# Outer Billiards, the Arithmetic Graph, and the Octagon

Richard Evan Schwartz [*]

July 1, 2010

## 1   Introduction

B. H. Neumann [**N**] introduced outer billiards in the late 1950s and J. Moser [**M1**] popularized the system in the 1970s as a toy model for celestial mechanics. Outer billiards is a discrete self-map of $\boldsymbol{R}^2 - P$, where $P$ is a bounded convex planar set as in Figure 1.1 below. Given $p_1 \in \boldsymbol{R}^2 - P$, one defines $p_2$ so that the segment $\overline{p_1 p_2}$ is tangent to $P$ at its midpoint and $P$ lies to the right of the ray $\overrightarrow{p_1 p_2}$. The map $p_1 \to p_2$ is called *the outer billiards map*. The map is almost everywhere defined and invertible. See [**T1**] for a survey of outer billiards.
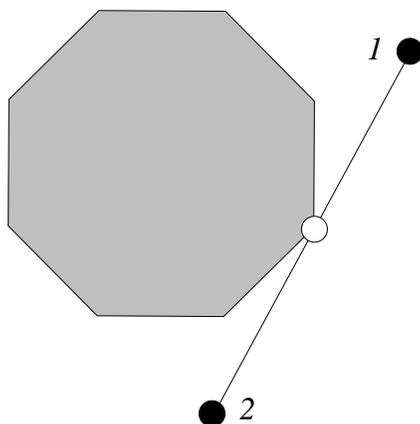


**Figure 1.1:** Outer billiards relative to $P$.

Usually, the main object of study is the orbit $\{p_n\}$ of the point $p_1$. These orbits are both complicated and interesting. When $P$ is a convex polygon, the orbits frequently have a fractal structure. Aside from a few cases, the structure of the orbits is not yet well understood.

It turns out that it is sometimes productive to study not the orbit itself but rather a certain *acceleration* of the orbit. Roughly speaking, the acceleration we have in mind amounts to considering the first return map to a certain infinite strip in the plane. We call this acceleration *the pinwheel dynamics*, because it is based on the pinwheel map that we studied in [**S3**]. We give a full account of the pinwheel map in §2.2 and recall some of the results of [**S3**] in §2.5. We point out, however, that this paper does not rely on any of the results in [**S3**]. The work here is self-contained.

The first purpose of this paper is to introduce an object called the *arithmetic graph*. The arithmetic graph is a polygonal path $\Gamma(P, p) \subset \boldsymbol{R}^n$ that one associates to the pinwheel dynamics of $p = p_1$. One can view this graph as a geometric incarnation of the symbolic coding of the pinwheel dynamics. The arithmetic graph is quite similar to the lattice paths studied by Vivaldi et. al. [**V**] in connection with interval exchange transformations. We will define the arithmetic graph in §2.3, right after defining the pinwheel map.

In case $P$ is a kite – i.e. a convex quadrilateral having a diagonal that is a line of symmetry – we have studied the arithmetic graph in great detail. In [**S1**] we analyzed the graph $\Gamma(P, p)$, where $P$ is the Penrose kite and $p$ is a specially chosen point. For this pair, we showed that $\Gamma(P, p)$ is *coarsely self-similar*, in the sense that a certain rescaled limit of $\Gamma(P, p)$ is a self-similar fractal curve. See §4.1 for a formal definition of what we mean by a *rescaled limit*.

We used the coarse self-similarity of the arithmetic graph in this case to conclude that the orbit of $p$ is unbounded relative to $P$. This provided the first example of an outer billiards system with unbounded orbits. In [**S2**] we studied the arithmetic graphs relative to kites in general, and showed that outer billiards has unbounded orbits when defined relative to any irrational kite. In §3.1 we will show some of the nicest pictures of arithmetic graphs associated to kites.

The irrational kites are the only known examples [1] of polygons for which outer billiards has unbounded orbits. Since the arithmetic graph turns out

---

[1]It is worth mentioning that Dolgopyat and Fayad show in [**DF**] that outer billiards also has unbounded orbits when defined relative to a semi-disk.

to be a decisive tool for establishing the unboundedness results for kites, we think that it will also be a useful tool for polygonal outer billiards more generally. Compare Theorem 2.3 (a result quoted from [**S3**]) and the discussion following it.

Even in the case when all the orbits are known to be bounded, as they are for so-called *quasi-rational* polygons – see [**VS**], [**K**], [**GS**] – we think that the arithmetic graph should shed light on the dynamics. (See the end of §2.5 for a definition of *quasi-rational*.) In particular, the arithmetic graph sheds light on the dynamics relative to the regular polygons. The regular polygons are nice examples of quasi-rational polygons.

The vertices of a regular polygon have coordinates that lie in a cyclotomic field. When $p$ lies in this same field, one can view certain projections of $\Gamma(p)$ as "Galois conjugates" of the orbits. In somewhat the same way that the full list of Galois conjugates of an algebraic number sheds light on the number, the arithmetic graph sheds light on the orbits associated to a polygon with algebraic vertices.

This brings us to the main purpose of our paper. We will concentrate on outer billiards for the regular octagon, a polygon that we scale so its vertices are the 8th roots of unity. The outer billiards dynamics for the regular octagon are well understood from certain points of view, but we will show that the study of the arithmetic graphs in the regular octagon case leads to a big surprise: There is a fractal loop $\Gamma$ in $\boldsymbol{R}^4$ which has the following three properties.

- One planar projection $\pi_3$ of $\Gamma$ is an embedded fractal loop reminiscent of the Koch snowflake. See Figure 1.2 below. We call this fractal *the snowflake*. We give a precise definition in §4.2.

- Another planar projection $\pi_2$ of $\Gamma$ is a fractal set reminiscent of the Sierpinski carpet. See Figure 1.3 below for an approximate picture. We call this fractal *the carpet*. We give a precise definition in §4.4.

- $\Gamma$ is the rescaled limit of the arithmetic graphs associated to any sequence of *odd* periodic orbits. We give a precise definition of *rescaled limit* in §4.1.

It turns out that all the periodic orbits intersect a certain strip in $3^k$ points for some integer $k$. See §7.4. We call the orbit *odd* if $k$ is odd. In Figure 1.2, the odd orbits are arithmetically closed orbits are lightly colored.

3

What we actually prove is slightly weaker than what we have just said. For technical reasons, we only consider odd orbits that lie outside the first layer of big grey octagons in Figure 1.3.
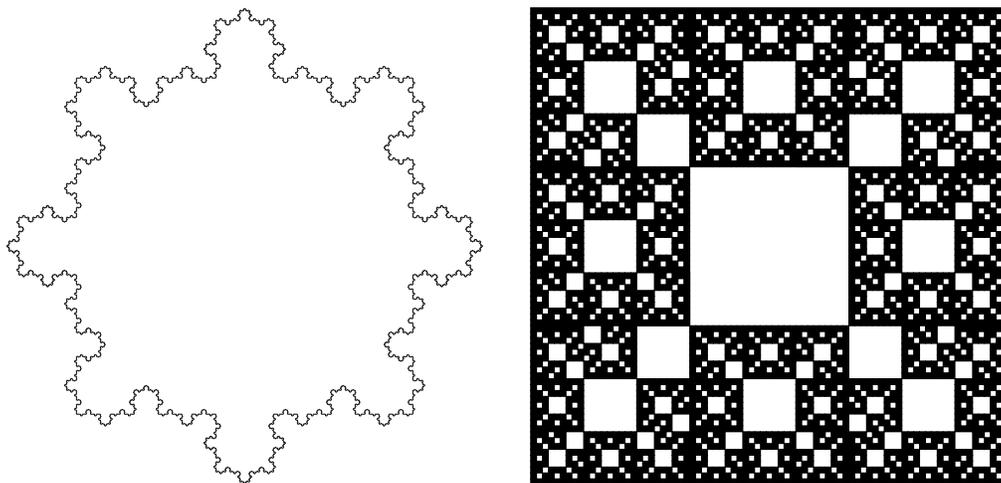


**Figure 1.2:** The snowflake and the carpet
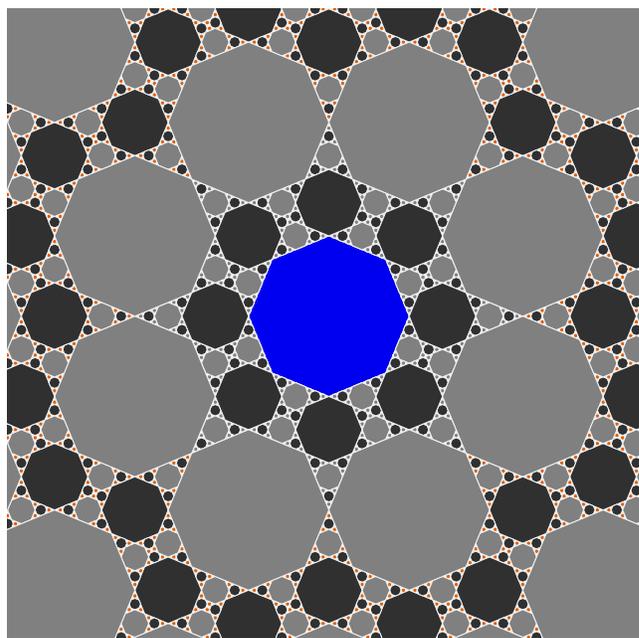


**Figure 1.3:** Periodic orbits associated to the octagon

By considering the correspondence between our two fractals, we produce a surjective continuous map from the snowflake to the carpet. This map is

"symmetry respecting" in a certain formal sense that we discuss in §4. See §4.6 for a description of the map from the snowflake to the carpet. What is surprising is that this canonical map between two seemingly unrelated fractals actually arises "naturally", in connection with outer billiards on the octagon.

For the sake of comparison, we now discuss the situation for other regular polygons.For $n = 3, 4, 6$, outer billiards on the regular $n$-gon is rather trivial and easy to describe. In all these cases, there is a familiar periodic tiling of the plane, such that the dynamics permutes the tiles. In the case $n = 4$ we get the usual square tiling. In the cases $n = 3, 6$ we get a tiling by hexagons and triangles.

For the case $n = 5$, Tabachnikov [**T2**] analyzes the situation in detail. The recent paper [**BC**] builds on the work in [**T2**] and describes the symbolic dynamics for the periodic orbits in great detail. We are interested in something different, namely the pinwheel dynamics, but the two are closely related. In the case of the regular pentagon, our construction produces a pair of embedded fractal curves, both akin to the Koch snowflake, that arise as geometric limits of arithmetic graphs associated to periodic orbits. See §3.2. For the sake of brevity, we do not present proofs of the statements we make for the regular pentagon. We are mainly interested in the octagon, and the analysis we give in the case of the octagon could be fairly easily redone for the pentagon.

The four cases $n = 5, 10, 8, 12$ are all quite similar to each other, and possible to understand completely. In these cases, outer billiards on the regular $n$-gon has an efficient renormalization scheme that allows one to give (at least in principle) a complete description of what is going on. The special properties of these values of $n$ is that the corresponding $n$th roots of unity lie in quadratic number fields. In the case $n = 10$, we get the same two snowflakes as in the case $n = 5$. We leave the case $n = 12$, which is mildly more complicated than any of the other quadratic cases, to the experimentally minded reader.

For any other positive integer $n$, outer billiards on the regular $n$-gon is not well understood at all. For the sake of comparison, we will draw some pictures of arithmetic graphs associated to the regular 7-gon. See §3.4. In this case, the pictures reveal an explosion of complexity in a geometrically striking way. Our 4 figures in §3.4 give only the faintest hint of the complexity.

Here is the plan of the paper.

- **Basic Definitions:** In §2 we will define the arithmetic graph, and give explicit formulas in the case of the regular octagon.

- **A Picture Gallery:** In §3 we draw the arithmetic graphs for many examples, including the regular octagon. After reading §3.3, the reader should have a good intuitive idea of what our main theorem will say, even though we defer the formal statement until §5.1.

- **Carpet and Snowflake:** In §4 we define the snowflake and carpet fractals precisely, and discuss the relationship between them.

- **Reduction to a Small Region:** We state our main result, Theorem 5.2, at the beginning of §5. In §5 we reduce Theorem 5.2 to the study of periodic points in a very small region $R_1$.

- **Toy Example of Renormalization:** In §6 we study a simple and well-known polygon exchange map that arises in connection with outer billiards on the regular octagon. The material in this chapter is not strictly necessary for our formal proof, but the system we study here is closely related to the one that comes from the pinwheel map.

- **Pinwheel Dynamics:** In §7, we describe the dynamics of the pinwheel map on the region $R_1$. Actually, we reduce an even smaller region, $R$, which is the top half of $R_1$. The dynamical system on $R$ is the main dynamical system of interest to us. Both the pinwheel map and the system from §6 have a period 3 renormalization that is the key to their analysis.

- **Substitution Scheme:** In §8 we study the arithmetic graphs produced by the dynamics on $R$. Using the renormalization scheme, we describe an efficient 2-part method for generating the projections of the arithmetic graph of interest to us. The first part of the method is combinatorial, and involves a substitution scheme for numerical sequences. The second part is geometrical, where we replace each number in the created string by a certain vector. We call this second half the *vector assignment*.

- **Fixed Point of Renormalization:** In §9 we modify the construction presented in §8, keeping the combinatorial part the same but changing the vector assignment. We see that the substitution scheme produces a kind of *renormalization operator* defined on the vector space of all possible vector assignments. This fixed point corresponds to the Perron-Frobenius eigenvector. By taking the fixed point of our operator as the new vector assignment, we produce a nicer family of curves that has the same rescaled limit.

- **Self-Similar Pattern Matching:** In §10 we check by a combination of direct calculation and induction that our improved curves from §9 have the snowflake and the carpet as their limit. The key idea is that the improved curves and the fractals from §4 are self-similar in compatible ways. The final analysis completes the proof of the Main Result.

The reader might be interested to know that we have made two java programs, *OctoMap 1* and *OctoMap 2*, which illustrate the mathematics in this paper. OctoMap 1 illustrates the snowflake and carpet fractals in great detail, and gives an interactive demonstration of the material in §4. OctoMap 2, a much more extensive program, starts with the dynamical system produced in §7 and illustrates all the key ideas that go into the proof of the Main Theorem. In particular, one can use OctoMap 2 to survey all the computations we describe in §7-10. One can find our applets, respectively, at the following address:

- http://www.math.brown.edu/~res/Java/OctoMap/Main.html

- http://www.math.brown.edu/~res/Java/OctoMap2/Main.html

We strongly encourage the reader to look at these applets while reading this paper. They relate to this paper the same way that a cooked meal relates to a recipe.

# 2 The Arithmetic Graph

## 2.1 Strip Maps

Let $\Sigma \subset \mathbf{R}^2$ be an infinite strip. Let $V$ be a vector whose tail lies on one edge of $\Sigma$ and whose head lies on the other. All this is shown in Figure 2.1.
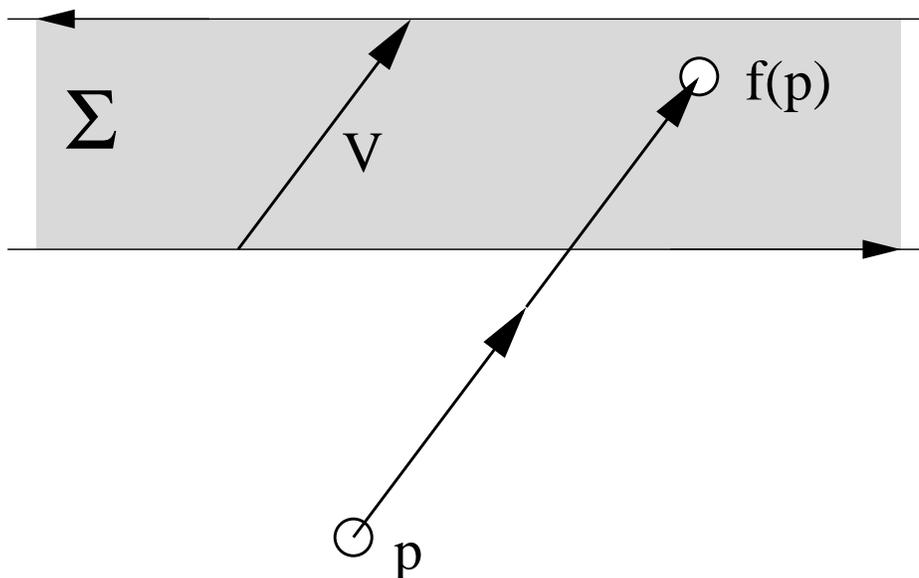


**Figure 2.1:** A strip map

To $(\Sigma, V)$ we associate a *strip map* $f : \mathbf{R}^2 \to \Sigma$, defined by $f(p) = p + nV$, where $n \in \mathbf{Z}$ is chosen so that $p + nV \in \Sigma$. In the example shown, we have $n = 2$.

**Remarks:**
(i) $f$ is not everywhere defined. It is not defined on a certain countable family of lines that are parallel to $\Sigma$. In particular $f$ is not defined on the boundary of $\Sigma$.
(ii) The pair $(\Sigma, -V)$ gives rise to the same map as the pair $(\Sigma, V)$.
(iii) In case $\Sigma$ is the horizontain strip bounded by the lines $y = 0$ and $y = 1$, and $V = (0, 1)$, the map map $f$ has the formula $f(x, y) = (x, [y])$, where $[y]$ is the fractional part of $y$. In this case, $f$ is not defined on the horizontal lines of integer height. Any other strip map is conjugate to this one by some affine transformation.

## 2.2 The Pinwheel Map

Let $P$ be a convex $n$-gon. To $P$ we associate $n$ special *pinwheel strip*. Each pinwheel strip $\Sigma$ is such that one component of $\partial\Sigma$ contains an edge $e$ of $P$, and vertices of $P$ farthest from this boundary component lie on the centerline of $\Sigma$. See Figure 2.2
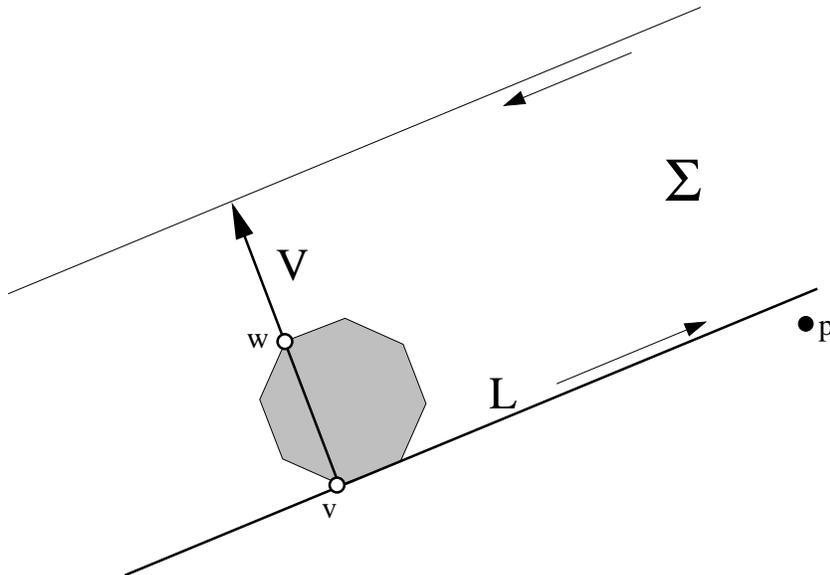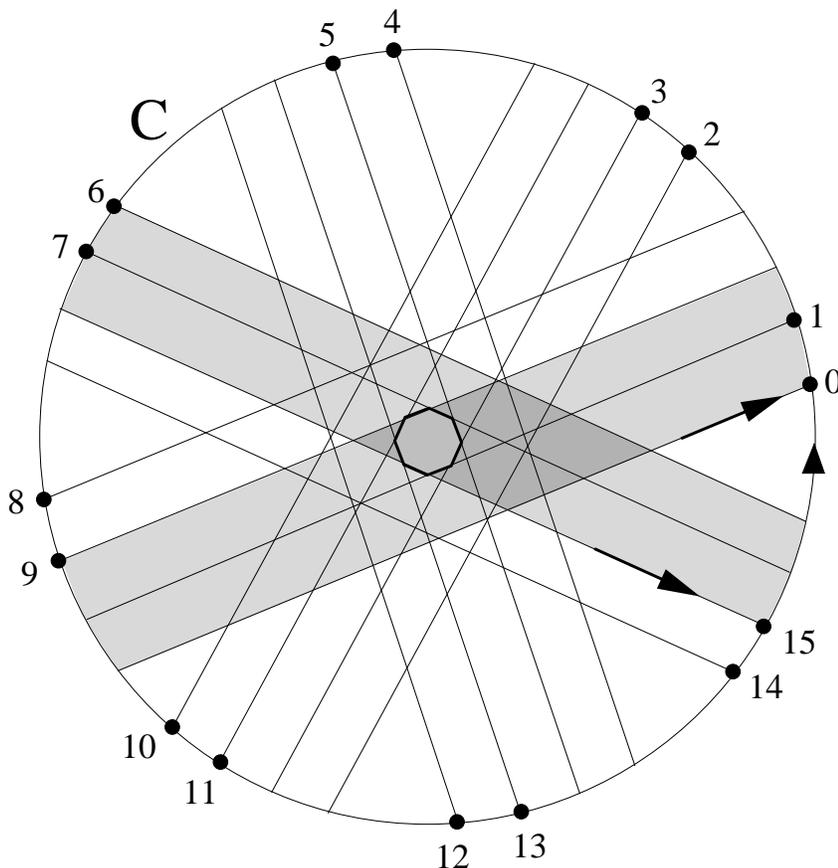


**Figure 2.2:** A pinwheel strip associated to a regular octagon

We orient the boundaries of a pinwheel strip $\Sigma$ so that a person walking along a boundary component would see $\Sigma$ on the left. Thus, the two boundary components are oriented in opposite directions. Say that a *pointed strip* is a strip, together with a choice of boundary component. We think of a pointed strip as pointing in the direction of the orientation on its preferred boundary component. To each $n$-gon we associate $2n$ pointed strips, 2 per pinwheel strip. We denote pointed strips by a pair $(\Sigma, L)$, where $L$ is a boundary component of the strip $\Sigma$.

To the pair $(\Sigma, L)$ we associate a vector $V$ as follows. We choose $\epsilon > 0$ and let $p$ be a point which is $\epsilon$ units from $L$, outside of $\Sigma$, and $1/\epsilon$ units away from the origin. Of the two possible locations for $p$ that the above conditions determine, we choose the one toward which $L$ points. Figure 2.2 shows what we mean. We then let $V = \phi^2(p) - p$. Here $\phi^2$ is the square of the outer billiards map. Our definition is independent of $\epsilon$, provided that $\epsilon$ is sufficiently small.

9

In general, $V = 2(w - v)$, where $v$ is the vertex bisecting the segment joining $p$ to $\phi(p)$ and $w$ is the vertex bisecting the segment joining $\phi(p)$ to $\phi^2(p)$. It is not hard to check that $\Sigma$ and $V$ are related as in §2.1.



**Figure 2.3:** The pointed strips associated to a regular octagon

Thus, we have associated to $P$ a total of $2n$ triples $(\Sigma_k, L_k, V_k)$. Here $(\Sigma_k, L_k)$ is a pointed pinwheel strip and $V_k$ is the associated vector. We cyclically order these $2n$ triples in the following way. We choose a large circle $C$ centered at the origin, and orient $C$ counterclockwise. Each pair $(\Sigma_k, L_k)$ determines a point $c_k \in C$, namely the intersection point of $L_k \cap P$ towards which $L_k$ points. We order our triples so that the points $c_1, ..., c_{2n}$ occur in counterclockwise order along $C$. Figure 2.3 shows these points, in case $P$ is the regular octagon. In Figure 2.3 we highlight the pointed strips $\Sigma_{15}$ and $\Sigma_0 = \Sigma_{16}$.

