensional version which comes from looking at the whole pinwheel map and not just the restriction to Σ, has a renormalization scheme like the octagonal PETs. This would be a different way to explain some of the self-similar properties of the arithmetic graph. This kind of renormalization scheme would bring the study of outer billiards on kites more in line with the study of the octagonal PETs.

Going further, I think it would be great to have a renormalization scheme for double lattice PETs in general. There is some hope, because these objects are closely associated with the general linear groups. I imagine something along the lines of multi-dimensional continued fractions, a higher dimensional

version of Rauzy induction.

In summary, here is my idea about a 4-step plan to understand the question of unbounded orbits in polygonal outer billiards.

1. Replace outer billiards with the pinwheel map. See [**S8**].

2. Compactify the pinwheel map. The result, at least generically, is a double lattice PET. See [**S6**].

3. Look for general renormalization scheme for double lattice PETs, along the lines of Rauzy induction for interval exchange transformations. See [**S0**] for the Penrose kite.

4. Encode the dynamics of the double lattice PET using the arithmetic graph. See [**S6**] and also a deep exploration of this in [**S4**] for kites.

5 References

[**BC**], N. Bedaride, J. Cassaigne, *Outer billiard outside regular polygons*, Journal of the London Math Society (2011)

[**DF**] D. Dolyopyat and B. Fayad, *Unbounded orbits for semicircular outer billiards*, Annales Henri Poincaré, to appear.

[**FM**], G. Forni and C. Mateus, *Introduction to Teichmüller Theory and its applications to dynamics of interval exchange transformations, flows of surfaces, and billiards*, arXiv:1311.2758

[**G**] D. Genin, *Regular and Chaotic Dynamics of Outer Billiards*, Pennsylvania State University Ph.D. thesis, State College (2005).

[**Go**], A. Goetz, *Piecewise Isometries: An Emerging Area of Dynamical Systems* (2003)

[**GS**] E. Gutkin and N. Simanyi, *Dual polygonal billiard and necklace dynamics*, Comm. Math. Phys. **143**:431–450 (1991).

[**H**] G. Hughes, *Outer Billiards on Regular Polygons*, (2013) arXiv:1311.6763

[**Ke**] M. Keane, *Interval exchange transformations*, Math Z. **141** (1975) pp. 25-31

[**Ko**] Kolodziej, *The antibilliard outside a polygon*, Bull. Pol. Acad Sci. Math. **37**:163–168 (1994).

[**KRTZ**], A. Kanel-Belov, P Rukhovich, V. Timorin, V. Zgurskii, *Aperioric points for dual billiards* (2023) arXiv: 2311.09643

[**M1**] J. Moser, *Is the solar system stable?*, Math. Intelligencer **1**:65–71 (1978).

[**M2**] J. Moser, *Stable and random motions in dynamical systems, with special emphasis on celestial mechanics*, Ann. of Math. Stud. 77, Princeton University Press, Princeton, NJ (1973).

[**N**] B. H. Neumann, *Sharing ham and eggs*, Summary of a Manchester Mathematics Colloquium, 25 Jan 1959, published in Iota, the Manchester University Mathematics Students' Journal.

[**R**], G. Rauzy, *Èchanges d'intervalles et transformations induites*, Acta Arithmetica **34(4)** (1979) pp 315-328

[**S1**] R. E. Schwartz, *Survey Lecture on Billiards*, Proceedings of the International Congress of Mathematicians (2022)

[**S2**] R. E. Schwartz, *Continued Fractions and the Four Color Theorem*, Journal of Experimental Mathematics, Issue 1 (2023) to appear.

[**S2**] R. E. Schwartz, *Outer Billiards on Kites*, Annals of Mathematics Studies **171** (2009)

[**S4**] R. E. Schwartz, *The Plaid Model*, Annals of Mathematics Studies **198** (2018)

[**S5**] R. E. Schwartz, *Unbounded Orbits for Outer Billiards*, J. Mod. Dyn. **3**:371–424 (2007). [**T1**] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library 30, Amer. Math. Soc. (2005).

[**S6**] R. E. Schwartz, *Polytope Exchange Transformations and Quarter Turn Compositions*, Preprint (2011)

[**S7**] R. E. Schwartz, *The Octagonal PETs*, A. M. S. Research Monograph (2013).

[**S8**] R. E. Schwartz, *Outer Billiards and the Pinwheel Map*, J. Modern Dynamics **2** (2011) pp 255-283

[**S9**] R. E. Schwartz, *Outer Billiards, the Arithmetic Graph, and the Octagon*, preprint (2010)

[**S10**] R. E. Schwartz, *Outer Billiards and the Penrose Kite: Compactification and Renormalization*, J. Modern Dynamics **3**] (2012) pp 473-581

[**T1**] S. Tabachnikov, *A proof of Culter's theorem on the existence of periodic orbits in polygonal outer billiards*, Geometriae Dedicata **129**(1):83–87 (2007).

[**T2**] S. Tabachnikov, *Billiards*, Société Mathématique de France, "Panoramas et Syntheses" 1, 1995

[**VL**] F. Vivaldi and J. H. Lowenstein, —it Arithmetical properties of a family of irrational piecewise rotations, *Nonlinearity* **19**:1069–1097 (2007).

[**VS**] F. Vivaldi and A. Shaidenko, *Global stability of a class of discontinuous dual billiards*, Comm. Math. Phys. **110**:625–640 (1987).

[**Z**] A. Zorich, *Flat Surfaces*, Frontiers in Number Theory, Physics, and Geometry, Vol 1, ed. P. Cartier; B. Julia; P. Moussa; P. Vanhove, Spr.-Verlag (2006)