Let $f_k$ be the strip map associated to the pair $(\Sigma_k, V_k)$. We say that the *pinwheel map* is the composition

$$\Phi : f_{2n} \circ \ldots \circ f_0 : \Sigma_0 \to \Sigma_0 \qquad (1)$$

We say that $p \in \Sigma_0$ is *preferred* if, for all $x \in P$, the vector $p - x$ has positive dot product with vectors parallel to $L_0$. We call the subset of preferred points the *preferred half* of $\Sigma_0$ and denote it by $H_0$. We usually restrict $\Phi$ to $H_0$. Note that $\Phi$ need not carry $H_0$ to $H_0$. However, if $p \in H_0$ is sufficiently far from $\Sigma_0 - H_0$ then $\Phi(p) \in H_0$.

**Remark:** For our purposes here, the boundary of $H_0$ (that separates $H_0$ from $\Sigma_0 - H_0$) is not important. What is important is the *end* of $\Sigma_0$ that $H_0$ determines. Below, and just for the case of the regular octagon, we will replace $H_0$ by a similar set, namely $\Sigma_0^1$ below, that is invariant under $\Phi$.

Let $\phi$ denote the outer billiards map and let $\phi^2$ be the square of the outer billiards map. Let $O_+(\phi^2, p)$ denote the forwards $\phi^2$-orbit of $p$.

**Lemma 2.1** *The following is true for all points $p \in H_0$ that lie sufficiently far from $P$. Let $q$ be the first point in $O_+(\phi^2, p)$ that lies in $H_0$. Then $q = \Phi(p)$.*

**Proof:** For each point $p \in \mathbf{R}^2 - P$ for which $\phi^2$ is well-defined, there is a vector $V(p)$ such that $\phi^2(p) - p = V(p)$. The function $p \to V(p)$ is locally constant. Let $\Delta$ be the disk bounded by the large circle $C$, as in Figure 2.2. Let $W_k$ be the open acute wedge shaped region of $\mathbf{R}^2 - \Delta$ bounded by the lines $L_{k-1}$ and $L_k$. The vector $V_k$ has the property that $V(p) = V_k$ for at least some points $p \in W_k$. But one can check easily that $\phi^2$ is defined on all of $W_k$ provided that the circle $C$ is taken large enough. Hence $V(p) = V_k$ for all $p \in W_k$.

Starting at $p = p_0 \in W_1 \cap H_0$, the successive points in $O_+(\phi^2, p_0)$ have the form $p_0 + mV_1$ for $m = 1, 2, 3....$ This continues until we reach a point $p_1 = p_0 + m_1 V_1 \in \Sigma_1$. But then $p_1 = f_1(p_0)$. Starting at $p_1 \in W_2 \cap \Sigma_1$, the successive points in $O_+(\phi^2, p_1)$ have the form $p_1 + mV_2$ for $m = 1, 2, 3....$ This continues until we reach a point $p_2 = p_1 + m_2 V_2 \in \Sigma_2$. But then $p_2 = f_2(p_1)$. Continuing in this way, we get $p_n \in \Sigma_0$. But $p_n$ and $p_0$ lie in opposite components of $\Sigma - \Delta$. We have gone halfway around. Continuing the process, we finally arrive at $p_{2n} = q = \Phi(p_0) \in H_0$, as claimed. ♠

11

## 2.3　The Arithmetic Graph

Let $p = p_0 \in \Sigma_0$ be a point. As in the proof of Lemma 2.1, there are integers $m_1, ..., m_{2n}$ such that

$$p_k := f_k(p_{k-1}) = p_{k-1} + m_k V_k \in \Sigma_k. \qquad (2)$$

Let $v_1, ..., v_n$ be the vertices of $P$, ordered clockwise around $P$. Let $e_1, ..., e_n$ be the standard basis of $\boldsymbol{R}^n$. For each vector $V = V_k$ there are indices $i$ and $j$ such that $V_k = 2(v_i - v_j)$. Of course, $i$ and $j$ depend on $k$, but we are suppressing this from our notation. We define

$$\widetilde{V} = 2(e_i - e_j); \qquad k = 1, ..., 2n. \qquad (3)$$

We define

$$\widetilde{\Phi}(p) = \sum_{k=1}^{2n} m_k \widetilde{V}_k; \qquad \widetilde{\Phi}^{k+1}(p) = \widetilde{\Phi}(\widetilde{\Phi}^k(p)). \qquad (4)$$

We define $\Gamma(P, p)$ to be the polygonal path in $\boldsymbol{R}^n$ that starts at 0 and has consecutive vertices $\widetilde{\Phi}(p)$, $\widetilde{\Phi}^2(p)$, etc.

**Remarks:**
(i) It might seem at first that $\widetilde{\Phi}$ is somehow a "lift" of $\Phi$ to $\boldsymbol{R}^n$. This is only partially true. What is true is that $\widetilde{\Phi}^k$ is a map from $\Sigma_0$ to $\boldsymbol{R}^n$. This map is only defined on points where $\Phi^k$ is well defined.
(ii) By construction the arithmetic graph is a *lattice path* in that its vertices lie in $\boldsymbol{Z}^n$. As remarked in the introduction, one should compare the lattice paths of [**V**].
(iii) There is a constant $C'$, depending only on $P$, such that $|m_{k+n} - m_k| < C'$. The reason is that the successive points $p_1, ..., p_{2n}$ are the vertices of a convex polygon that is centrally symmetric to within a bounded error. Hence, the edges of $\Gamma(P, p)$ have length at most $C$, for some constant $C$ depending only on $P$. This property is important when we take rescaled limits of the graph.
(iv) There is an obvious projection $\pi : \boldsymbol{R}^n \to \boldsymbol{R}^2$ given by $\pi(e_k) = v_k$. By construction, $p + \pi(\Gamma(P, p))$, meaning the translation of $\pi(\Gamma(P, p))$ by $p$, is exactly the forwards $\Phi$ orbit of $p$. What makes $\Gamma(P, p)$ interesting is that sometimes other projections are much more revealing. We will illustrate this in the next chapter with many examples.

## 2.4 Formulas for the Octagon

Since we are concentrating on the regular octagon, it seems worthwhile giving explicit formulas in this case. Recall that the vertices of the regular octagon are $\omega^k$ for $k = 0, ..., 7$. We let $(ij)$ denote the line through $\omega^i$ and $\omega^k$, oriented from $i$ to $j$. We let $k(ij)$ denote the image of $(ij)$ after we apply reflection in $\omega^k$. Finally, we let $[ij]$ denote the vector that points from $\omega^i$ to $2\omega^j - \omega_i$. The vector $-[ij]$ is literally the negative of $[ij]$.
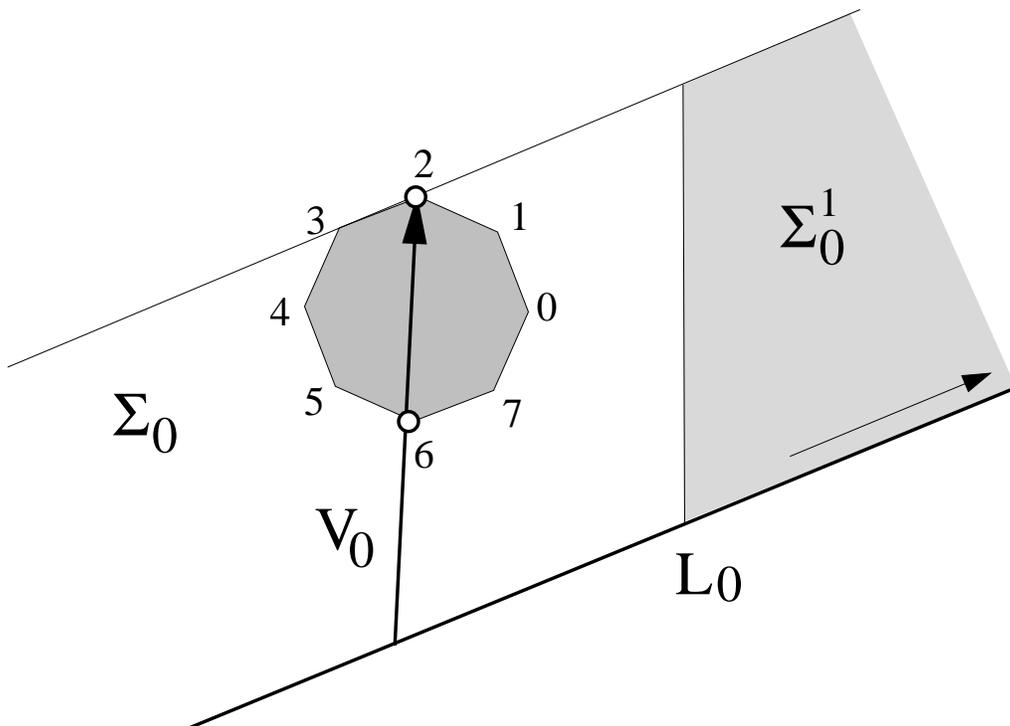


**Figure 2.4:** The triple $(\Sigma_0, L_0, V_0)$.

**Definition:** The shaded region in Figure 2.4 is the subset of $\Sigma_0$ to the right of the line $y = 2 + \sqrt{2}$. We let $\Sigma_0^1$ denote this set.

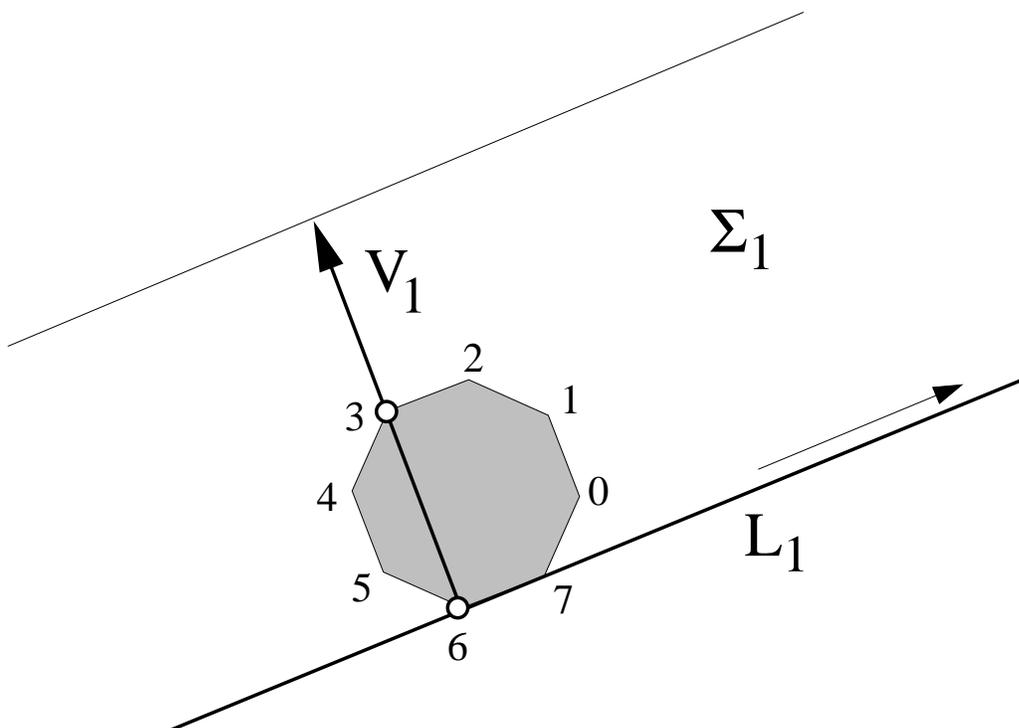To illustrate our notation by way of example, The triple $(\Sigma_0, L_0, V_0)$ is specified by

$$\mathbf{0}: \qquad 6(23) \qquad (23) \qquad -[26]. \tag{5}$$

We first list $L_0$, then the other component of $\Sigma_0$, and then $V_0$. See Figure 2.4.

13

The listing for $(\Sigma_2, L_2, V_2)$ is obtained from the one for $(\Sigma_0, L_0, V_0)$ simply by incrementing all the indices by 1. That is

$$\mathbf{2}: \qquad 7(34) \qquad (34) \qquad - [37].$$

And so on.



**Figure 2.5:** The triple $(\Sigma_1, L_1, V_1)$.

The triple $(\Sigma_1, L_1, V_1)$ is specified by

$$\mathbf{1}: \qquad (67) \qquad 3(67) \qquad [63] \qquad\qquad (6)$$

See Figure 3.4.

The listing for $(\Sigma_3, L_3, V_3)$ is obtained from the one for $(\Sigma_1, L_1, V_1)$ simply by incrementing all the indices by 1. That is

$$\mathbf{3}: \qquad (70) \qquad 4(70) \qquad [74]$$

And so on.

14

## 2.5 Discussion

Lemma 2.1 gives a precise relationship between the pinwheel map and the outer billiards map for points in $H_0$ that are far from $P$. The relationship between $\phi^2$ and $\Phi$ for points near $P$ is rather subtle, and it was the purpose of the paper [**S3**] to explain the relationship. Here we summarize some of the results from [**S3**]. These results are not needed for this paper, but they are worth knowing.

Let $O_+(\Phi, p)$ denote the forward $\Phi$-orbit of $p \in H_0$. Here is a consequence of our work in [**S3**]

**Theorem 2.2** *Suppose that all our constructions are made with respect to a convex polygon with no parallel sides. Then, the following holds for any sufficiently large disk $\Delta$ centered at the origin. Let $p \in \Sigma_0 - \Delta$. Then $O_+(\phi^2, p)$ returns to $H_0 - \Delta$ if and only if $O_+(\Phi, p)$ returns to $H_0 - \Delta$, and the two points of return are the same.*

There are two subtle points to Theorem 2.2. First, the forward orbits of $p$, in either case, might wind many times around $P$ before returning to $H_0 - \Delta_0$. Second, Lemma 2.1 only applies to points of $H_0$ that are sufficiently far from $P$. After our orbits exit $H_0 - \Delta$, then might wander quite close to $P$ and only "come back out" much later on. However, magically, they "come back out" in exactly the same way. One consequence of Theorem 2.2 is

**Theorem 2.3** *Suppose that all our constructions are made with respect to a convex polygon with no parallel sides. The pinwheel map $\Phi$ has unbounded orbits relative to $P$ if and only if outer billiards has unbounded orbits relative to $P$.*

We only proved these results for polygons without parallel sides because we wanted to avoid certain technical complications. We fully expect that the same results hold for all convex polygons. However, we have not yet worked out the details. For the present paper, which deals (rigorously) just with the regular octagon, we will prove something stronger than Theorem 2.2. Namely, in §5.3 we prove

**Lemma 2.4 (Invariance)** *For $k = 1, 2, 3...$ let $\Sigma_0^k \subset \Sigma_0$ be the subset consisting of points lying to the right of the line $x = k(1 + \sqrt{2})$ . Then $\Sigma_0^k$. $\Phi$-invariant for all $k \geq 1$*

The argument in Lemma 2.1 works directly for all orbits in $\Sigma_0^1$. Accordindly, for points in $\Sigma_0^1$, we can be sure that the pinwheel dynamics and the first return dynamics of the square outer billiards map coincide.

The Invariance Lemma implies that all forward orbits of both $\phi$ and $\Phi$ are bounded. By symmetry, the same goes for the backwards orbits. Hence, the Invariance Lemma implies Theorem 2.3 for the regular octagon. Compare §5.3.

The Invariance Lemma is a special case of a general result concerning quasi-rational polygons. The polygon $P$ is called *quasi-rational* if it may be scaled so that

$$\text{area}(\Sigma_k \cap \Sigma_{k+1}) \in \boldsymbol{Z} \cup \infty$$

for all $k = 1, ..., 2n$. Here indices are taken mod $2n$, as usual. In case no sides of $P$ are parallel, the above areas are all finite. For regular polygons, one can scale so that all the finite areas are 1. Hence, regular polygons are quasi-rational. Likewise, polygons with rational vertices are quasi-rational.

As we mentioned in the introduction, it is proved in [**VS**], [**K**], and [**GS**] that all outer billiards orbits are bounded for a quasi-rational polygon. In [**S3**] we give a self-contained proof of this result, in case $P$ has no parallel sides.

# 3 Examples of Arithmetic Graphs

## 3.1 Rational Kites

As in [**S2**] we let $K(A)$ be the kite with vertices

$$(-1, 0); \qquad (0, 1); \qquad (0, -1); \qquad (A, 0); \qquad A \in (0, 1). \qquad (7)$$

The case $A = 1$ corresponds to the square, a trivial case we ignore. $K(A)$ is called (ir)rational iff $A$ is (ir)rational.

When $A = p/q$, we call the point $(1/q, 1)$ the *fundamental point*, for reasons we explain in great detail in [**S2**] and briefly as follows. The orbit of any point $(x, 1)$ has the same combinatorics as the point $(1/q, 1)$ for any $0 < x < 2/q$. For this reason, the point $(1/q, 1)$ is a representative of the dynamics of points "arbitrarily close" to the top vertex of $K(p/q)$ and on the same horizontal line as the vertex.

We will consider the arithmetic graphs

$$\Gamma_{p/q} = \Gamma(K(p/q), (1/q, 1)) \qquad (8)$$

for various choices of $p/q$. We will draw a certain projection of this graph into the plane. We will not specify the projection we use explicitly because it is tedious to do so. The fact is that $\Gamma_{p/q}$ lies in a thin neighborhood of a 2-plane in $\boldsymbol{R}^4$ and a random projection will produce pictures very similar to ours. Indeed, any projection looks about the same, up to an affine transformation of $\boldsymbol{R}^2$.

Consider the rational sequence $\{a_n\}$ which starts $0, 1$ and obeys the rule $a_n = 4a_{n-1} + a_{n-2}$. This sequence starts out
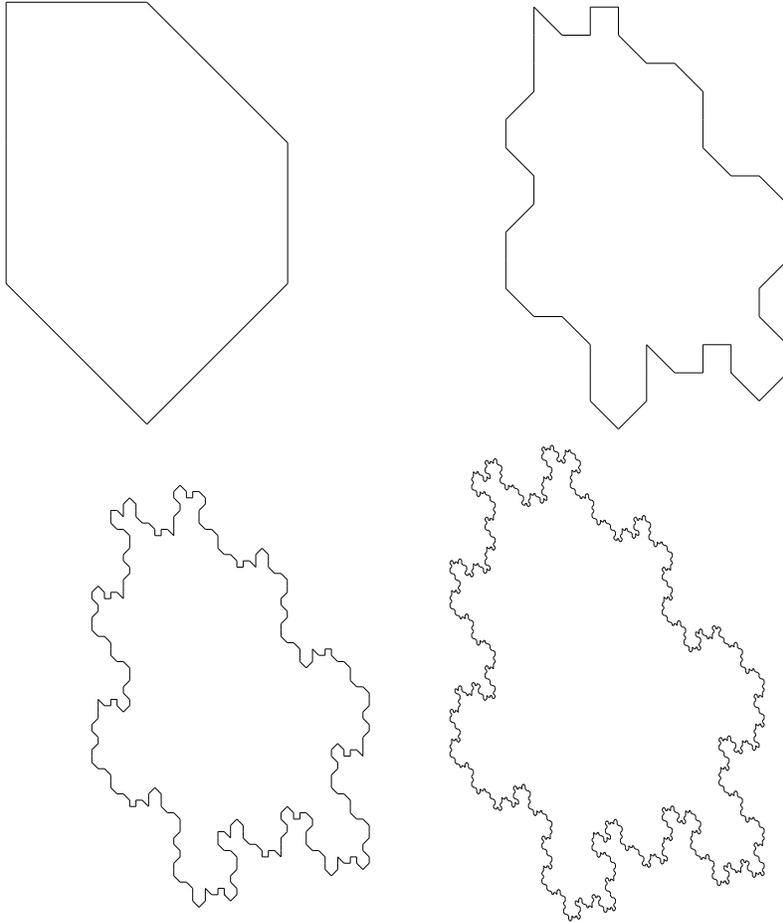
$$0, 1, 4, 17, 72, 305, \ldots$$

We have

$$\lim_{n \to \infty} a_{n-1}/a_n = \sqrt{5} - 2 = \phi^{-3}.$$

Here $\phi$ is the golden ratio. The kite $K(\phi^{-3})$ is affinely equivalent to the Penrose kite. The first four quotients are

$$1/4 \qquad 4/17 \qquad 17/72 \qquad 72/305 \ldots$$

Figure 3.1 shows $\Gamma_A$ for the first four of these quotients.

**Figure 3.1:** $\Gamma_A$ for $A = 1/4, 4/17, 17/72, 72/305$.

We want to emphasize that our projection is such that the vertices of these polygons all lie in $\mathbf{Z}^2$. Were the polygons drawn to scale, each one would be about $\phi^3 \approx 4.23$ times as large as the predecessor. However, when we rescale them so that they are about the same size, a fractal structure emerges. Later in the paper, we will formalize what we mean to taking the *rescaled limit* of a sequence of arithmetic graphs, but we hope that the above informal discussion makes the general idea clear.

Another nice sequence of pictures is given by the quotients of the rational sequence $\{a_n\}$ where $a_0 = 1$ and $a_1 = 2$ and $a_n = 2a_{n-1} + a_{n-2}$. In this case $\lim a_{n-1}/a_n = \sqrt{2} - 1$. One of the quotients is $169/408$. Figure 3.2 shows the picture of the corresponding arithmetic graph. One can see the fractal

18

structure emerging just from this single polygon.
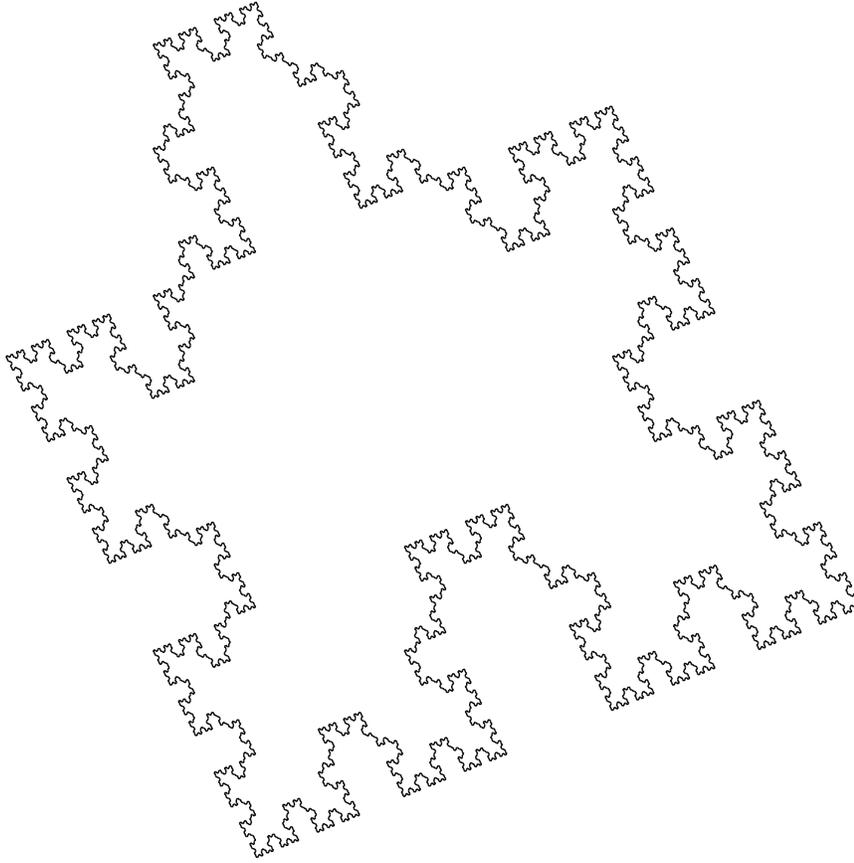


**Figure 3.2:** $\Gamma_A$ for $A = 169/408$.

There is one interesting feature of $K_{p/q}$ that one sees almost immediately. If $pq$ is even, then $K_{p/q}$ is a closed embedded loop. If $pq$ is odd, then $K_{p/q}$ is an open embedded polygonal curve. The difference derives from the fact that the orbit of $(1/q, 1)$ is stable when $pq$ is even and unstable when $pq$ is odd. By this we mean that a small change in the kite parameter destroys the orbit of $(1/q, 1)$ when $pq$ is odd, but not when $pq$ is even. We prove the above statements in [**S2**].

My program Billiard King [2] allows the reader to draw these pictures for any (smallish) rational parameter. Our monograph [**S2**] discusses these graphs in great detail. In fact, one could consider [**S2**] as an exploration of the arithmetic graphs associated to kites.

---

[2]You can download this program from my website http://www.math.brown.edu/~res

## 3.2 The Regular Pentagon

Let $P$ be the regular pentagon. We scale $P$ so that its vertices have the form $\omega^k$ for $k = 0, 1, 2, 3, 4$. Here $\omega = \exp(2\pi i/5)$ is the usual primitive 5th root of unity. In this case, we have maps $\pi_k : \boldsymbol{R}^5 \to \boldsymbol{C}$ given by

$$\pi_k(X) = \sum_{j=0}^{4} x_j \omega^{kj}. \tag{9}$$

Here $X = (x_0, x_1, x_2, x_3, x_4)$. The map $\pi_1$ is just the projection mentioned above. Note that $\pi_k$ and $\pi_{5-k}$ agree up to complex conjugation. Thus, the only remaining interesting projection is $\pi_2$.

Relative to any convex polygon, any periodic point $x$ lies in a convex polygon $P_x$ consisting of points that have exactly the same dynamics. The outer billiards map moves the tile $P_x$ around isometrically. The pictures we show of outer billiards will draw the orbits of these tiles, rather than the orbits of individual points, because the pictures are more revealing. Figure 3.3 shows the picture for the regular pentagon.
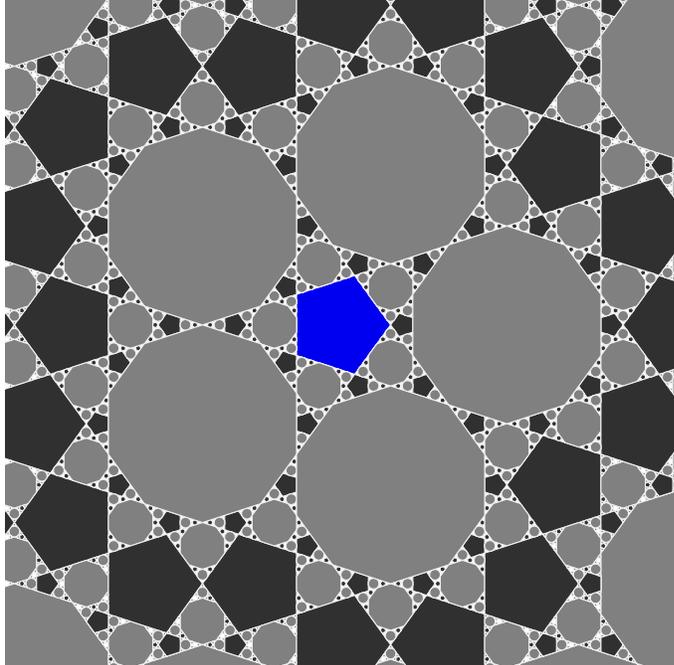


**Figure 3.3:** Some of the periodic orbits for the regular pentagon

In the case of the regular pentagon, there are two kinds of orbits, those lying in regular pentagon tiles and those lying in regular 10-gon tiles. This is worked out in [**T2**], and also discussed in [**BC**]. The smaller the tile, the longer the periodic orbit.

The left half of Figure 3.4 shows a plot of $\pi_2(\Gamma(P, p))$, where $p$ is a point of $\Sigma_0$ that lies inside a pentagonal tile. We choose the period to be fairly long, so that the actual size of the graph is quite large. One can see the emerging fractal structure in the rescaled picture. The right half of Figure 3.4 shows the corresponding picture for a periodic point that lies inside a 10-gon tile. The reader can probably see that these tiles somehow fit together "hand-in-glove".
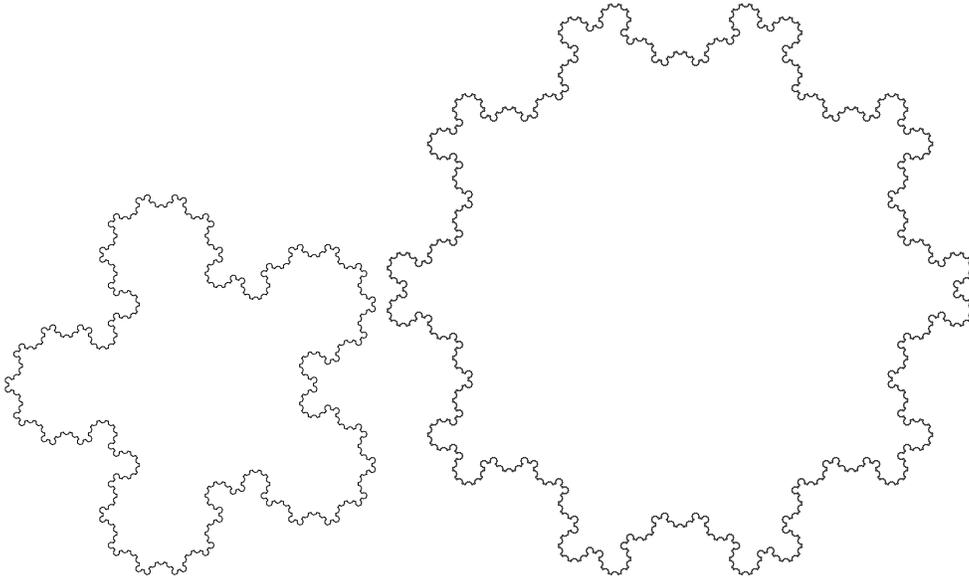


**Figure 3.4:** Some of the periodic orbits for the regular pentagon

Any choice of long periodic orbit will yield a picture that is essentially identical to one of the two above. In the case of the regular pentagon, the arithmetic graphs lie within a thin tubular neighborhood of a 2-plane in $\boldsymbol{R}^5$. Any other projection of the graph would look like an affine image of one of the two pictures shown above. This is why we say that, in the regular pentagon case, there are two fractal curves that somehow control the structure of the pinwheel dynamics.

As we remarked in the introduction, we do not offer proofs of these results, because we will give complete proofs of the analogous results for the regular octagon.

## 3.3 The Regular Octagon

Let $P$ be the regular octagon. We scale $P$ as in §2.4, and we define the projection maps as in §3.2. In the case, there are two interesting projections, $\pi_2$ and $\pi_3$. (It turns out that $\pi_4$ always maps the graph into a line segment.)

As we discuss in the introduction, in connection with Figure 1.3, we call a periodic orbit *odd* if it intersects $\Sigma_0^1$, the half-strip from Figure 2.4., in $3^k$ points, for $k$ odd. It turnso out that "half" the periodic orbits that start in $\Sigma_0^1$ have this property. The odd periodic orbits are the lightly colored ones in Figure 3.5 that happen to intersect $\Sigma_0^1$. This includes all the lightly colored octagons that lie outside the first large ring of 8 octagons.
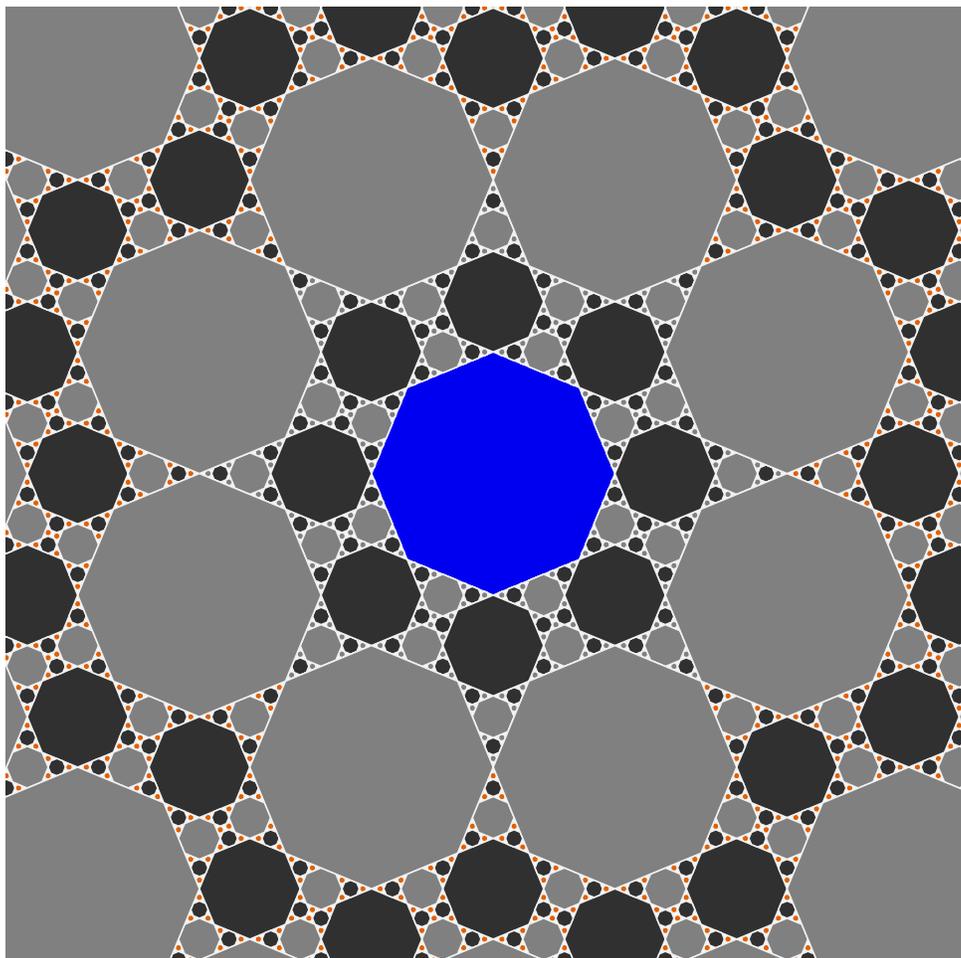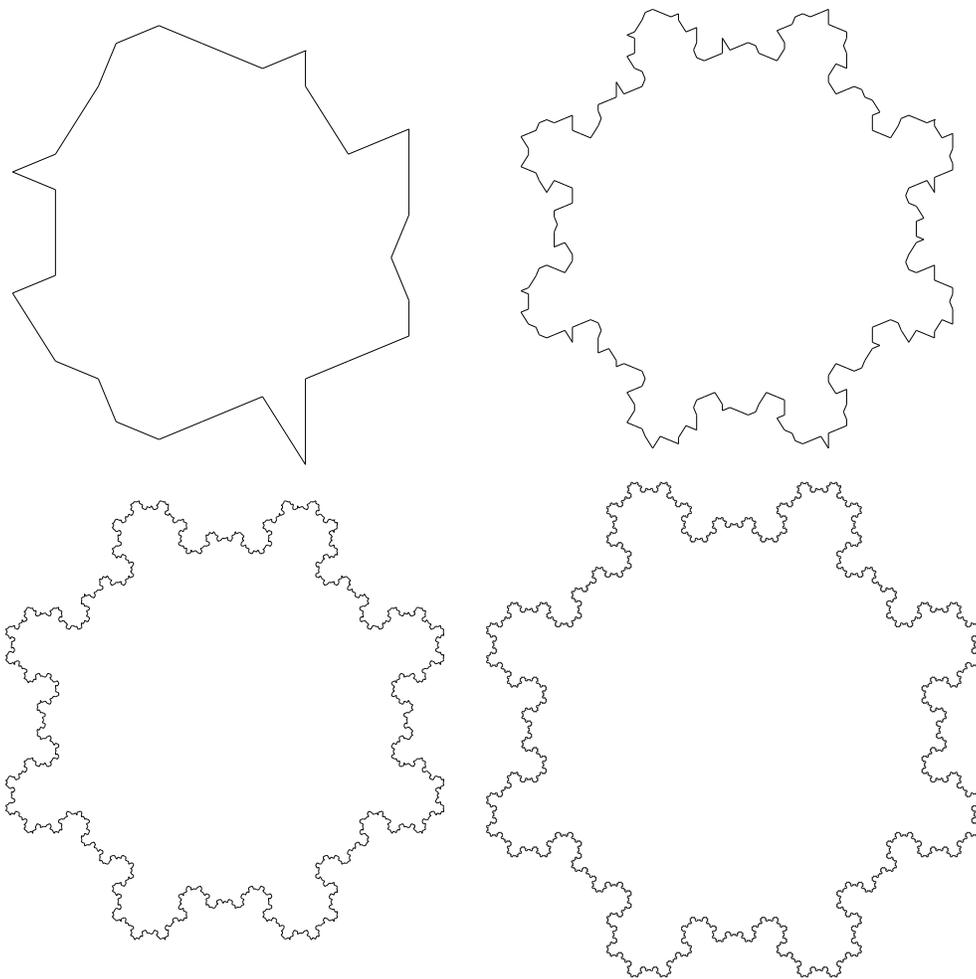


**Figure 3.5:** The tiling associated to the octagon

Figure 3.6 shows some $\pi_3$ projections of the arithmetic graphs associated to odd orbits. For $\pi_3$, the projections corresponding to the even orbits look similar. Note that Figure 1.2, which was drawn using the subdivision method discussed in §4.2, emerges as we rescale the pictures.



**Figure 3.7:** The snowflake emerges.

Now we are going to show the $\pi_2$ projections of the same orbits.
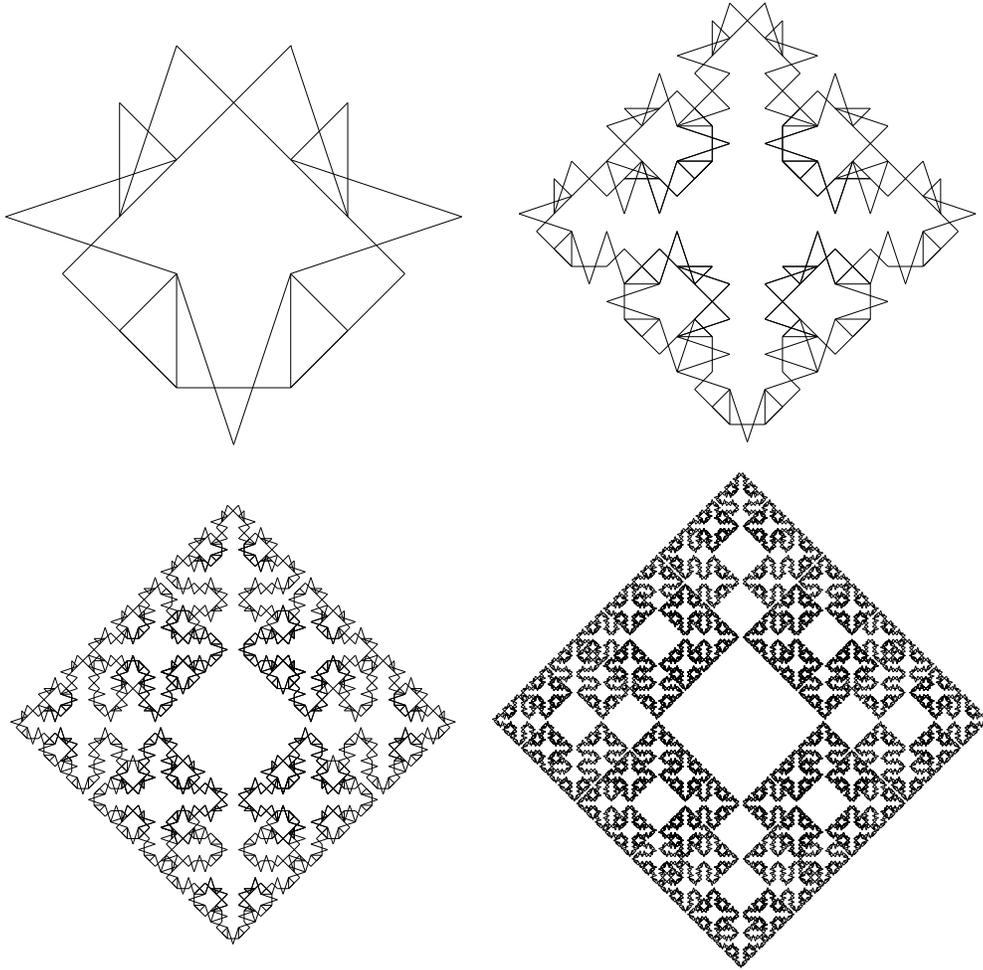
**Figure 3.7:** The carpet emerges in the odd case.

For the even orbits, the $\pi_2$ projection is an open curve that looks fairly similar. We refer the reader to the program OctoMap 2, where one can see much better pictures, in both the odd and even cases.

## 3.4   The Regular Heptagon

We can make the same constructions for the regular heptagon as we made for the regular pentagon and octagon. The analogue of Figure 3.3 exists, but is not yet understood. The two interesting projections in this case are $\pi_2$ and $\pi_3$. Unlike the two cases we considered above, there is an explosion

of distinct pictures in this case. Sometimes the $\pi_2$ projection looks nicer and sometimes the $\pi_3$ projection looks nicer. Sometimes both projections look incomprehensibly complex. Here are two simple examples, both $\pi_2$ projections.
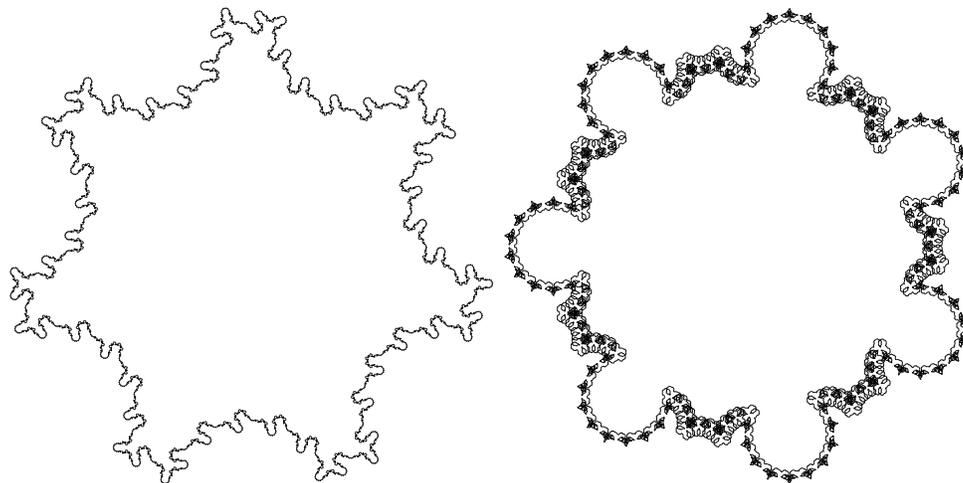


**Figure 3.10:** Some examples associated to the regular 7-gon.

These examples do not even begin to capture the vast array of pictures one sees for the regular 7-gon. We encourage the experimentally-minded reader to explore the situation for themselves.

# 4 The Snowflake and the Carpet

## 4.1 The Hausdorff Topology

Let $K_1, K_2 \subset \boldsymbol{R}^2$ be two compact subsets. We define $\delta(K_1, K_2)$ to be the smallest $\epsilon$ such that $K_1$ is contained in the $\epsilon$-tubular neighborhood of $K_2$ and *vice versa*. The function $\delta$ is known as the *Hausdorff distance* between $K_1$ and $K_2$. Given any compact subset $\Omega \subset \boldsymbol{R}^2$ we let $M(\Omega)$ denote the set of compact subsets of $\Omega$. Equipped with the function $\delta$, the space $M(\Omega)$ is itself a compact metric space.

When we have a sequence of uniformly bounded compact subsets $\{K_n\}$ and we say that it converges, we mean that it converges in $M(\Omega)$ for some large compact $\Omega$ that contains all the individual sets. The set $\Omega$ just serves as a kind of container, so that we can speak about the convergence as taking place within a compact metric space. The notion of convergence we get is independent of the choice of $\Omega$. Indeed, we could say more simply that $K_n \to K$ iff $d(K_n, K) \to 0$. We call $K$ the *Hausdorff limit* of $\{K_n\}$.

One case of interest to is is when we have a sequence $\{P_n\}$ of polygons whose diameter tends to $\infty$. If we can a compact subset $K$ and a sequence of similiarities $\{S_n\}$ such that $S_n(P_n) \to K$ in the Hausdorff topology, then we call $K$ a *rescaled limit* of $\{P_n\}$. In this case, the contraction factor of $S_n$ necessarily tends to 0.
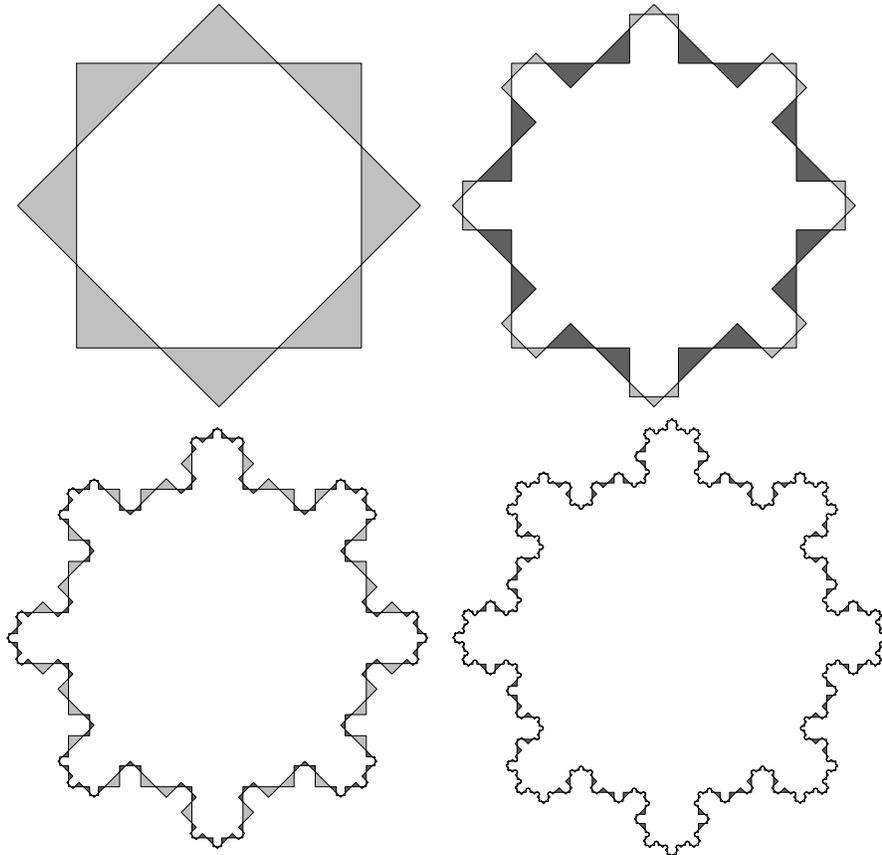
## 4.2 The Snowflake

We call the snowflake $I_3$. Figure 4.1 shows the subdivision of a right-angled isosceles triangle into 5 smaller ones. If the long side of the triangle has length 1, then the square on the right hand side of the figure has side length $\sqrt{2} - 1$. The three small triangles in the middle have the same size.
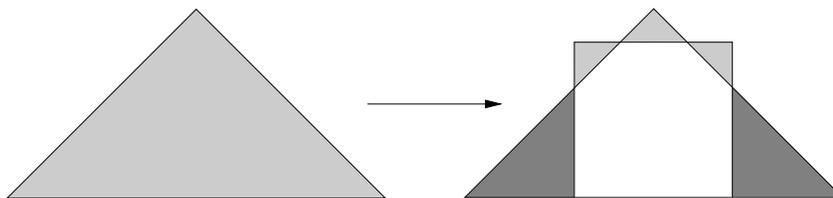


**Figure 4.1:** Subdivision rule

To produce the snowflake, we start with the union of 8 isosceles triangles shown at the top right in Figure 4.2. Then we apply the subdivision rule

iteratively. The Hausdorff limit is our snowflake. We denote this limit by $I_3$.



**Figure 4.2:** Producing the snowflake by iteration.

One nice feature of our construction is that $I_3$ contains all the vertices of the triangles at every stage of the construction, and the union of these vertices is dense in $I_3$. Moreover, all these vertices lie in $\mathbf{Z}[\omega]$, where $\omega$ is the usual primitive 8th root of unity. The triangles in our pictures are 2 colored in a natural way, according as to whether they point inward or outward. Assuming that we label a triange 0 if it points inward and 1 if it points outward, our subdivision rule changes the colors according to the following rule:

$$0 \to 10001; \qquad 1 \to 01110. \tag{10}$$

We will see the significance of this structure below when we examine the carpet.

## 4.3 A Hidden Symmetry of the Snowflake

For any right isosceles triangle $T$, let $L(T)$ denote the curve one obtains by iterating the basic snowflake substitution rule and taking a limit. Assuming that $T$ has a distinguished side $\sigma$, let $L(T, \sigma)$ denote the portion of $L(T)$ whose endpoints are the endpoints of $\sigma$.
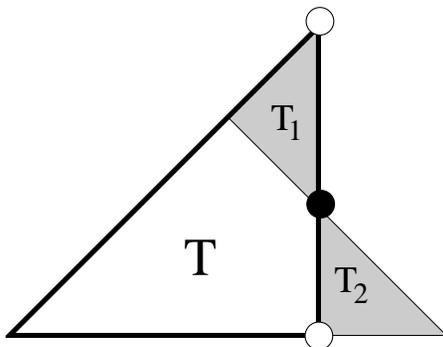


**Figure 4.3:** Some special triangles

Let $T$, $T_1$, and $T_2$ be the three right isosceles triangles shown in Figure 4.3. The two triangles $T_1$ and $T_2$ have the same size as each other, and are $\sqrt{2} - 1$ times as large as $T$. Let $A = L(T, \sigma)$, where $\sigma$ is the edge of $T$ bounded by the two white vertices. Let $A_1 = L(T_1, \sigma_1)$, where $\sigma_1$ is edge of $T_1$ bounded by the white and black vertex. Define $A_2 = L(T_2, \sigma_2)$ similarly.

**Lemma 4.1** $A = A_1 \cup A_2$.

**Proof:** Let $d$ be the Hausdorff distance between $A$ qnd $A_1 \cup A_2$. The substitution of $T$ consists of 5 smaller triangles, $T'_1, ..., T'_5$, with $T'_1 = T_1$ and $T'_2 \cup T'_3$ related to $T_2$ just as $T_1 \cup T_2$ is related to $T$. There are edges $\sigma'_2$ and $\sigma'_3$ of $T'_2$ and $T'_3$ such that $\sigma_2 = \sigma'_2 \cup \sigma'_3$. The remaining triangles $T'_4$ and $T'_5$ lie to the left of the bottom white vertex of $T$.

Since $T_1 = T'_1$, we have $A_1 \subset A$. At the same time, the pair of arcs $(A, A_1 \cup A_2)$ is similar to the pair of arcs $(A_2, A'_2 \cup A'_3)$, where $A'_2 = L(T'_2, \sigma'_2)$ and $A'_3 = L(T'_3, \sigma'_3)$. From this we see that the Hausdorff distance between $A$ and $A_1 \cup A_2$ is the same as the Hausdorff distance between $A_2$ and $A'_2 \cup A'_3$. But, by scaling, the latter distance is $(\sqrt{2} - 1)d$. So, we have the equation

$$d = (\sqrt{2} - 1)d,$$

which of course forces $d = 0$. ♠

## 4.4 The Carpet

We call the carpet $I_2$. The carpet is produced by substitution rule that is combinatorially identical to the one that produces the snowflake. This time, we have two shapes, parallelograms and trapezoids. The initial shapes have their vertices in $\mathbf{Z}^2$. Our figures below are accurate.

Figure 4.4 illustrates how the parallelogram $P$ is replaced by the sequence

$$S_P(P) = T_1 \cup P_2 \cup P_3 \cup P_4 \cup T_5$$
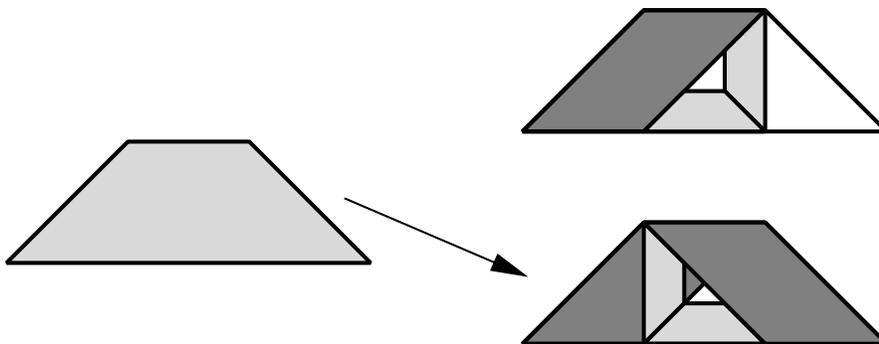
of 5 smaller parallelograms and trapezoids.
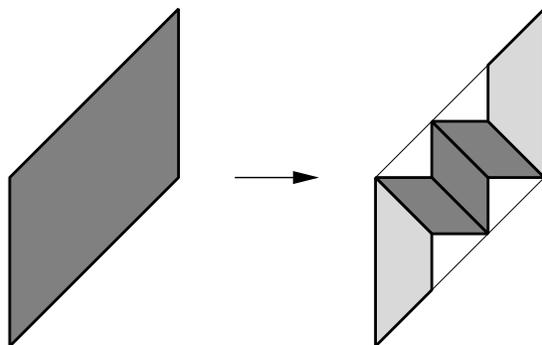


**Figure 4.4:** Substitution rule for the parallelogram

Figure 4.5 shows how the trapezoid $T$ is replaced by the sequence

$$S_T(T) = P_1 \cup T_2 \cup T_3 \cup T_4 \cup P_5$$

of smaller parallelograms and trapezoids. We say *sequence* here rather than *union* because the pieces in the sequence $S_T(T)$ overlap each other. To make the substitution clear, we first draw $P_1 \cup T_2 \cup T_3$. Then we add $P_4$ and $P_5$. The partition is invariant under reflection in the vertical line of symmetry.



**Figure 4.5:** Substitution rule for the trapezoid

Let $X$ be a cyclically ordered finite list of parallelograms and trapezoids, similar to the ones above. We let $|X|$ be the union of shapes in $X$. We let $S_P(X)$ denote the result of substituting $S_P(P)$ for all parallelograms $P \in X$. Likewise we define $S_T(X)$. Note that $S_P(X)$ and $S_T(X)$ both have natural cyclic orders, inherited from the ordering on $X$. Our carpet is

$$I_2 = \lim_{n \to \infty} |I_2(n)| \qquad I_2(n) = S_P(S_T(I_2(0))). \tag{11}$$

Here $I_2(0)$ is a union of 8 copies of $T$ arranged in a square pattern that winds around twice. (We use 8 rather than 4 so that the seeds for $I_2$ and $I_3$ have the same cardinality.) Figure 4.6 shows the sets $I_2(n)$ produced by iterated substitution, for $n = 0, 1, 2, 3$. Since the pieces in overlap, there is more than one partition we could draw. We have chosen the partition in which every trapezoid is drawn above every parallelogram.
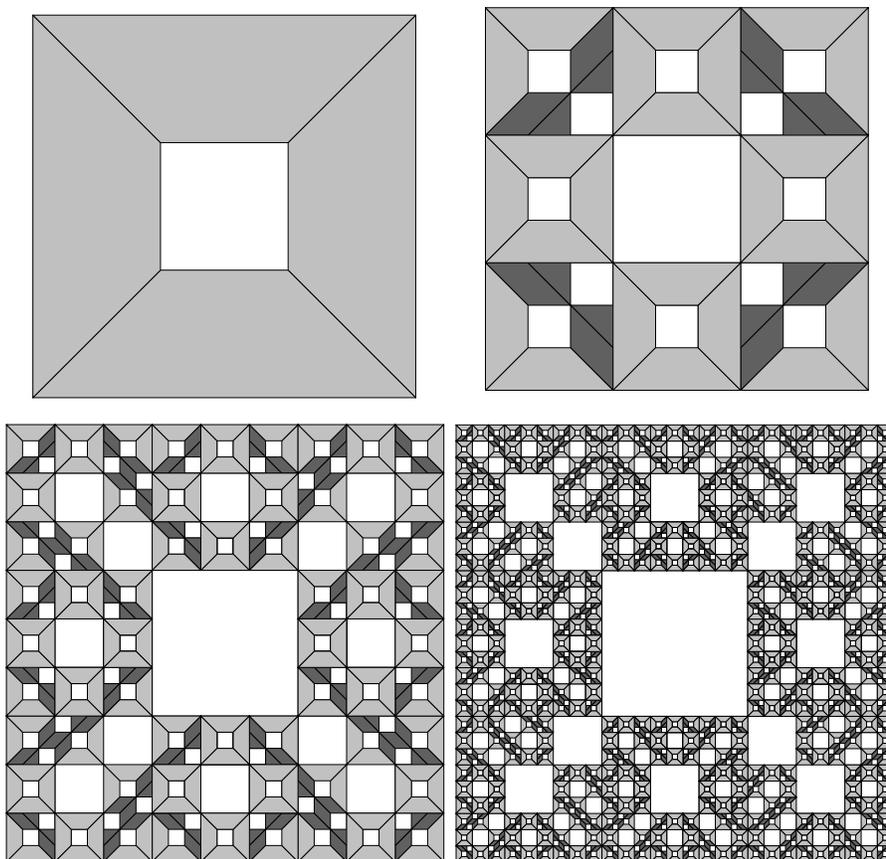


**Figure 4.6:** Iterating the substitution rule

## 4.5 Another View of the Carpet

Figure 4.7 shows another substitution rule that generates $I_2$. In Figure 4.7, we are showing how to replace a square by a subset of a square. Here we think of the dark triangles and quadrilaterals as markings that tell us how to orient the pieces when we iterate the basic rule. We obtained this rule by looking at Figure 4.6. The fractal obtained from this alternate rule is a subset of the Sierpinski carpet. It is obtained from the Sierpinski carpet by systematically deleting certain squares.
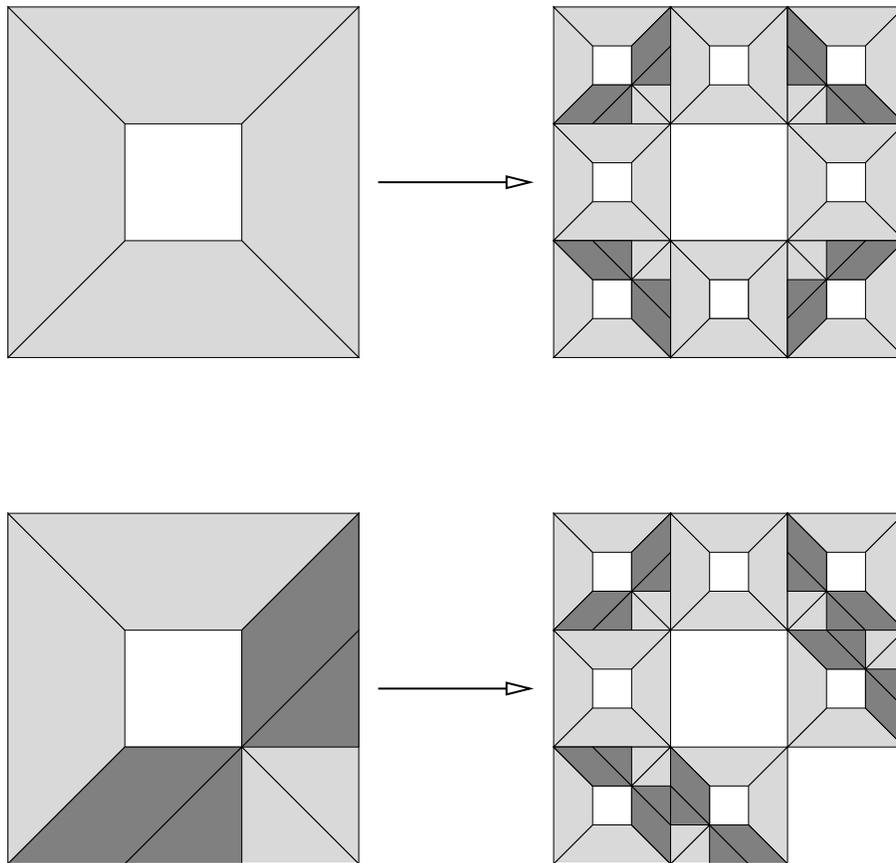


**Figure 4.7:** An alternate substitution rule

**Lemma 4.2** $I_2$ *is the fractal produced by the alternate substitution rule.*

**Proof:** We will start with the original rule and keep modifying it until we get to the alternate rule. Along the way, we show that the modifications don't change the limiting fractal.

31

For starters, we divide each parallelogram in half. In this way, we interpret the original substitution rule as a being generated by a rule for trapezoids and a rule for isosceles triangles. The left hand side of Figure 4.8 shows the new $S_P(P)$, The right hand side of Figure 4.8 shows the new $S_T(T)$.
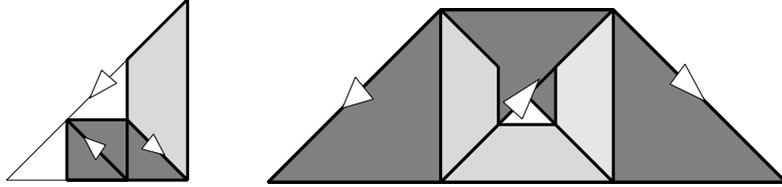


**Figure 4.8:** The first modification

We have redrawn the second halves of Figures 4.4 and 4.5. In the case of Figure 4.5, we have drawn all the trapezoids on top. In drawing this piece, we arbitrarily chose to draw the left (divided) parallelogram on top of the right one. The substitution rule for the isosceles triangle does not respect the symmetry of the triangle. We record the asymmetry by orienting the hypotenuse of each triangle, as shown. The new substitution rule clearly generates $I_2$. We still call our new rules $S_R$ and $S_T$. Here $R$ stands for a right isosceles triangle and $T$ still stands for the trapezoid.

Figure 4.9 shows two successive modifications of the rule for $S_T$.
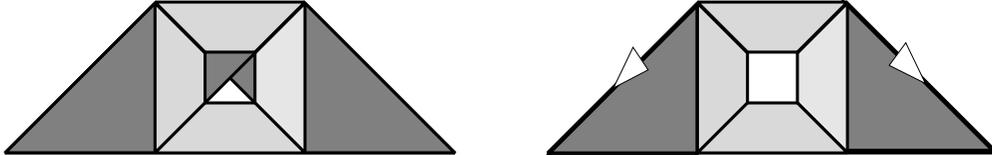


**Figure 4.9:** Two modifications of $S_T$.

For the first modification, we add another trapezoid to $S_T(T)$. This new trapezoid appears in $S_R(S_T(T))$, so the new rule rule generates the same limit. We give the same names to our new substitution rules. For the second modification, we delete two of the isosceles triangles. Call the deleted triangles $R_1$ and $R_2$. Observe that $S_R(R_k) \subset S_T(T)$ for $k = 1, 2$. Therefore, the latest subdivision rule has the same limit as the original. We again call the generators of the rule $S_T$ and $S_R$. This time $S_T$ and $S_R$ replace one piece by a union of pieces that have pairwise disjoint interiors.

From here it is a simple exercise to check that $S_R \circ S_T$ implements the rules shown in Figure 4.7. ♠

## 4.6 The Map from the Snowflake to the Carpet

The initial seed for $I_3$ is a cyclically ordered list $I_3(0)$ of 8 isosceles triangles $\tau_1, ..., \tau_8$. The initial seed for $I_2$ is a cyclically ordered list $I_2(0)$ of 8 trapezoids $T_1, ..., T_8$, which wrap around twice, as we have already mentioned. We have a correspondence $T_i \leftrightarrow \tau_i$ between the two seeds.

Since the two substitution rules are combinatorially identical, we inductively get a surjective map from the triangles in $I_3(n)$ to the quadrilaterals in $I_2(n)$. This map respects the cyclic ordering on both sets, and it also respects the relation of *descendence*. Say that a shape $\tau_n$ of $I_j(n)$ is a *descendent* of a shape $\tau_m$ of $I_j(m)$ if the substitution rule, applied $n - m$ times, to $\tau_m$, produces a list of shapes that contains $\tau_n$.

We define $\phi : I_3 \to I_2$ by taking the limit of our correspondence. Given $x \in I_3$ we can find a sequence $\{\tau_n\}$ of triangles, with $\tau_n \in X_3(n)$, that accumulates on $x$. We define $\phi(x)$ to be the limit of the corresponding quadrilaterals in $I_2$.

**Lemma 4.3** $\phi$ *is well-defined.*

**Proof:** Suppose that $\{\tau'_n\}$ is another sequence of triangles that converges to $x \in I_2$. Let $S_n$ and $S'_n$ be the shapes of $I_2(n)$ corresponding respectively to $\tau_n$ and $\tau'_n$. It suffices to prove that $\mathrm{diam}(S_n \cup S'_n) \to 0$.

Call two triangles in $I_3(m)$ *close* if they are either identical or else consecutive in the cyclic order. Let $d_n$ be the largest integer such that $\tau_n$ and $\tau'_n$ are descendents of close triangles in $I_3(d_n)$. Since the distance between $\tau_n$ and $\tau'_n$ converges to 0, and $I_3$ is evidently an embedded curve, $d_n \to \infty$.

We make the same definitions as above for $I_2$. We can find close shapes $\Sigma_n$ and $\Sigma'_n$ in $I_2(d_n)$ such that $S_n$ is a descendent of $\Sigma_n$ and $S'_n$ is a descendent of $\Sigma'_n$. Note that $\Sigma_n$ and $\Sigma'_n$ have intersecting boundaries in all cases. Also, the diameters of these shapes tends to 0 with $n$. Hence $\mathrm{diam}(\Sigma_n \cup \Sigma'_n) \to 0$. Finally, we have $S_n \subset \Sigma_n$ and $S'_n \subset \Sigma'_n$. Hence $\mathrm{diam}(S_n \cup S'_n) \to 0$ as desired. ♠

Essentially the same proof shows that $\phi$ is continuous. Since all the approximating correspondences are surjective, $\phi$ is surjective as well.

**Remark:** Note that we could get an equally canonical map if we compose $\phi$ with some isometry of $I_2$ or $I_3$. This basic (though trivial) ambiguity comes up in Statement 3 of our Main Theorem below.

## 4.7   Geometry of the Map

Now we give some geometrical information about the map $\phi : I_3 \to I_2$ defined in the previous section. Note that $I_3$ contains the vertices of triangles in $I_3(n)$ for all $n$. Say that a *special point* of $I_3$ is a right-angled vertex of $I_3(n)$ for some $n$. The set $A_3$ of special points is dense in $I_3$.

Say that a *special point* of $I_2$ is a midpoint of the long edge of some trapezoid in $I_2(n)$. We could equally define the special points to be the centers of the parallelograms. The two definitions coincide. Let $A_2$ denote the set of special points of $I_2$. The set $A_2$ is the set of midpoints of the edges of the countable family of squares that arises from the substitution rule explained in §4.5.

**Lemma 4.4**  $\phi(A_3) = A_2$.

**Proof:**  Let $v \in A_3$ be some point. Then $v$ is the right-angled vertex of a triangle $\tau_0$ of $I_3(n)$ for some $n$. Note that $v$ is also the right vertex of a triangle $\tau_k$ of $I_3(n + k)$ for $k = 1, 2, 3...$ The triangles $\tau_1, \tau_2, ...$ all have the same type, either inward pointing or outward pointing.

Consider the outward pointing case first. In this case the quadrilateral $T_k$ of $I_2(n+k)$ corresponding to $\tau_k$ is a trapezoid. Inspecting our subdivision rule, we see that $T_{k+1}$ is the middle trapesoid of $S_T(T_k)$. Looking at Figure 4.5, we see that $T_k$ and $T_{k+1}$ have a common distinguished point. This holds for all $j$. Hence $\bigcap T_k = \phi(v)$ is the common distinguished point of all these trapezoids. Hence $\phi(v) \in A_2$.

In the inward pointing case, the quadrilateral $P_k$ corresponding to $\tau_k$ is a parallelogram. Here $P_{k+1}$ is the middle parallelogram of $S_P(P_k)$. But these parallelograms have a common center. The intersection $\bigcap P_k = \phi(v)$ is this common center. Looking at Figure 4.5 again, we see that the centers of the parallelograms coincide with distinguished points of trapezoids. So, again $\phi(v) \in A_2$.

Our argument so far shows that $\phi(A_3) \subset A_2$. Any $x \in A_2$ is the distinguished point of some trapezoid $T$ of some $I_2(n)$. But then there is a triangle $\tau$ of $I_3(n)$ which corresponds to $\tau$. By construction $x = \phi(v)$, for the right vertex $v$ of $\tau$. Therefore $\phi(A_3) = A_2$. ♠

Let $B_3 \subset I_3$ denote the set of points that are acute vertices of triangles in $I_3(n)$ for all the different $n$. Then $B_3$ is dense in $I_3$. Let $B_2 \subset I_2$ denote

the set of vertices of quadrilaterals in $I_2(n)$ all the different $n$. The set $B_2$ is the set of corners of the distinguished squares mentioned above in connection with $A_2$. The curve $I_3$ looks "oscillatory" in the neighborhood of any $x \in B_3$. Any line through $x$ intersects $I_3$ infinitely often in every neighborhood of $x$.

**Lemma 4.5** $\phi(B_3) = B_2$.

**Proof:** The proof is similar to the one for the previous result, so we will be a bit sketchy. Any $x \in B_3(n)$ is the acute vertex of a triangle $\tau_k$ in $I_3(n+k)$. The triangle $\tau_{k+1}$ is either the first or last triangle in the list of 5 triangles in $I_3(n+k+1)$ that replaces $\tau_k$. Whether $\tau_{k+1}$ is "first" or "last" is independent of $k$.

Let $T_k$ be the quadrilateral of $I_2(n+k)$ corresponding to $\tau_k$. By construction, these quadrilaterals all share a common vertex, and $\phi(x)$ must be this common vertex. This shows that $\phi(B_3) \subset B_2$. The proof that $B_2 \subset \phi(B_3)$ is just as in the previous result. ♠

**Remark:** It is worth pointing out one more feature of the map $\phi$: It is far from 1 to 1. The set $I_2(n)$, considered as a list of quadrilaterals, contains multiple copies of the same trapezoid. Many trapezoids appear $2^n$ times in $I_2(n)$. From this we see that, for any $n$, there are $2^n$ segments of $I_3$ that $\phi$ identifies. For this reason, the canonical surjection from $I_3$ to $I_2$ is a many-to-one map. We think if it is a kind of fractal version of the universal covering map from the line to the circle.

## 4.8 A Hidden Symmetry of the Carpet

The snowflake has a hidden symmetry, as discussed in §4.3. Given the close correspondence between the carpet and the snowflake, we would expect a combinatorially identical hidden symmetry to appear for the carpet. This is indeed the case. In this section we describe the symmetry and sketch the proof.

The hidden symmetry of the carpet manifests itself in two ways, and this at first seems different from what happens with the snowflake. However, in the case of the snowflake, there are really two kinds of right-angled isosceles triangles, so the symmetry there actually does arise in two ways.
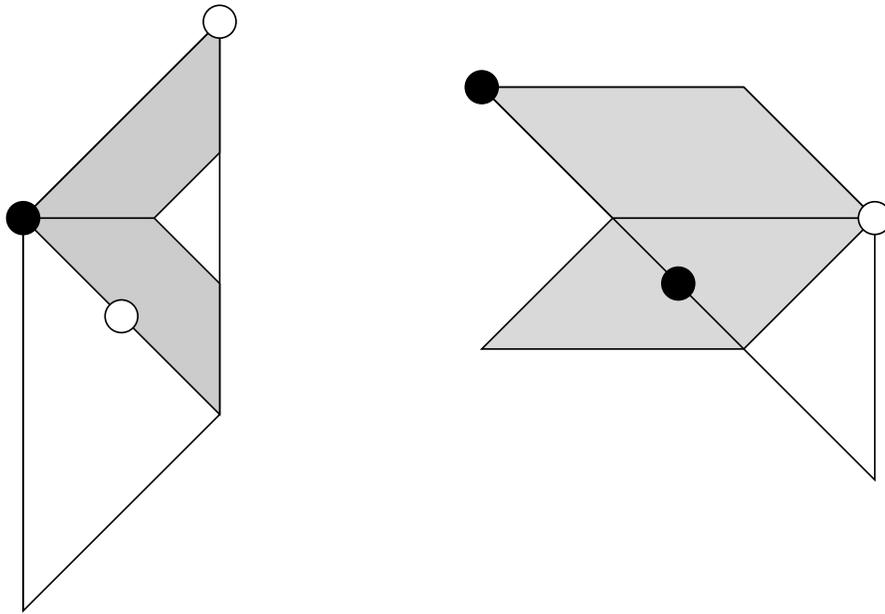
**Figure 4.7:** Some special quadrilaterals.

The left side of Figure 4.7 shows a large parallelogram and two smaller shaded trapezoids. We can generate a subset of a suitably scaled copy of the carpet starting with the large parallelogram as a seed and then considering the portion of the resulting limit that connects the two white dots in Figure 4.7. Let $A$ denote the resulting set. Likewise, we start with each shaded trapezoid as a seed, and consider the portion of the resulting limit that connects a black dot to a white dot. Let $A_1$ and $A_2$ be the resulting sets.

The right hand side of Figure 4.7 shows a trapezoid and two shaded parallelograms. We let $B$ and $B_1$ and $B_2$ be the portions of carpets produced by the same construction as we did for the $A$s, but interchanging the roles of *black* and *white*.

**Lemma 4.6** $A = A_1 \cup A_2$ *and* $B = B_1 \cup B_2$.

**Proof:** The proof here is formally isomorphic to what we did for the snowflake. We omit the details. ♠

# 5 The Main Result

## 5.1 Statement of the Result

Let $\omega = \exp(2\pi i/8)$ be the usual primitive 8th root of unity. Let $P$ be the regular octagon whose vertices are powers of $\omega$. Let $\Sigma_0$ be the pinwheel strip associated to $P$ as in §3.3. Let $\Sigma_0^1 \subset \Sigma_0$ denote the half-strip that lies to the right of the line $x = 1 + \sqrt{2}$. See Figure 5.1.
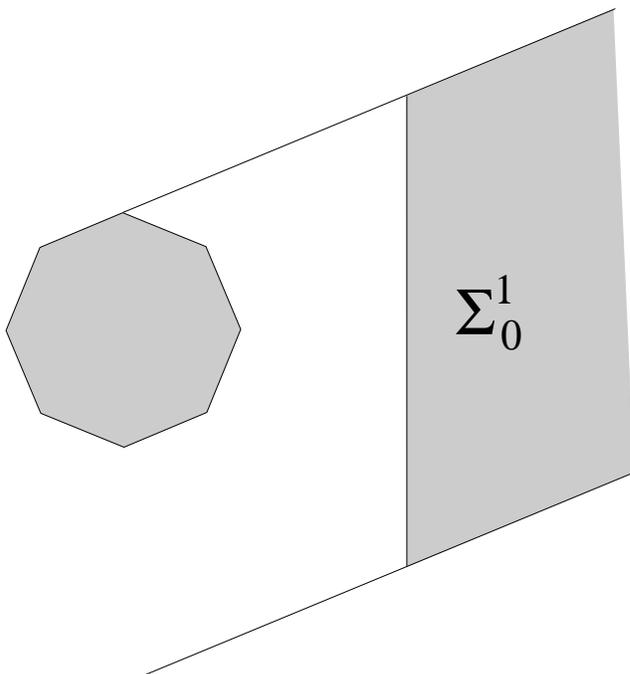


**Figure 5.1:** The relevant sets

§7.5 we prove

**Lemma 5.1** *Relative to the the pinwheel map for the regular octagon, any periodic orbit that intersects $\Sigma_0^1$ nontrivially does so in exactly $3^k$ points for some $k = 0, 1, 2....$*

Given a periodic point $x \in \Sigma_0^1$, we define $|x|$ to be the exponent $k$ such that the pinwheel orbit of $x$ intersects $\Sigma_0^1$ exactly $3^k$ times. We call a sequence $\{x_n\} \in \Sigma_0^1$ a *good sequence* if $|x_n| \to \infty$ as $n \to \infty$ and he congruence of $|x_n|$ mod 4 is odd, and independent of $n$.

37

Recall that $\pi_2$ and $\pi_3$ are the projections from $\mathbf{R}^8$ to $\mathbf{R}^2$ defined in the previous chapter. Define

$$s_2 = \sqrt{3}; \qquad s_3 = 1 + \sqrt{2}. \tag{12}$$

Let $\phi : I_3 \to I_2$ be the canonical surjection described above.

**Theorem 5.2 (Main Theorem)** *Let $\{x_n\}$ be a good sequence. Also, let $A_n \subset \mathbf{R}^8$ be the arithmetic graph of $x_n$. Then the following is true.*

1. *$\pi_2(A_n)$ and $\pi_3(A_n)$ are closed polygons for all $n$.*

2. *Let $S_{k,n}$ be dilation by $s_k^{-|x_n|}$ about the origin. Then the following limit exists.*
$$\Gamma_k = \lim_{n \to \infty} S_{k,n} \circ \pi_k(A_n).$$

3. *$\Gamma_3 = I_3$, where $I_3$ is scaled so that one of the $8$ isosceles triangles in its seed has vertices*
$$(0,0); \qquad (1/2, -s_3/2) \qquad (-s_3/2, -1/2).$$

4. *$\Gamma_2 = I_2$, where $I_2$ is scaled so that its center is $(3/2, 3/2)$ and one of its main corners is $(-3/2, 3/2)$.*

5. *Identifying points on $\pi_3(A_n)$ and $\pi_2(A_n)$ if they come from the same point on $A_n$, we get a polygonal surjectin*
$$\phi_n : S_{3,n}\pi_3(A_n) \to S_{2,n}\pi_2(A_n).$$

   *We have $\lim_{n \to \infty} \phi_n = F \circ \phi$. Here $F$ is an isometry of $I_2$ and $\phi : I_3 \to I_2$ is the canonical surjection from the snowflake to the carpet.*

**Remarks:**
(i) Experimentally, we see that Theorem 5.2 also holds when we consider sequences of points in $\Sigma_0 - \Sigma_0^1$, but these cases lead to annoying complications. We are interested in establishing a nice robust version of the phenomenon, but not necessarily the sharpest possible version.
(ii) The limits are exactly the same when $|x_n| \equiv 1 \bmod 4$ and when $|x_n| \equiv 3 \bmod 4$. The only thing that changes is the global isometry $F$. In the former case, $F$ is orientation preserving (and can be taken to be the identity) and in the other case $F$ is orientation reversing.
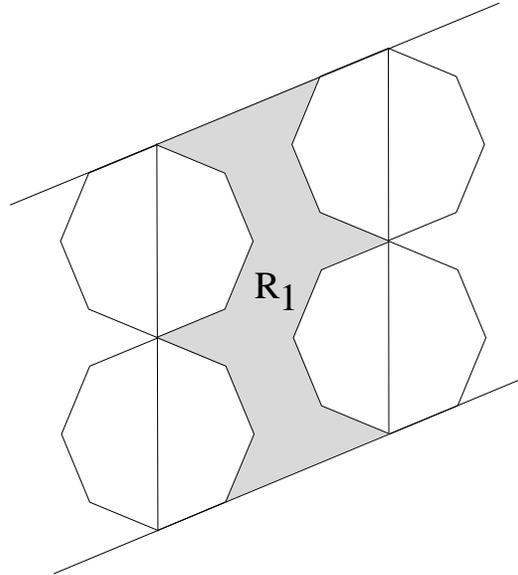
The rest of the paper is devoted to proving Theorem 5.2.

## 5.2 Restricting the Domain

Let $P$ be the regular octagon, and let $\Sigma_0$ be the pinwheel strip described in §2.4. Consider the octagons

$$O_k^\pm = P + k(\sqrt{2} + 1, \pm 1); \qquad k = 1, 2, 3 \ldots \tag{13}$$

$O_k^+$ and $O_k^-$ share a common vertex and fit inside $\Sigma_0$ as shown in Figure 5.2. Figure 5.2 shows the cases $k = 1, 2$, but the remaining cases look the same.



**Figure 5.2:** The region $R_1$.

We let $R_k$ be the region of $\Sigma_0$ bounded by $O_k^+ \cup O_k^-$ and $O_{k+1}^+ \cup O_{k+1}^-$. Figure 5.2 shows the region $R_1$, which is the main region of interest to us. Note that $R_k$ is translation equivalent to $R_1$ for all $k \geq 1$, so the picture in Figure 5.2 is typical.

The first purpose of this chapter is to prove the Invariance Lemma. This result implies that each $R_k$ is forward invariant under the pinwheel map. Following this, we will prove

**Lemma 5.3** *Each $R_k$ is an invariant set for the pinwheel map. Furthermore, if Theorem 5.2 is true for sequences in $R_1$ then it is true as originally stated.*

## 5.3 The Pattern of Octagons

let $P$ be as above. The octagons we described above are part of the infinite pattern of octagons shown in Figure 5.3. Figure 5.3 shows a part of an infinite set of octagons that, as it turns out, is invariant under the outer billiards map. Figure 5.3 also shows the strip $\Sigma_0$.
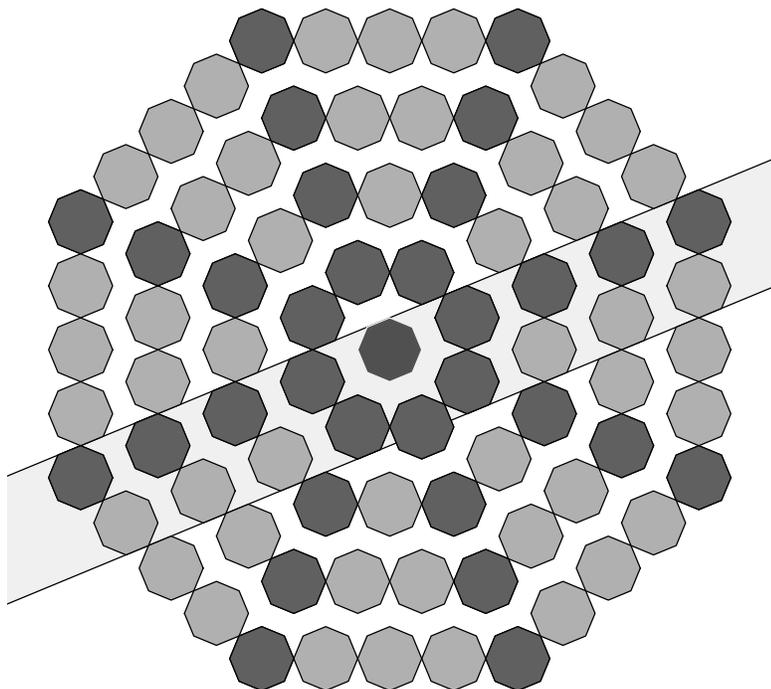


**Figure 5.3:** Octagonal orbits

The 8 dark rows of octagons are obtained by translating the central one by a vector of the form

$$k\omega^n(\sqrt{2}+1,1); \qquad k,n \in \mathbf{Z}; \qquad \omega = \exp(2\pi i/8). \qquad (14)$$

The remaining octagons interpolate between these dark ones. The octagons are situated in such a way that the outer billiards map is entirely defined on the interior of each one. The octagons come in an infinite family of rings, or *necklaces* $N_1, N_2, N_3, ....$ The octagons $O_k^{\pm}$ are precisely the two octagons of $\Sigma_0 \cap N_k$. Compare [**GS**] and [**K**].

**Lemma 5.4** *Each necklace is an outer billiards orbit. Moreover, the region between consecutive necklaces is invariant under the outer billiards map.*

**Proof:** We change notation slightly. Let $O(n, k; 0)$ be the octagon we get by translating the central one by the vector in Equation 14. Let $O(n, k; i)$ denote the octagon that is $i$ "clicks" away from $O(n, k; 0)$ going counterclockwise around $N_k$.

The outer billiard map has the same action on all the octagons lying on the vertical line segment between $O(-1; k; 0)$ and $O(1, k; -1)$ because such octagons all lie in the sector shown in Figure 5.4. The outer billiards map reflects each of the octagons under construction through the apex of the sector in Figure 5.4. We check easily that the action is as follows.

$$O(-1, k; 0) \to O(4, k; 1); \qquad O(1, k; -1) \to O(5, k; 0). \qquad (15)$$
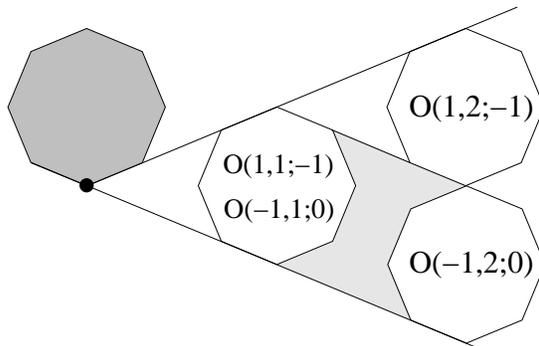


**Figure 5.4:** Octagons in a sector

The octagons $O(4, k; 1)$ and $O(5, k; 0)$ have their centers on the same vertical line segment $L$, as shown for $k = 2$ in Figure 5.4. By symmetry, the octagons between $O(-1, k; 0)$ and $O(1, k; -1)$ map to the octagons of the grid that have their centers on $L$. These octagons all belong to $N_k$. For the remaining octagons in $N_k$, the result follows from the 8-fold symmetry of the picture.

Call two octagons $O_1 \in N_k$ and $O_2 \in N_{k+1}$ *close* if they are as close together as possible. Two close octagons determine a polygonal region "between them" like the one that is shaded in Figure 5.4. Call such a region a *buffer*. If the two octagons belong to the sector from Figure 5.4, then the buffer is mapped between $N_k$ and $N_{k+1}$. But then, by symmetry, all buffers are mapped between $N_k$ and $N_{k+1}$. But every point between $N_k$ and $N_{k+1}$ is contained in a buffer. ♠

## 5.4 Invariant Domains in the Strip

Let $\Sigma_0'$ denote those points in $\Sigma_0$ that lie to the central octagon $P$. A point $p$ lies in $\Sigma_0'$ if every vector pointing from $P$ to $p$ has positive $x$ coordinate. Let $\Phi' : \Sigma_0' \to \Sigma_0'$ be the first return map of the $\phi^2$, the square outer billiards map.

**Corollary 5.5** *For $k \geq 2$ the set $R_k$ is invariant under $\Phi'$.*

**Proof:** For $k \geq 2$, the region between $N_k$ and $N_{k+1}$ intersects $\Sigma_0'$ exactly in the set $R_k$. ♠

**Remark:** The region between $N_1$ and $N_2$ intersects $\Sigma_0$ in a set that is somewhat larger than $R_1$., as one can see by looking carefully at Figure 5.3. It is for this reason that Corollary 5.5 fails for $R_1$.

**Corollary 5.6** *For $k \geq 4$ the set $R_k$ is invariant under $\Phi$, the pinwheel map.*

**Proof:** For $k \geq 4$ the argument in Lemma 2.1 goes through, and shows that $\Phi = \Phi'$. ♠

Below we will prove the following result.

**Lemma 5.7** *Let $\mu = 1, 2, 3...$ Let $p \in R_1$ and*

$$q = p + 2\mu(\sqrt{2} + 1, 1) \in R_{1+\lambda}.$$

*Then the arithmetic graphs $\Gamma(p)$ and $\Gamma(q)$ coincide.*

**Corollary 5.8** *The set $R_k$ is $\Phi$-invariant for all $k \geq 1$.*

**Proof:** By Lemma 5.7, translation by the vector $4(\sqrt{2} + 1, 1)$ conjugates the action of $\Phi$ on $R_1$ to the action of $\Phi$ on $R_5$. We already know that $R_5$ is invariant under $\Phi$ and the conjugation property implies the same result for $R_1$. A similar argument takes care of $R_2$ and $R_3$. We have already handled the remaining cases. ♠

**Remark:** Note that $\Phi$ and $\Phi'$ do not coincide on $R_1$, because Corollary 5.5 is false for $R_1$ whereas Corollary 5.8 is true. These maps in fact coincide for $k = 2, 3, 4...$.

**Proof of the Invariance Lemma:** We want to show that $\Sigma_0^k$ is $\Phi$ invariant. Let $N_k^+$ be the set of half-octagons obtained by taking the right halves of all the octagons in $N_k$. Since $\phi^2$ is a piecewise translation, the set $N_k^+$ is invariant under $\phi^2$. But then the set

$$S_k = N_k^+ \cap \Sigma_0$$

is $\Phi'$ invariant for $k \geq 4$. Hence, by the argument in Lemma 2.1, the set $S_k$ is $\Phi$-invariant for $k \geq 4$. As in the proof of Corollary 5.8, we now conclude that $S_k$ is $\Phi$ invariant for all $k = 1, 2, 3...$ Finally

$$\Sigma_0^k = S_k \cup R_{k+1} \cup S_{k+1} \cup R_{k+2} \dots.$$

We have decomposed $\Sigma_0^k$ into $\Phi$ invariant sets. Hence $\Sigma_0^k$ is also $\Phi$ invariant. ♠

**Proof of Lemma 5.3:** Suppose we want to prove Theorem 5.2 for a sequence of periodic points $\{x_n\}$ in $\Sigma_0^1$. Since $\Phi$ is the identity on each of the orbits $N_k$, we have $x_n \in R_{k_n}$ for large $n$. Chopping off the initial portion of the sequence, we can assume that $x_n \in R_{k_n}$ for all $n$.

By Lemma 5.7, we can find a point $y_n \in R_1 \cup R_2$ such that $x_n$ and $y_n$ have the same arithmetic graph. Hence, it suffices to prove Theorem 5.2 for the sequence $\{y_n\} \in R_1 \cup R_2$.

The intersection $R_1 \cup R_2$ is a single point, namely $\zeta = (2 + 2\sqrt{2}, 1)$. We call two points *specially related* if reflection through $\zeta$ interchanges the two points. We check by hand the following property, which we call *Property P*: If $y$ and $z$ are specially related then $\Phi(y)$ and $\Phi(z)$ are specially related and $\widehat{\Phi}(y) + \widehat{\Phi}(z) = 0$.

From Property P we see that the arithmetic graphs of $y$ and $z$ are isometric to each other. Using this basic principle, we can replace each $y_n \in R_2$ by the specially related point $z_n$ without changing the isometry type of the rescaled limits of the arithmetic graphs. ♠

**Remark:** In §7 we will explain in detail how we actually check Property P.

## 5.5 Equivalent Points in Strips

The rest of the chapter is devoted to the proof of Lemma 5.7. In this section we make a general observation about strip maps. In the section following this one, we will apply our observation to the case of the regular octagon.

Let $\Sigma_1$ and $\Sigma_2$ be non-parallel strips. Let $V$ and $W_1, W_2$ be the three vectors shown in Figure 5.3. We say that $p, q \in \Sigma_1$ are *related* if $p - q = kW_1$ for some integer $k$. Likewise, we say that $p, q \in \Sigma_2$ are *related* if $p - q = kW_2$ for some integer $k$.
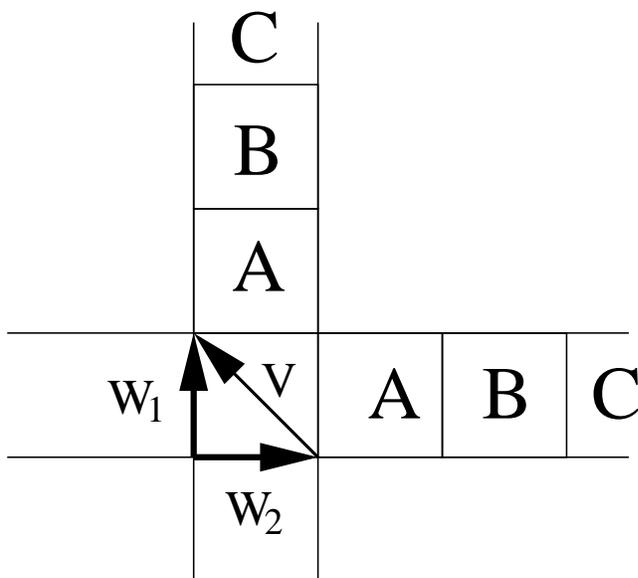


**Figure 5.4:** A Strip map

**Lemma 5.9** *Let $f$ be the strip map associated to $(\Sigma_2, V)$. If $p, q \in \Sigma_1$ are related then $f(p), f(q) \in \Sigma_2$ are related.*

**Proof:** Our statement is an affinely invariant one. So, we normalize our strips so that they are perpendicular and each have width 1. In this case, $f$ translates the horizontal squares $A, B, C, ...$ to the vertical squares with the same labels. If $p, q \in \Sigma_1$ are related then they are in the same positions relative to the horizontal squares that contain them. But then $f(p), f(q) \in \Sigma_2$ are in the same position relative to the vertical squares that contain them. Hence $f(p)$ and $f(q)$ are related. ♠

## 5.6 Proof of Lemma 5.7

We first prove Lemma 5.7 in case that $\lambda = 2\mu$ is even.

We define
$$W_{2n} = W_{2n+1} = 2\omega^n(\sqrt{2}+1, 1). \tag{16}$$

The vector $W_0 = W_1$ is parallel to both pointed strips $\Sigma_0$ and $\Sigma_1$. The vector $W_2 = W_3$ is parallel to both $\Sigma_2$ and $\Sigma_3$, and so on.

We say that two points $p, q \in \Sigma_{k-1}$ are *equivalent*

$$q - p = \mu W_{k-1}; \qquad \mu \in \mathbf{Z}. \tag{17}$$

We let $\mu_{k-1}(p, q)$ be the integer in Equation 17.

Let $f_k$ be the strip map associated to $(\Sigma_k, V_k)$, as in §2.3. Let $m_k(p)$ be as in Equation 2. That is,

$$f_k(p) = p + m_k(p)V_k. \tag{18}$$

**Lemma 5.10** *Let $f = f_k$. If $p, q \in \Sigma_{k-1}$ are equivalent and $k$ is odd, then $f(q), f(q)$ are equivalent in $\Sigma_k$ and*

$$\mu_k(f(p), f(q)) = \mu_{k-1}(p, q); \qquad m_k(p) = m_k(q).$$

**Proof:** By rotational symmetry, it suffices to consider the case when $k = 1$. Suppose that $k = 1$. Note that $\Sigma_0 \cap \Sigma_1$ is a strip parallel to $W_0 = W_1$. So, $p \in \Sigma_0 \cap \Sigma_1$ if and only if $q \in \Sigma_0 \cap \Sigma_1$. Suppose that $p, q \in \Sigma_0 \cap \Sigma_1$. Then $f(p) = p$ and $f(q) = q$. Also, $W_1 = W_0$. Hence $f(p)$ and $f(q)$ are equivalent in $\Sigma_1$ and

$$\mu_1(f(p), f(q)) = \mu_0(p, q); \qquad f_1(p) = f_1(q) = 0.$$

If $p, q \in \Sigma_0 - \Sigma_1$ then $p + V_1 \in \Sigma_1$ and likewise $q + V_1 \in \Sigma_1$. Here

$$f_1(q) - f_1(p) = q - p.$$

Hence
$$\mu_1(f(p), f(q)) = \mu_0(p, q); \qquad f_1(p) = f_1(q) = 1.$$

♠

45

**Lemma 5.11** *Let $f = f_k$. If $p, q \in \Sigma_{k-1}$ are equivalent and $k$ is even, then $f(q), f(q)$ are equivalent in $\Sigma_k$ and*

$$\mu_k(f(p), f(q)) = \mu_{k-1}(p, q); \qquad m_k(q) = m_k(p) + \mu_{k-1}(p, q).$$

**Proof:** By rotational symmetry, it suffices to consider the case $k = 2$. In this case, the strips $\Sigma_1$ and $\Sigma_2$, as well as the vectors $W_1$ and $W_2$ and $V = V_2$ are related precisely as discussed in §5.5. Lemma 5.9 now implies that $p$ and $q$ are equivalent in $\Sigma_2$ and $\mu_2(p, q) = \mu_1(p, q)$.

Referring to Figure 5.4, which is an affine image of the octagon picture, the number $\mu_1(p, q)$ counts the number of parallelograms one needs to shift in order to move $p$ over to $q$. For instance, if $p$ were in parallelogram $A$ and $q$ were in parallelogram $C$ in Figure 5.4, then $\mu_0(p, q) = 2$. From Figure 5.4, we see clearly that $m_2(q) = m_k(p) + \mu_1(p, q)$. ♠

Now we can finish the proof of Lemma 5.7. Suppose that $p, q \in \Sigma_0$ are equivalent. Let $\mu = \mu_0(p, q)$. Then, referring to Equation 4, we have

$$\widehat{\Phi}(q) - \widehat{\Phi}(p) = \sum_{i=1}^{8} (m_k(q) - m_k(p))\widetilde{V}_k =^1 \sum_{k \text{ even}} \mu \widetilde{V}_k =^2 0. \qquad (19)$$

Equality 1 comes from Lemmas 5.10 and 5.11, and Equality 2 comes from the fact that $\widetilde{V}_6- = \widetilde{V}_2$ and $\widetilde{V}_8 = -\widetilde{V}_4$.

# 6    A Toy Model

## 6.1    The Map

The work in the previous chapter reduces the proof of Theorem 5.2 to the study of the pinwheel map $\Phi : R_1 \to R_1$, where $R_1$ is shown in Figure 5.2. This first return map turns out to be a polygon exchange map. In this chapter, we will briefly discuss a simpler but related polygon exchange map.

Though the simpler map is not the one we will ultimately study in our proof of Theorem 5.2, we think that it will prepare the reader for the more complicated system we do study. First, the basic shapes that arise in the simple system here arise in the more complicated system we study later. Second, both the system here and the one we do study have a *renormalization scheme*. It is the renormalization scheme for the first return map that leads to the self-similar nature of the arithmetic graphs.

The dynamical system we study here is well known, and indeed also arises in the study of outer billiards: It is the first return map to a certain invariant domain for a dynamical system generated by the outer billiards map and the rotational symmetry group of the regular octagon. Compare [**T2**] and [**BC**]. The system is a self map of the kite-shaped region $X$, shown as the shaded region in Figure 6.1. The region $X$ is best defined in terms of a large regular octagon. The smaller regular octagon, which is a subset of $X$, is drawn in for later reference.
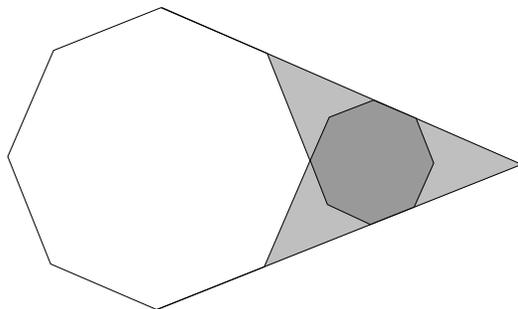


**Figure 6.1:** The region $X$.

As shown in Figure 6.2, the region $X$ can be partitioned in two ways into right angled isosceles triangles,

$$X = A_1 \cup B_1 = A_2 \cup B_2. \tag{20}$$

There is a unique orientation-preserving isometry $\phi_A : A_1 \to A_2$. The map $\phi_A$ rotates $3\pi/4$ radians clockwise about the center of the small octagon in Figure 6.1. Likewise, there is a unique orientation preserving isometry $\phi_B : B_1 \to B_2$. The map $\phi_B$ rotates $\pi/4$ radians counterclockwise about the center of the big octagon in Figure 6.1. We have a map $\phi : X \to X$ which restricts to $\phi_A$ on $A_1$ and $\phi_B$ on $B_1$. The map $\phi$ is defined except on the segment common to $A_1$ and $B_1$.
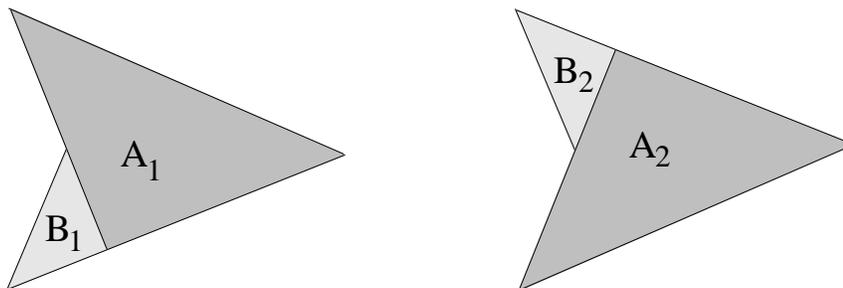


**Figure 6.2:** Two partitions of $X$.

## 6.2 The Renormalization Scheme

Let $O \subset X$ denote the inner octagon in Figure 6.1. By construction, the center of $O$ is a fixed point of $\phi$, and every point of $O$ has period 8 with respect to $\phi$.
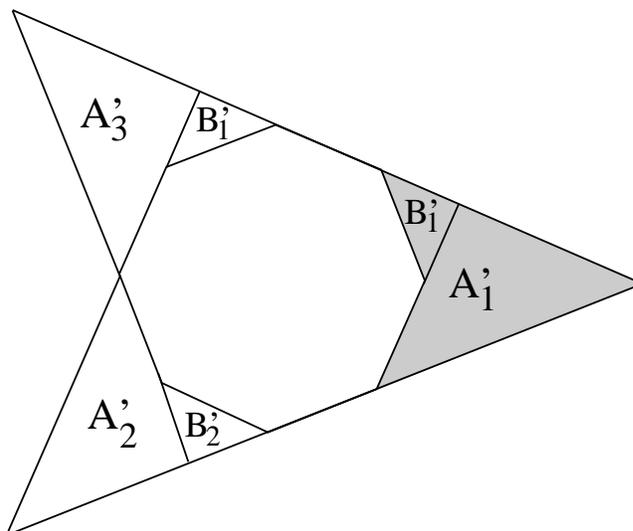


**Figure 6.3:** Renormalization Scheme

Let $X'$ be the shaded component of $X - O$ shown in Figure 6.3. We have partitioned $X'$ into two pieces, which we call $A_1'$ and $B_1'$. The third power $\phi^3$ maps $X'$ to itself. Figure 6.2 shows the pieces $A_2' = \phi(A_1')$ and $A_3' = \phi(A_2')$ and $B_2' = \phi(B_1')$ and $B_3' = \phi(B_2')$.

We have not drawn $A_4', B_4' \subset X'$, but we observe that $A_4'$ and $B_4'$ relate to $A_1'$ and $B_1'$ in the same say that $A_2$ and $B_2$ relate to $A_1$ and $B_1$. More precisely, there is an orientation reversing dilation $\Theta : X' \to X$ such that

$$\Theta^{-1} \circ \phi \circ \Theta = \phi^3 \tag{21}$$

on $X'$. The map $\Theta$ is the *remormalization map*. Having $\Theta$ by itself is nice, but we have a second property that really pins things down. Namely,

$$X - O = X' \cup \phi(X') \cup \phi^2(X'). \tag{22}$$

In other words, every point of $X - O$ lies in the orbit of a point in $X'$.

## 6.3 The Periodic Points

Our remormalization scheme allows us to get a complete understanding of the periodic points of $\phi$. The octagon $O$ consists of points having period 8. (The center is the one point of $O$ that has period 1.) For our purposes, it is nicer to think of $O$ as having period 1 *as a tile*. The set $O$ is preserved by $\phi$. We set $O_0 = O$.

By equation 21, the region $X'$ contains a small octagon

$$O_0' = \Theta^{-1}(O_0) \tag{23}$$

that has tile-period 3. We set

$$O_{k+1}' = \phi(O_k'). \tag{24}$$

The tiles $O_k'$ for $k = 0, 1, 2$ all have tile-period 3. They comprise a tile-orbit.

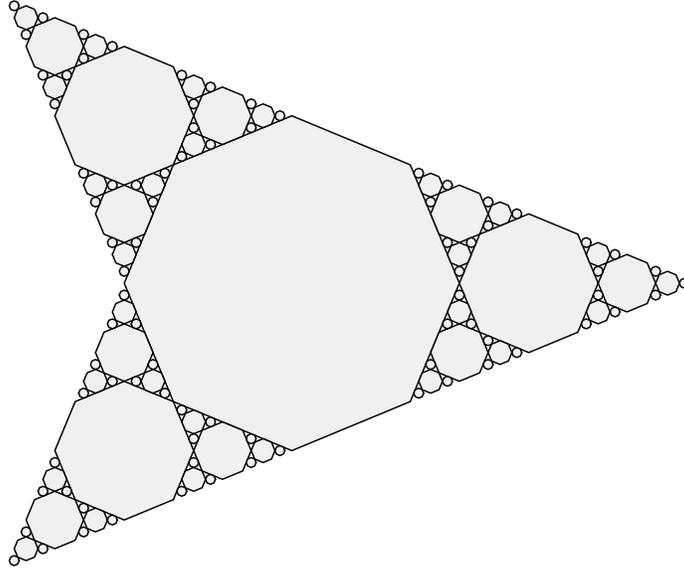Now we iterate. By Equation 21, the three octagons

$$O_{3k}'' = \Theta^{-1}(O_k'); \qquad k = 0, 1, 2 \tag{25}$$

all have tile-period 9. Setting

$$O_{k+1}'' = \phi(O_k''), \tag{26}$$

we produce 9 small octagons $O_0'', ..., O_8''$ that all have tile-period 9. Next, we produce 27 small octagons $O_k^{(3)}$ for $k = 0, ..., 26$, all having tile-period 27. Figure 6.4 shows some of these octagons.

**Figure 6.4:** Cascade of periodic tiles

Thus, we have produced, for each $n = 0, 1, 2, 3...$, a periodic octagonal tile having tile-period $3^n$. We claim that every periodic point is contained in one of these octagons. To see this, we use Equation 22. Clearly, the only periodic points of order 1 lie in $O_0$. Suppose $N$ is the smallest positive integer for which we have not proven our result. Let $y'$ be a periodic point of period $N$.

Since $N > 1$, we have $y' \in X - O$. By Equation 22, the point $y'$ lies in the orbit of some point $x' \in X'$. Let $x = \Theta(x')$. Note that $\phi(X')$ and $\phi^2(X')$ are disjoint from $X'$. Since $\phi^N(x') = x'$, we must therefore have $N = 3M$ for some integer $M$. But then, by Equation 21, we see that $x$ has period $M$. By induction, $x$ lies in one of our octagonal tiles. But then, by construction, so does $x'$. Now we know that $x'$ is one of the periodic points we have already studied. So, the same goes for $y'$, the original point of interest to us.

Using the renormalization scheme, we have been able to classify all the periodic points of the system. Incidentally, we note that the complement of the octagons has dimension larger than 1, and the set of points where some iterate of $\phi$ is undefined has dimension 1. Therefore, $\phi$ has some aperiodic points. This is the same argument Tabachnikov [**T2**] gives for the regular pentagon.

# 7 The Pinwheel Dynamics

## 7.1 The Partition

We will work with the region $R_1$ from Lemma 5.3. Both sides of Figure 7.1 show $R_1$. Note that $R_1$ is the union of 4 copies of the region $X$ from the previous chapter.
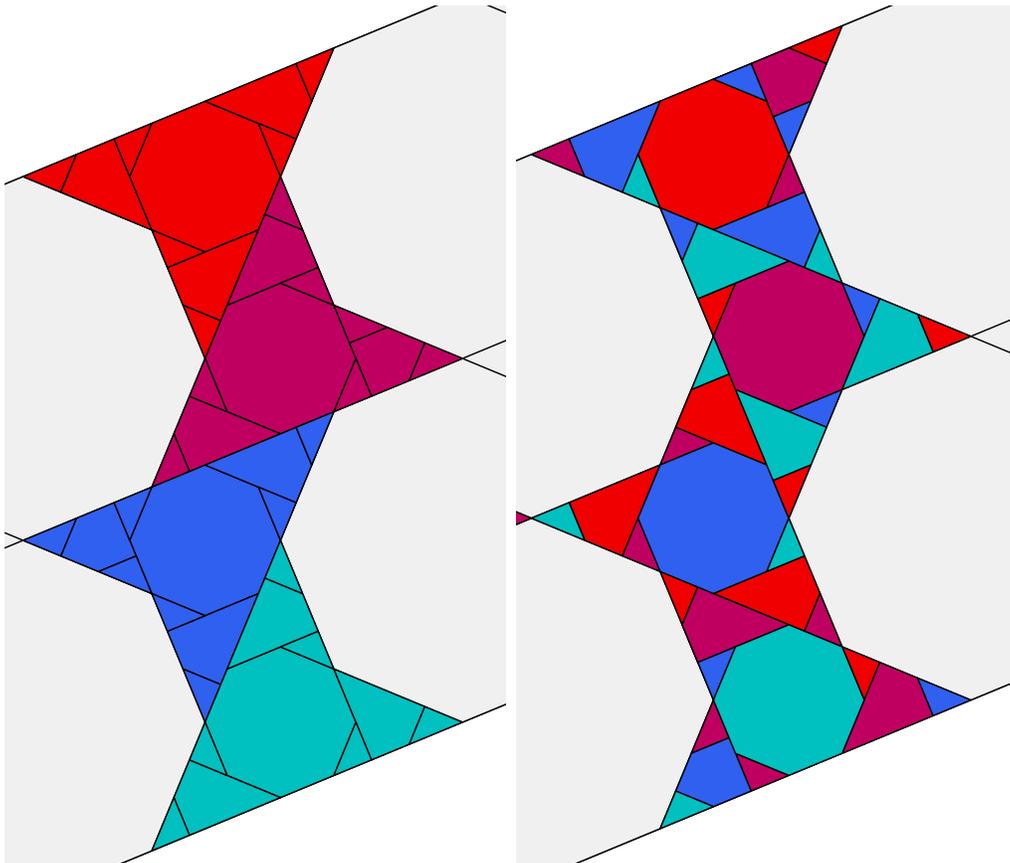


**Figure 7.1:** The region $R_1$, inside the strip $\Sigma_0$.

Figure 7.1 shows the action of $\Phi$ on $R_1$. The left hand side of the figure shows the partition of $R_1$ into the maximal pieces on which $\Phi$ and $\widehat{\Phi}$ are constant, and the right hand side shows the images of these pieces under $\Phi$. We will give coordinates for these tiles below.

Just knowing that $\Phi$ is a piecewise translation, we can nearly pin down the action of $\Phi$ from Figure 7.1. The problem is that there are some pairs

of triangles, having identical shading, that are translates of each other. Such pairs are specially situated: One triangle in each such pair lies on the left hand side of $R_1$ and the other triangle lies in the right hand side. To pin down the action of $\Phi$ exactly, we mention that $\Phi$ maps any triangle on the left (respectively right) hand side of $R_1$ to a triangle on the right (respectively left) hand side of $R_1$.

**Remark:** It seems worth also mentioning that $\Phi$ maps any quadrilateral or pentagon on the left (respectively right) hand side of $R_1$ to another quadrilateral or pentagon on the left (respectively right) hand side of $R_1$. Finally, $\Phi$ fixes each of the central octagons.

$\Phi$ has one obvious symmetry and one hidden symmetry. Let $\rho_1$ be the order-2 rotation about the center point of $R_1$. Then $\Phi$ commutes with $\rho$.

To describe the hidden symmetry, we observe that it is better to think of $R_1$ as a subset of an infinite cylinder. Define $\sigma(x, y) = (x, y+2)$. The squared-map $\sigma^2$ maps the bottom boundary of $\Sigma_0$ to the top one. The quotient, obtained by identifying the two boundary components, is a cylinder. We think of $R_1$ as a subset of this cylinder, so that certain pieces on the top of $R_1$ are contiguous with pieces on the bottom.

This point of view makes the picture look more symmetric. Consider the magenta pentagon $K'$ shown on the middle left hand side of Figure 7.1. Let $K''$ be the magenta triangle such that $K = K' \cup K''$ is a kite, isometric to the rest of the kites in the picture. $\Phi(K')$ lies at the very top of $R_1$ and $\Phi(K'')$ lies at the very bottom. Under the identification we have been discussing, $\Phi(K'')$ is just the continuation of $\Phi(K')$. So, when we think of $R_1$ as a subset of a cylinder, $K$ is a single tile on which $\Phi$ is constant. The other pentagon/triangle pair on the left hand side of Figure 7.1 has the same analysis. Considered this way, we see that the partition of $R_1$ consists just of kites, triangles, and octagons.

Finally, we mention the hidden symmetry. Once we interpret $R_1$ as being a subset of the cylinder, $\Phi$ commutes with $\sigma$.

**Remark:** It turns out that $\widehat{\Phi}$ does not share the same symmetries. That is, it is not true that the maps $\widehat{\Phi}$ and $\widehat{\Phi} \circ \rho$ and $\widehat{\Phi} \circ \sigma$ coincide. However, as we will see, there is some relation between these maps.

## 7.2 Notation and Coordinates

This section does not contribute to the mathematics at all. We include it for readers who would like to see explicit coordinates to that they can reproduce our experiments and calculations on their own. Let

$$s = \sqrt{2} + 1. \tag{27}$$

Let $O_3$ be the red octagon in Figure 7.1. The octagon $O_3$ has radius $1/s$. By this we mean that the distance from the center of $O_3$ to any vertex is $1/s$. We point out another octagon, $O_4$, of radius $1/s^2$. The octagon $O_4$ is homothetic to $O_3$, and the rightmost vertex of $O_4$ coincides with the leftmost vertex of $O_3$. Both $O_3$ and $O_4$ are periodic tiles for $\Phi$. See Figure 7.2 below. Now we distinguish 4 special points.

- The point $c_1 = (3s, 1)/2$ is the center of $R_1$.

- The point $c_2 = (3s, 3)/2$ is the center of the top half of $R_1$.

- The point $c_3 = (2 + \sqrt{2}, 2)$ is the center of $O_3$.

- The point $c_4 = (2\sqrt{2}, 2)$ is the center of $O_4$.

Let $\rho_k$ denote counterclockwise rotation by $2\pi k/8$ about $c_k$.
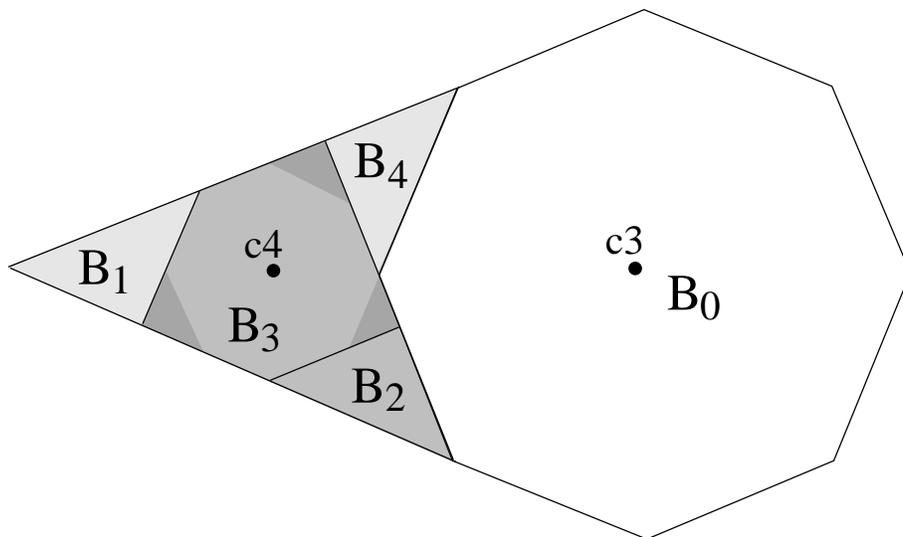


**Figure 7.2:** Some polygons of interest.

Figure 7.2 shows 4 shaded polygons. For these polygons, we will use a second labelling scheme in which $O_4 = B_0$. Figure 7.2 shows the $B$ labels. The union $B_2 \cup B_3$ is the kite that lies to the left of the red octagon in Figure 7.1. We have tried to shade $B_3$ in such a way that we point out $O_4 \subset B_3$.

The vertices of $2B_3$ are

$$(5s-6, 2s-1) \quad (4s-4, 4s-6) \quad (3s-2, 2s-1) \quad (s+3, 3s-3) \quad (2s+1, s+2). \tag{28}$$

We divide these coordinates in half to get the coordinates of $B_3$.

The vertices of $2B_1$ are given by

$$(2s, 4); \qquad (3+s, 3s-3); \qquad (3s-2, 2s-1). \tag{29}$$

We divide these coordinates in half to get the coordinates of $B_3$.

We can deduce the remaining coordinates just from what we have already given. For instance,

$$B_4 = \rho_4^5(B_1); \qquad B_2 = \rho_4^3(B_1). \tag{30}$$

The remaining polygons in the partition of $R_1$ all have the form

$$\rho_1^{a_1} \circ \rho_2^{a_2} \circ \rho_3^{a_3}(B); \qquad a_1, a_2 \in \{0,1\}; \qquad a_3 \in \{3,5\}. \tag{31}$$

Here either $B = B_j$ or $B = B_2 \cup B_3$.

Now we explain briefly how we verify that the return map $\Phi$ is as stated. We explain this for $B_3$. The check is the same for the remaining files. Recall that $\Phi = f_{16} \circ ... \circ f_1$, where $f_k$ is the strip map associated to the pair $(\Sigma_k, V_k)$. We verify that there are integers $m_1, ..., m_{16}$ such that

- $B_3^1 := B_3 + m_1 V_1 \in \Sigma_1$.

- $B_3^2 := B_3^1 + m_2 V_2 \in \Sigma_2$.

- $B_3^3 := B_3^2 + m_3 V_3 \in \Sigma_3$.

And so on. By convexity we just have to check this on the vertices of $B_3$. This is a short calculation we omit. This calculation shows not only that $\Phi(B_3) = B_3^{16}$, but also that $\widehat{\Phi}$ is constant on the interior of $B_3$.

## 7.3  The Compressed System

Let $R \subset R_1$ denote the top half of $R_1$. The region $R$ is colored red and magenta on the left half of Figure 7.1. Note that $R$ is a fundamental domain for the action of $\sigma$ on $R_1$. We define a new dynamical system $\Psi : R \to R$ as follows: $\Psi(p) = \Phi(p)$ if $\Phi(p) \in R$ and $\Psi(p) = \sigma \circ \Phi(p)$ if $\Phi(p) \notin R$. In other words, we use the action of $\sigma$ to move everything into $R$.
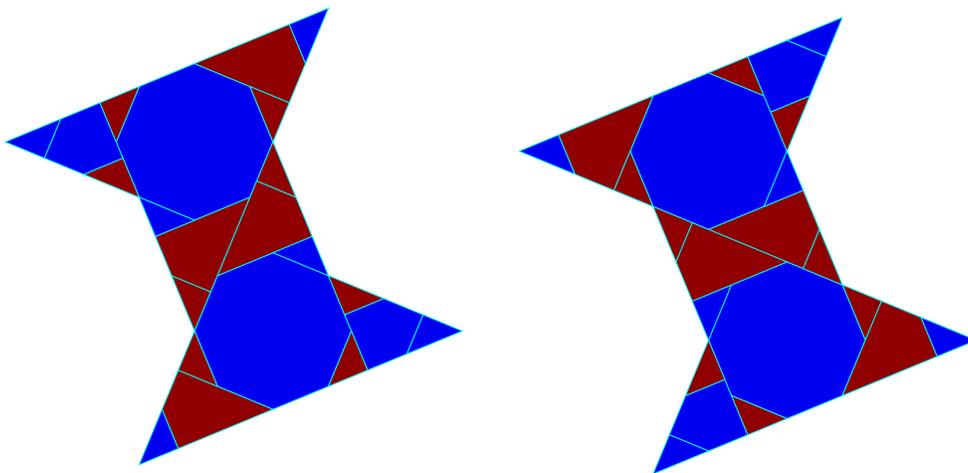


**Figure 7.3:** The compressed system

The compressed system has 22 regions. The regions $B_0, ..., B_4$ are as in Figure 7.2. For the record, here are the remaining regions.

- $B_5 = \rho_3^5(B_1)$ and $B_6 = \rho_3^5(B_2 \cup B_3)$ and $B_7 = \rho_3^5(B_4)$.

- $B_8 = \rho_3^2(B_1)$ and $B_9 = \rho_3^2(B_2 \cup B_3)$ and $B_{10} = \rho_3^2(B_4)$.

- $B_{k+11} = \rho_2^4(B_k); \qquad k = 0, ..., 10.$

The regions $B_5, B_6, B_7$ all lie in the upper right corner of $R$. The regions $B_8, B_9, B_{10}$ lie in the middle of $R$ on the left side. The remaining 11 regions are symmetrically placed with respect to the first 11. Again for the record, the red colored regions are

$$B_{i+11j}; \qquad i = 2, 4, 6, 7, 8, 9; \qquad j = 0, 1. \tag{32}$$

We will use this notation system in the next chapter, just so that we can list things in a completely explicit way.

Figure 7.3 does for $\Psi$ what Figure 7.1 does for $\Phi$. The left hand side shows the partition of $R$ into maximal regions on which $\Psi$ is constant. The blue polygons are the ones on which $\Psi = \Phi$ and the red polygons are the ones on which $\Psi = \sigma \circ \Phi$. The right hand side shows the images of these pieces under $\Psi$. Once we mention that $\Psi$ is a piecewise translation that maps all the triangles on the left (respectively right) side of $R$ to triangles on the right (respectively left) side of $R$, the partitions alone determine $\Psi$. We also remark that $\Psi$ maps the kites and pentagons on the left (respectively right) side of $R$ to the left (respectively right) side of $R$.

The map $\Psi$ has a symmetry that it inherits from $\Phi$, and also a new one. First, $\Psi$ commutes with the rotation $\rho_2$ about the center point of $R$. Second, the partition on the right hand side of Figure 7.3 is a mirror image of the one on the left. (The line of symmetry joins the centers of the red octagons.) Finally, $\Psi$ moves the one partition to its mirror image by way of translations.

## 7.4   Classification of Periodic Orbits

**Lemma 7.1** *$p \in R$ is a periodic point for $\Psi$ if and only if $p$ is a periodic point for $\Phi$.*

**Proof:** We write $\phi_0 = \psi_0 = p$. Let $\phi_n$ be the $n$th iterate of $\phi_0$ under $\Phi$ and let $\psi_n$ be the $n$th iterate of $\psi_0$ under $\Psi$. Since $\Phi$ commutes with $\sigma$, we have

$$\phi_n = \sigma^{d_n}(\psi_n) \tag{33}$$

where $d_n$ counts the number of $k \in \{0, ..., (n-1)\}$ for which $\psi_k$ lies in a blue region in the partition of $R$ (on the left side of Figure 7.3.)

Suppose that $p$ has $\Phi$-period $n$. Then $\phi_n = \phi_0$. This means that $\sigma^{d_n}(\psi_n) = \psi_0$. Since both $\psi_n$ and $\psi_0$ belong to $R$, we must have $d_n$ even. Hence $\psi_n = \psi_0$. This shows that the $p$ is periodic for $\Psi$. On the other hand, suppose that $p$ has $\Psi$-period $n$. Then $\psi_n = \psi_0$. If $d_n$ is even then $\phi_n = \phi_0$. If $d_n$ is odd, then $d_{2n} = 2d_n$ because $\psi_{n+k} = \psi_k$ for each $k = 0, ..., n$. Hence $\phi_{2n} = \phi_0$. Hence $p$ is $\Phi$-periodic. ♠

Figure 7.4 shows the beginning stages of a packing of octagons in $R$. As we will see momentarily, every periodic point of $\Psi$ lies in one of the octagons in this packing.
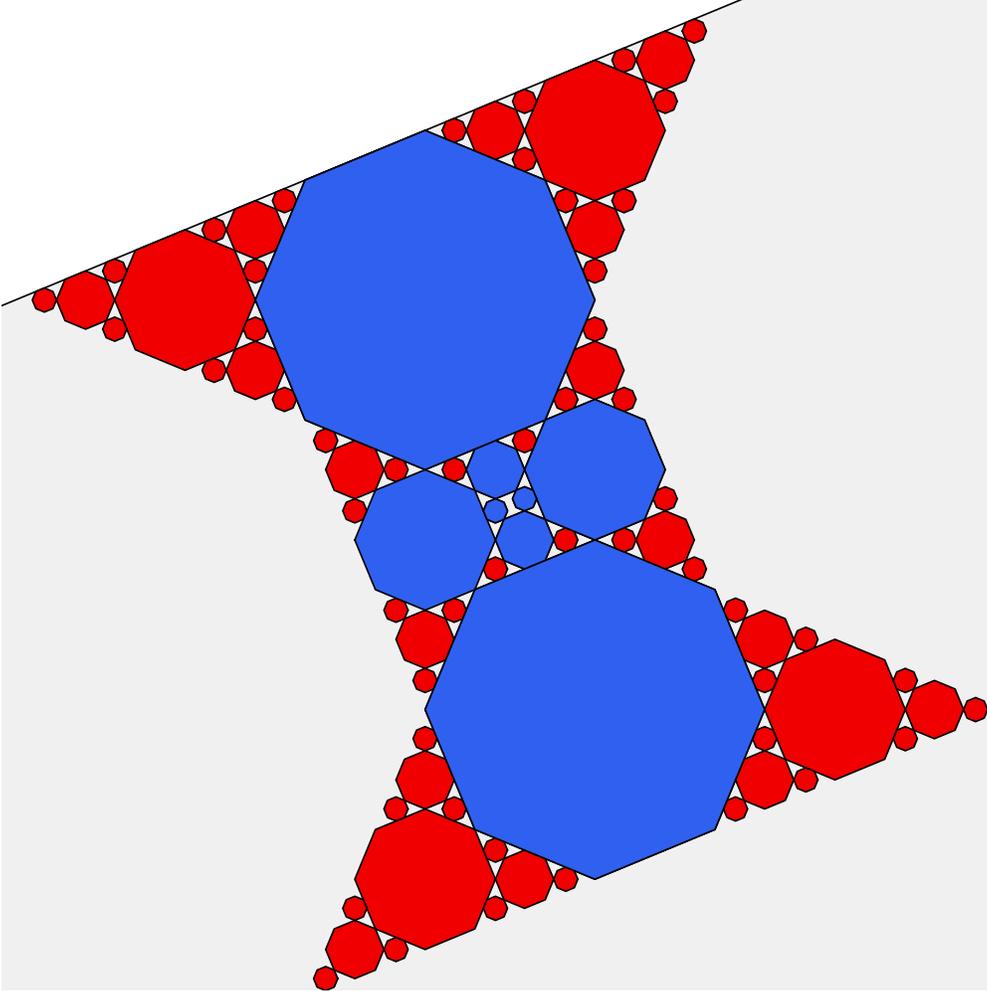
**Figure 7.4:** Image of the central portion

We will establish the result about the packing using a renormalization scheme that is very similar to the one used in the previous chapter. The remormalization scheme is based on the dilation

$$\Theta(z) = \overline{az + b}; \qquad a = (s+1) + i(s+1); \qquad b = -3i(s+1) \qquad (34)$$

We have $\Theta(S) = R$, where $S \subset R$ is the portion of $R$ between the two octagons $O_1, O_2 \subset R$. The left hand side of Figure 7.4 below shows $R$ in blue. We will use the renormalization scheme to classify the periodic points of $\Psi$.

**Lemma 7.2** *Let $O_1^k = \Theta^{-k}(O_1)$ and $O_2^k = \Theta^{-k}(O_2)$. Then*

- *There is a self-similar packing of regular octagons in $S$, each of which is translation equivalent to $O_1^k$ (or $O_2^k$) for some $k$. (See Figure 7.4.)*

- *Every periodic point of $\Psi$ belongs to one of the octagon interiors, and its period is the same as the tile-period of the octagon containing it.*

- *Every octagon in the packing has tile period $3^k$ for some $k = 0, 1, 2, ...,$ and such an octagon either lies in the orbit of $O_1^k$ or the orbit of $O_2^k$. These are the blue octagons in Figure 7.4.*

**Proof:** We see by inspection that $\phi^3(S') \subset S'$. Indeed, the left hand side of Figure 7.5 shows $S$ in blue, the middle picture shows $\Psi(S)$ in blue, and the right hand side shows $\Phi^2(S)$ in blue.
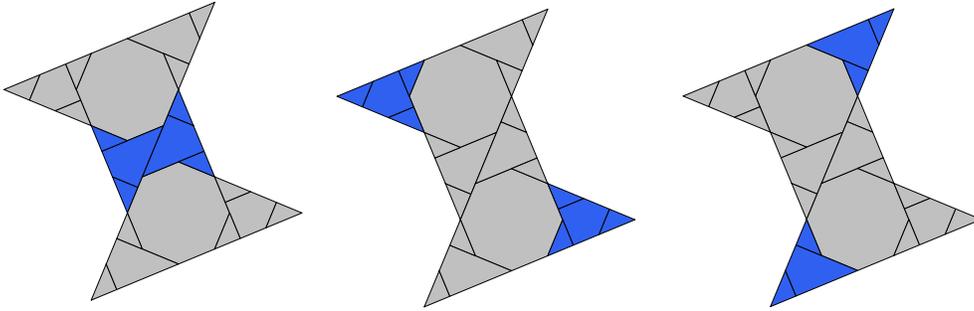


**Figure 7.5:** Image of the central portion

**Lemma 7.3** *Every $\Psi$-periodic point in $R$ has period $3^k$ for some $k = 0, 1, 2...$*

**Proof:** A short calculation establishes the following two facts.

$$\Theta^{-1} \circ \Psi \circ \Theta|_S = \Psi^3; \qquad R - O_1 - O_2 = S \cup \Psi(S) \cup \Psi^2(S). \qquad (35)$$

The same analysis as in the previous chapter produces the packing of octagons, together with the claims about their tile periods. The only new feature of the situation here is that the map $\Psi$ is a piecewise translation, whereas the map from the previous chapter had a rotational component. Since $\Psi$ is a piecewise translation, each point in a octagonal tile has exactly the same period as the tile itself. ♠

## 7.5  Proof of Lemma 5.1

Thanks to the analysis done in §5 it suffices to prove that every periodic point of $R_1$ has $\Phi$-period $3^k$ for some $k = 0, 1, 2, \dots$ We have already worked out the structure of the periodic points of the $\Psi$-system, and we will use this to get the desired information about the periodic points of the $\Phi$-system.

There is a 2 to 1 map $F : R_1 \to R$. which, by construction, carries periodic points of $\Phi$ to periodic points of $\Psi$. Moreover, this map conjugates the action of $\Phi$ on $R^1$ to the action of $\Psi$ on $R$. From this, we see that every $\Phi$-periodic orbit has period either $3^k$ or $2 \times 3^k$.

A parity check rules out the second possibility. Given a $\Psi$-periodic orbit $p_0, p_1, \dots, p_{N-1}$, we let $\sigma_i = 0$ if $p_0$ lies in a blue region of $R$ and other wise $\sigma_i = 1$. We just need to check that $\sum_{i=0}^{N-1} \sigma_i$ is even. We check that the parity of the sum is preserved under the renormalization operation. Since the sum obviously works out for the large octagons – i.e., when $N = 1$, it works out in general.

The parity check is a finite calculation, as we now explain. Suppose we start with a point $z \in R$. Let $w_0 = \Theta^{-1}(R)$. Let $w_j = \Psi^j(w_0)$. We just have to prove that

$$\sigma \equiv \sigma_0 + \sigma_1 + \sigma_2 \mod 2. \tag{36}$$

Given the nature of $\Psi$, we only need to check this equation for 1 point in each of the 22 regions. Here are the results for the first 11 regions

$$
\begin{array}{rcccc}
0: & 0 & 1 & 1 & 1 \\
1: & 0 & 0 & 0 & 0 \\
2: & 1 & 0 & 0 & 1 \\
3: & 0 & 0 & 0 & 0 \\
4: & 1 & 1 & 0 & 0 \\
5: & 0 & 1 & 0 & 1 \\
6: & 1 & 1 & 1 & 1 \\
7: & 1 & 1 & 1 & 1 \\
8: & 1 & 1 & 1 & 1 \\
9: & 1 & 1 & 1 & 1 \\
10: & 0 & 1 & 0 & 1 \\
\end{array}
$$

The pattern repeats exactly for the second 11 regions.

This completes the proof of Lemma 5.1.

# 8 Generating the Arithmetic Graph

## 8.1 The Basic Approach

Let $R_1$ be our basic region, partitioned as on the left hand side of Figure 7.1. To each region $r \in R_1$ we choose and $p \in r$ and define

$$g_k(r) = \pi_k(\widehat{\Phi}(p)). \tag{37}$$

Here $\widehat{\Phi}(p) \in \mathbf{Z}^8$ is the vector from Equation 4. Since $\widehat{\Phi}$ is constant on the region $r$, we see that our definition is independent of the choice of $p \in r$. If $p$ is part of some orbit, then $g_k(p) \in \mathbf{R}^2$ is the side of the edge of $\pi_k(\Gamma)$ that corresponds to $(p, \Phi(p))$.

Given an initial point $p_0 \in R_1$ we consider the forward orbit $p_0, p_1, p_2, ...$ We have a sequence of regions $r_0, r_1, r_2, ...$ and a corresponding sequence $g_k(r_0), g_k(r_1), g_k(r_2), ...$ These vectors are the successive sides of the projection $\pi_k \circ \Gamma(p)$.

## 8.2 The Compressed Approach

Given the nice renormalization scheme we discussed in the previous chapter, we prefer to work with the compressed system $\Psi : R \to R$, as defined in §7.3. The region $R$ is colored as in Figure 7.3.

Recall that $R$ is the top half of $R_1$. Let $R'$ be the bottom half. Let $p_0, p_1, p_2, ...$ be a $\Psi$ orbit in $R$. Let $q_0, q_1, q_2$ be a $\Phi$ orbit in $R_1$ such that $p_0 = q_0$. Let $\sigma_i = 0$ if $p_i$ lies in a blue region of $R$ on the left hand side of Figure 7.3, and $\sigma_i = 1$ otherwise.

We define the *accumulated parity*

$$s_n = \left( \sum_{i=0}^{n-1} \sigma_i \right) \bmod 2. \tag{38}$$

That is, we count the number of times that the orbit lands in a red region on the left hand side of Figure 7.3. The sequence $\{s_n\}$ is a binary sequence that keeps track of whether $p_n = q_n$ or not.

The accumulated parity is the additional information that lets us "lift" an $\Psi$ orbit on $R$ to a $\Phi$ orbit on $R_1$.

**Lemma 8.1** $s_n = 0$ *if and only if* $p_n = q_n$.

**Proof:** This follows from two facts. First, $\sigma_i = 0$ iff $\Phi(q_i)$ and $q_i$ lie in the same half of $R_1$, Second, $\Phi : R_1 \to R_1$ commutes with the map interchanging $R$ and $R'$. ♠

To each region $r \in R$, we associate 4 vectors according to the following rule.

$$g_k(r, 0) = g_k(r); \qquad g_k(r, 1) = g_k(r'); \qquad r' = r - (0, 1) \in R'. \qquad (39)$$

Given the orbit $p_0, p_1, p_2 \in R$ we have the corresponding regions $r_0, r_1, r_2, ...$ and the corresponding accumulated parities $s_0, s_1, s_2, ....$ By construction, $g_k(r_n, s_n)$ is the $n$th edge of $\pi_k \Gamma(p_0)$.

For reference, we list the vector assignments. It is convenient to let $k$ stand for the (region,parity) pair $(k, 0)$ and let $k + 22$ stand for the (region,parity) pair $(k, 1)$. The indices $0, 11, 22, 33$ correspond to the biggest octagons and never arise, so we leave them off [3] We consider the $g_2$ vectors first. Letting $T(x, y) = (-y, -x)$, we have the symmetries

$$g_2(k + 11) = T(g_2(k)); \qquad g_2(k + 22) = -g_2(k); \qquad k = 0, ..., 10. \quad (40)$$

So, the following assignment determines all 44 vectors.

1. $g_2(1) = (6, 2)$.

2. $g_2(2) = (2, -2)$.

3. $g_2(3) = (2, -2)$.

4. $g_2(4) = (-4, 4)$.

5. $g_2(5) = (-6, 2)$.

6. $g_2(6) = (0, -4)$.

7. $g_2(7) = (2, 6)$.

8. $g_2(8) = (4, 4)$.

9. $g_2(9) = (2, 2)$.

10. $g_2(10) = (-2, 6)$.

---

[3]For the record, $g_2(0) = (0, 8)$ and $g_3(0) = (0, 0)$.

In the case of $g_3$ we have the symmetries

$$g_3(k+22) = g_3(k); \qquad g_3(k+11) = -g_3(k) + \sigma_k(0,4); \qquad k = 0, ..., 10. \quad (41)$$

Here $\sigma_k \in \{0,1\}$ is the parity of the $k$th region. We have

$$\sigma_k = 1 \qquad \Longrightarrow \qquad k \in \{2,4,6,7,8,9\}. \quad (42)$$

In light of these symmetries, the following list of 10 vector assignments determines the whole thing.

1. $g_3(1) = (0,-1,-2,-1)$.

2. $g_3(2) = (2,1,0,-1)$.

3. $g_3(3) = (2,1,0,-1)$.

4. $g_3(4) = (-2,-2,2,0)$.

5. $g_3(5) = (-2,-1,0,-1)$.

6. $g_3(6) = (0,0,0,-2)$.

7. $g_3(7) = (-2,-1,4,1)$.

8. $g_3(8) = (-2,-2,2,0)$.

9. $g_3(9) = (-2,-1,0,-1)$.

10. $g_3(10) = (0,-1,2,1)$.

Our notation is as follows.

$$(a,b,c,d) = (a + b\sqrt{2}, c + d\sqrt{2}). \quad (43)$$

In the next section we will explain how these vector assignments combine with a combinatorial substitution rule to generate the graphs.

## 8.3 The Substitution Approach

Suppose that we have some orbit $p_0, p_1, p_2, ....$ We let

$$p_0' = \Theta^{-1}(p_0)$$

and consider the renormalized orbit $p_0', p_1', p_2', ....$ Let $s_j = \sigma_0 + ... + \sigma_{j-1}$ be the accumulated parity of the original orbit. Likewise define $s_j'$. Let $r_j$ be the region of $R$ containing $p_j$, and likewise define $r_j'$.

In §7.5 we explained the sense in which the renormalization scheme respects the accumulated parities. This fact implies that

$$s_{3j}' = s_j \tag{44}$$

Given equation 44, the pair $(s_j, r_j)$ determines the triple

$$(r_{3j}', s_{3j}') \quad (r_{3j+1}', s_{3j+1}') \quad (r_{3j+2}', s_{3j+2}').$$

Recall that we have chosen a particular listing for these 44 possible (region, parity) pairs. What we are saying is that the numerical code for the original orbit determines the numerical code for the renormalized orbit, by way of a substitution scheme in which one number is replaced by 3 numbers. Here is one quarter of the substitution rule.

$$
\begin{array}{rcccc}
0 & \to & 9 & 25 & 39 \\
1 & \to & 10 & 14 & 5 \\
2 & \to & 10 & 12 & 17 \\
3 & \to & 10 & 12 & 16 \\
4 & \to & 9 & 23 & 27 \\
5 & \to & 8 & 36 & 28 \\
6 & \to & 8 & 35 & 17 \\
7 & \to & 9 & 24 & 6 \\
8 & \to & 9 & 26 & 6 \\
9 & \to & 9 & 26 & 7 \\
10 & \to & 9 & 25 & 40 \\
\end{array}
$$

The rest of the pattern can be deduced from symmetry and from what we have already written. We point out two involutions on the set $\{0, ..., 43\}$. The first of these has the action $n \to n \pm 22$. The second one has the action $n \to$

$n \pm 11$, where the + option is taken if $n \in \{0, ..., 10\}$ or $n \in \{0, ..., 10\} + 22$. The substitution rule is invariant under these two involutions. For instance,

$$
\begin{array}{ccccc}
4 & \rightarrow & 9 & 23 & 27 \\
15 & \rightarrow & 20 & 34 & 38 \\
26 & \rightarrow & 31 & 1 & 5 \\
37 & \rightarrow & 42 & 12 & 16
\end{array}
$$

Now we can explain a nice combinatorial way to generate the arithmetic graph projections. Referring to Lemma 7.2, the two orbits $O_1^k$ and $O_2^k$ both have period $3^k$. Here $O_1^0$ and $O_2^0$ respective are the top and bottom biggest octagons in Figure 7.4. We will concentrate on the orbits $O_1^k$. The other half of the orbits have a very similar treatment.

We start with 0, and iterate the substitution rule $k$ times. Then we replace each number by the relevant vector. The resulting list of vectors is the set of edges of the projection of the arithmetic graph. We call the resulting path $G_2(k)$ or $G_3(k)$, depending on whether we use the $G_2$ vectors or the $G_3$ vectors.

Here are two examples. For $G_2(2)$ we make the substitution $0 \rightarrow 9, 25, 39$, and then plug in the $G_2$ vectors:

- $9 \rightarrow (2, 2)$.

- $25 \rightarrow (-2, 2)$.

- $39 \rightarrow (0, -4)$.

We then get the vertices of $G_2(1)$ by accumulating the vectors:

$$(0, 0), (2, 2), (0, 4), (0, 0).$$

So, $G_2(1)$ is a closed path – a triangle.

To generate $G_2(2)$ we go one more step.

$$0 \rightarrow 9, 25, 39 \rightarrow 9, 26, 7, 32, 34, 38, 41, 2, 28.$$

Then we make all the substitutions and accumulate the vectors. We discover in this case that $G_2(2)$ is not closed. We check that $G_2(k)$ is closed if $k = 1, 3, 5, 7$ and open if $k = 2, 4, 6, 8$. In §9.3 we will prove that this pattern persists for all $k$. Our proof there, in particular, implies Statement 1 of the Main Theorem.

# 9 Taking a Better Limit

## 9.1 Subsequential Limits

Note that $G_2(k)$ and $G_3(k)$ have natural parametrizations, coming from the ordering on the edges. Define

$$\Gamma_2(k) = (\sqrt{3})^{-k} G_2(k); \qquad \Gamma_3(k) = (1 + \sqrt{2})^{-k} G_3(k). \qquad (45)$$

We notice that $\Gamma_2(k)$ and $\Gamma_2(k+\beta)$ closely follow each other as parametrized curves when $\beta = 4$. This property does not hold for smaller $\beta$. When $\beta = 2$, the two curves $\Gamma_2(k)$ and $\Gamma_2(k + \beta)$ are close in the Hausdorff metric, but not close as parametrized curves. We also notice that $\Gamma_3(k)$ and $\Gamma_3(k + \beta)$ closely follow each other closely when $\beta = 2$ but not when $\beta = 1$.

Note that $\beta = 4$ is the lowest number that works well for both curve families. This accounts for the dependence on the mod 4 congruence that appears in the Main Theorem, which involves the subsequential limits

$$\lim_{j \to \infty} \Gamma_i(j + 4k); \qquad i = 2, 3; \qquad j = 1, 3. \qquad (46)$$

The 4th power of our substitution rule is our basic object of study. With this rule, each number is replaced by 81 numbers. In geometric terms, each $G_2$ vector $g_2(n)$ is replaced by a polygonal path $g_2'(n)$ of length 81. This path consists of 81 other $G_2$ vectors, scaled down by $\sqrt{3}$. Likewise, each $G_3$ vector $g_3(n)$ is replaced by a polygonal path $g_3'(n)$ of length 81.

Let $Rg_2(n)$ denote the vector that points from the tail of $g_2'(n)$ to the head of $g_2'(n)$. Likewise define $Rg_3(n)$. We notice that $g_i(n)$ and $Rg_i(n)$ are nearly the same vector for each choice $i = 2, 3$ and each $n = 0, ..., 43$. We think of $R$ as a *renormalization operator*, which replaces one vector by another one that is very close to it. It makes sense to define $R^k g_i(n)$ as follows:

- We iterate the substitution rule $4k$ times.

- We replace each number by the appropriate $g_i$ vector.

- We rescale by $(s_i)^{-k}$, where $s_2 = \sqrt{3}$ and $s_3 = 1 + \sqrt{2}$.

The key to analyzing our limits is to looked at the fixed point of the renormalization operator.

## 9.2 The Fixed Point of Renormalization

For motivational purposes (only) we mention the Perron-Frobenius Theorem.

**Theorem 9.1 (Perron-Frobenius)** *Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation represented by a matrix with all positive entries. Then $T$ has (up to scale) a unique unit positive eigenvector $u$. Moreover, if $v$ is any positive vector $v$, then we have $v_n \to u$, where $v_n = T^n(v)/\|T^n(v)\|$.*

It turns out that $g_i(n)$ and $Rg_i(n)$ are not in general the same vector. The Perron-Frobenius Theorem suggests to us that we consider the alternate assignment

$$n \to \lambda_i(n) := \lim_{s \to \infty} R^s g_i(n). \qquad (47)$$

Doing things this way is analogous to finding the Perron-Frobenius eigenvector of a transformation $T$ by taking the limit of iterates of $T$ applied to some conveniently chosen initial vector $v$. In case the limit $u$ is expected to have nice properties – e.g. algebraic coordinates – one might hope to guess $u$ by looking at the decimal expansions of a fairly high iterate. Once we have made a guess, it is trivial to verify that it works. This is what we did.

Our new assignment for $G_2$ has the same symmeties as in Equation 40, so the 10 vectors we list suffices to describe the whole thing.

1. $\lambda_2(1) =^* (6, 2)$.

2. $\lambda_2(2) = (3, -3)$.

3. $\lambda_2(3) =^* (2, -2)$.

4. $\lambda_2(4) = (-3, 3)$.

5. $\lambda_2(5) =^* (-6, 2)$.

6. $\lambda_2(6) = (-3, -1)$.

7. $\lambda_2(7) = (3, 5)$.

8. $\lambda_2(8) = (5, 3)$.

9. $\lambda_2(9) = (3, 1)$.

10. $\lambda_2(10) =^* (-2, 6)$.

This assignment has the same symmetries as the original one, and these symmetries let one deduce the remaining vectors in the assignment. The starred equalities indicate that no change has been made. Curiously, we have

$$\text{new} = \text{old} + \epsilon(1, -1); \qquad \epsilon \in \{-1, 0, 1\}. \tag{48}$$

In the case of $G_3$, the new assignment has nicer symmetries than the original namely

$$\lambda_3(k) = -\lambda_3(k + 11) = \lambda_3(k, 1). \tag{49}$$

That is, the factor of $\sigma_k$ is replaced by 0 in Equation 41. In light of these symmetries, our 10 vectors below determine the whole thing.

1. $\lambda_3(1) =^* (0, -1, -2, -1)$.

2. $\lambda_3(2) = (2, 1, -2, -1)$.

3. $\lambda_3(3) =^* (2, 1, 0, -1)$.

4. $\lambda_3(4) = (-2, -2, 0, 0)$.

5. $\lambda_3(5) =^* (-2, -1, 0, -1)$.

6. $\lambda_3(6) = (0, 0, -2, -2)$.

7. $\lambda_3(7) = (-2, -1, 2, 1)$.

8. $\lambda_3(8) = (-2, -2, 0, 0)$.

9. $\lambda_3(9) = (-2, -1, -2, -1)$.

10. $\lambda_3(10) =^* (0, -1, 2, 1)$.

Again, the starred equalities indicate the ones where there is no change. This time we have

$$\text{new} = \text{old} + \epsilon(0, 2); \qquad \epsilon \in \{-1, 0, 1\}. \tag{50}$$

A direct calculation shows that

$$R\lambda_i(n) = \lambda_i(n); \qquad i = 2, 3; \qquad n = 1, ..., 43. \tag{51}$$

As usual, we are omitting $n = 0, 11, 22, 33$.

Let $\Lambda_i(k)$ denote the version of $\Gamma_i(k)$ produced by the new vector assignments. Our new assignments are nicer, but we need to prove that the $\Lambda$ paths have the same limits as the $\Gamma$ paths. Once we know this, we can throw out the arithmetic graphs, and work with the much nicer $\Lambda$ paths. The rest of the chapter is devoted to analyzing and equating the relevant limits.

Figure 9.2 shows $\Gamma_3(3)$ and $\Gamma_3(7)$. Notice that these curves follow each other pretty well, but that a few of the vertices of $\Gamma_3(7)$ do not lie on $\Gamma_3(3)$. Figure 9.3 shows the improved $\Lambda_3(3)$ and $\Lambda_3(7)$. These curves are better related to each other. Note also that $\Gamma_3(7)$ and $\Lambda_3(7)$ are really just about the same curve.
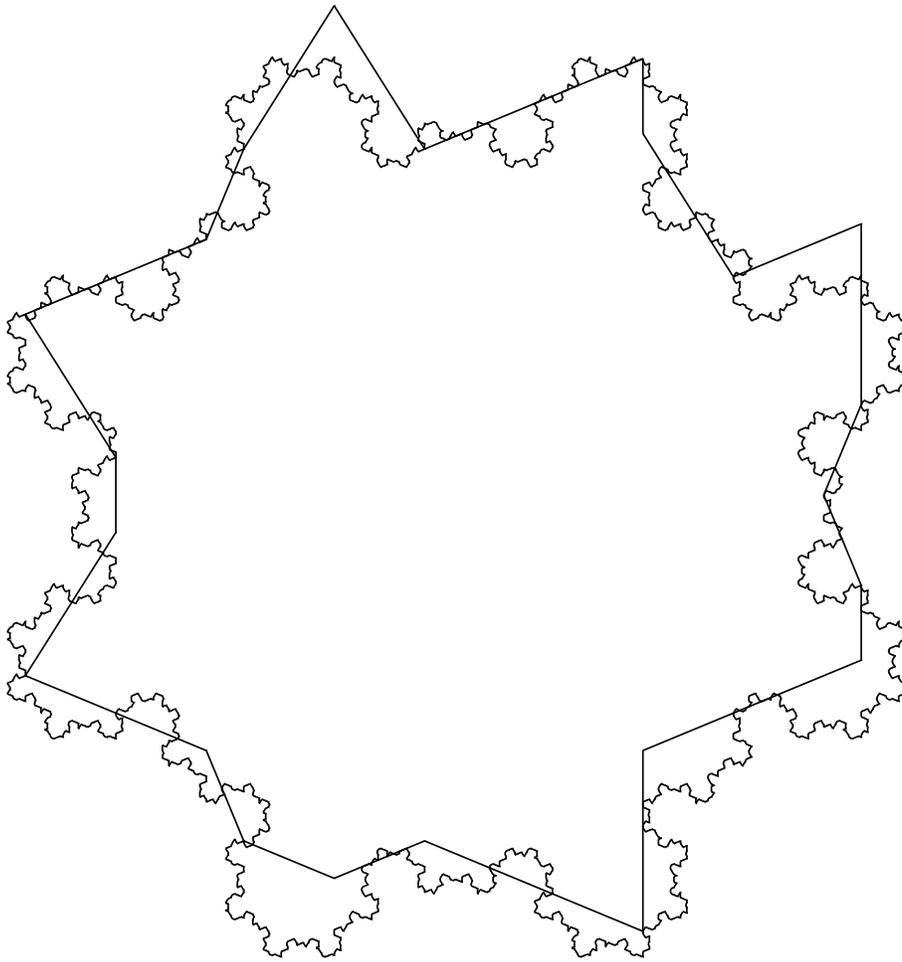


**Figure 9.2:** $\Gamma_3(3)$ and $\Gamma_3(7)$.
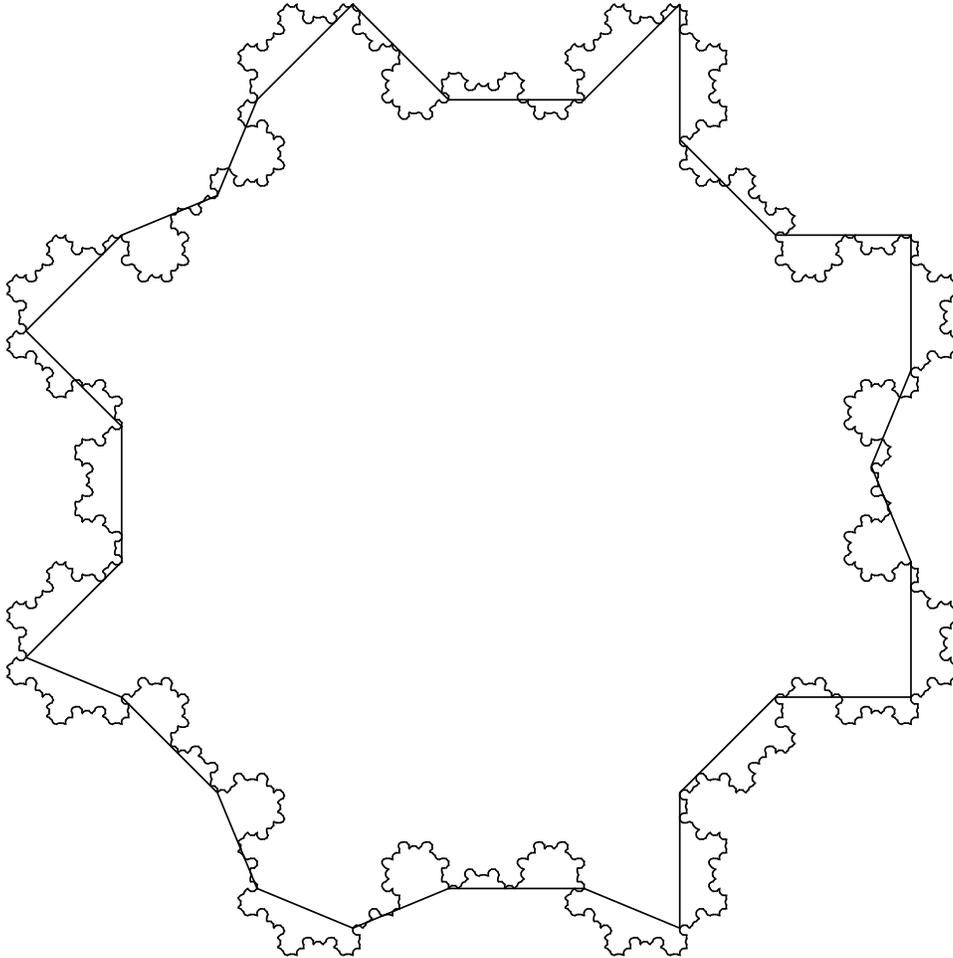
**Figure 9.3:** $\Lambda_3(3)$ and $\Lambda_3(7)$.

## 9.3   Proof of Statement 1

Statement 1 of the Main Theorem, which only deals with the odd case, follows from the more complete result we now prove.

**Lemma 9.2** *The following is true.*

- $G_3(k)$ *is a closed polygon for all $k = 1, 2, 3, ...$*

- $G_2(k)$ *is a closed polygon if and only if $k$ is odd.*

**Proof:** Let $L_k$ be the unscaled version of $\Lambda_k$. The main idea in our proof is showing that $G_k$ and $L_k$ have the same endpoints. Assuming this result, it suffices to prove our result for the $L$ curves. But the $\Lambda$ curves are just scaled version of the $L$ curves. So, it suffices to prove our result for the $\Lambda$ curves. Given the endpoint-preserving property of our ideal substitution rule – see Equation 51 – we see that $\Lambda_i(k)$ is closed if and only if $\Lambda_i(k+4)$ is closed. We check the conclusion of this lemma for the first few cases, and then apply the induction step to take care of the remaining cases.

Now we come to the interesting part of the proof, showing that each $G$ curve has the same endpoints as the corresponding $L$ curve. We treat $G_2$ and $L_2$ first. All paths start at the origin, so we just have to see that the other endpoints match. That is, we need to prove that

$$\sum_{i=0}^{N-1} \delta(n_i) = 0; \qquad \delta(n_i) = g_2(n_i) - \lambda_2(n_i). \tag{52}$$

Here $\{n_i\}$ is the numerical code for the relevant pair of paths, and $N$ is the total combinatorial length.

As we remarked above – see Equation 48 – we always have $\delta(n) = \epsilon(1, -1)$ for $\epsilon \in \{-1, 0, 1\}$. We check by hand the following 44 equations.

$$\delta(n) = \delta(m_1) + \delta(m_2) + \delta(m_3). \tag{53}$$

Here $(m_1, m_2, m_3)$ is the substitution for $n$. Equation 53 implies that the sum in Equation 52 is invariant under substitution. We check that the sum is 0 in the first 4 cases, and then we see by induction that it is always 0.

The proof for $G_3$ and $L_3$ works in the same way. This time, we have $\delta(n) = \epsilon(0, 2)$ for $\epsilon \in \{-1, 0, 1\}$. ♠

## 9.4   Proof of Statement 2

Let $G$ and $L$ be the unscaled versions of $\Gamma$ and $\Lambda$ respectively, as in the proof of Lemma 9.2. Say that a *distinguished strand* of $G$ is one that comes from repeated substitution applied to a single edge. Likewise define distinguished strands of $L$. Such strands consist of $3^k$ segments. Any distinguished strand $\sigma$ determines a vector $V(\sigma)$ that points from the starting point to the ending point.

**Lemma 9.3** *Let $g$ and $l$ be corresponding distinguished strands of $G(n)$ and $L(n)$. Then $V(g)$ and $V(l)$ differ by at most $2$ units.*

**Proof:** Let $g'$ and $l'$ be the distinguished strands of the paths $G(n-1)$ and $L(n-1)$ which give rise to $g$ and $l$, respectively, upon substitution. The same argument as in Lemma 9.2 shows that

$$V(l') - V(g') = V(l) - V(g).$$

By induction, we are reduced to the case of single edges. Now we apply Equations 48 and 50. ♠

**Corollary 9.4** *The distance between any corresponding vertices of $G_i(n)$ and $L_i(n)$ is at most $4\log_3(n)$.*

**Proof:** Both $G_i(n)$ and $L_i(n)$ start at $(0,0)$. Suppose we want to consider the $k$th vertices. Let $\gamma$ be the arc of $G_i$ connecting $(0,0)$ to the $k$th vertex. Likewise define $\lambda$.

Using the base 3 expansion of $k$ as a guide, we can decompose $\gamma$ and $\lambda$ into at most $2\log_3(n)$ distinguished arcs. Our lemma now follows from the previous result and the triangle inequality. ♠

**Lemma 9.5** *Let $i$ be $2$ or $3$. The sequences $\{\Lambda_i(1+4k)\}$ and $\{\Lambda_i(3+4k)\}$, considered as a sequence of parametrized curves, converge in the $L_\infty$ norm as $k \to \infty$.*

**Proof:** We will consider the first of these sequences. The second one has the same proof. The edges of $\Lambda_i(4k+5)$ are obtained by replacing each edge of $\Lambda(4k+1)$ by a length 81 polygonal path that has the same endpoint. Define

$$D_k = \sup_{s \in [0,1]} \text{dist}\Big(\Lambda_i(4k+1;s), \Lambda_i(4k+5;s).\Big). \tag{54}$$

Each edge of $\Lambda(4k+5)$ is

$$1/9 = (\sqrt{3})^{-4}$$

71

times as long as an edge of the same type in $\Lambda(4k + 1)$. By similarity, we have

$$D_{k+1} = \frac{1}{\sqrt{3}} D_k, \tag{55}$$

once $k$ is large enough that $\Lambda_i(1 + 4k)$ contains all the edge types. It suffices to take $k = 1$, in fact.

Equation 55 implies that the distance between two successive curves in our sequence decreases exponentially. Hence, our curves form a Cauchy sequence in the space $\text{Map}(S^1, \boldsymbol{R}^2)$, equipped with the $L_\infty$ topology. Hence, our sequence converges. ♠

The following result implies Statement 2 of the Main Theorem.

**Lemma 9.6** *The sequences $\{\Gamma_i(1 + 4k)\}$ and $\{\Gamma_i(3 + 4k)\}$, considered as a sequence of parametrized curves, converge in the $L_\infty$ norm as $k \to \infty$. The limits in this case coincide with the limits in Lemma 9.5.*

**Proof:** We prove this first for $i = 2$. The $i = 3$ case is essentially the same. By Corollary 9.4, the distance between corresponding points of $G_2(1+4k)$ and $L_2(1 + 4k)$ is at most $4\log(8k)$. But then, by scaling, the distance between corresponding points of $\Gamma_2(1+4k)$ and $\Lambda_2(1+4k)$ is at most $4\log_3(8k)(\sqrt{3})^{-k}$, a quantity that tends to 0 exponentially fast. ♠

# 10 The End of the Proof

## 10.1 Scaling Constants

In this chapter, we prove Statements 3, 4, and 5 of the Main Theorem.

Let $I_2$ and $I_3$ be suitably scaled versions of the sets discussed in §4. For $i = 2, 3$ and $j = 1, 3$ let $\Lambda_{ij}$ be the limit of the sequence $\{\Lambda_i(j + 4k)\}$. To finish the proof of the Main Theorem, we need to show that

$$\Lambda_{21} = \Lambda_{23} = I_2; \qquad \Lambda_{31} = \Lambda_{33} = I_3. \qquad (56)$$

Our first order of business is to explain how to scale $I_2$ and $I_3$.
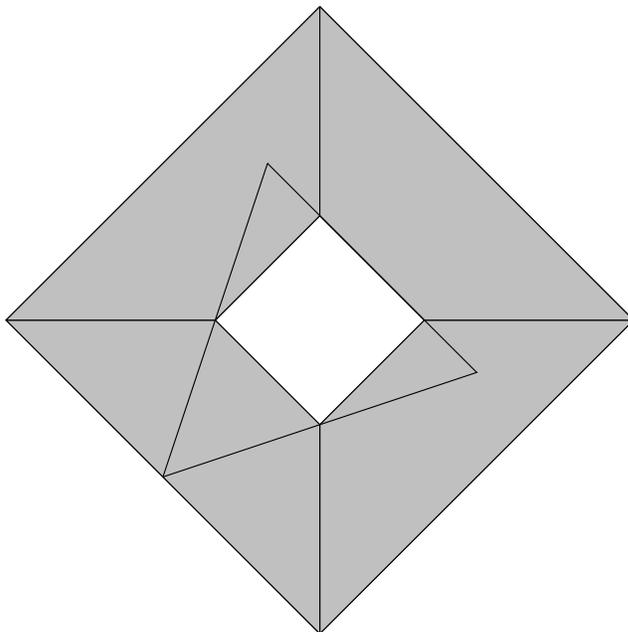


**Figure 10.1:** $I_2$ seed and $\Lambda_2(1)$.

We will deal with $I_2$ first. Figure 10.1 shows the initial seed for $I_2$, scaled so that it is well situated with respect to $\Lambda_2(1)$. By construction, $\Lambda_2(1)$ is an isosceles triangle with vertices $(0,0)$ and $(3,1)$ and $(1,3)$. Given the way that the $I_2$ seed sits with respect to this triangle, we deduce that the center of $I_2$ is $(3/2, 3/2)$ and the left corner is $(-3/2, 3/2)$. This gives us the copy of $I_2$ mentioned in the Main Theorem.

Now we deal with $I_3$. Figure 10.2 shows $\Lambda_2(1)$ and the union $I_3(1)$ of triangles obtained by applying the substitution operator to the $I_3$ seed.
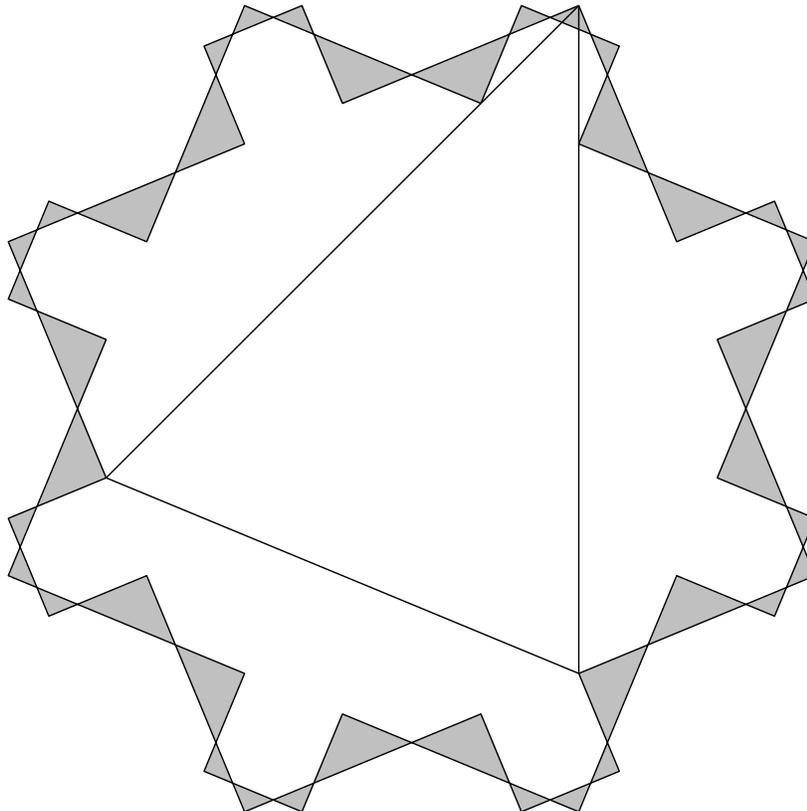


**Figure 10.1:** $I_3(1)$ and $\Lambda_3(1)$.

$\Lambda_3(1)$ is an isosceles triangle, with vertices

$$(0,0); \qquad (-2-\sqrt{2}, -2-\sqrt{2}) \qquad (0, -2-2\sqrt{2}). \qquad (57)$$

The point $(0,0)$ is the apex of the triangle, the top vertex.

This is enough to determine the placement of $I_3$. Here is the nicest way we can think of describing the scaling. Let $s = 1 + \sqrt{2}$. This is the scaling factor that came up in the previous chapter. The $I_3$ seed consists in 8 isosceles triangles. One of the isosceles triangles has vertices $(0,0)$, $(1/2, -s/2)$ and $(-s/2, -1/2)$. This gives us the version of $I_3$ mentioned in the Main Theorem. Note that all the vertices of $I_3$, as we have scaled it, lie in $\boldsymbol{Q}[\sqrt{2}]$.

## 10.2 Proof of Statement 3

Let $I_2$ and $I_3$ be scaled as in the previous section. By Equation 51, we see that $\Lambda_{ij}$ contains the vertices of all the polygons of which it is a limit. Moreover, by scaling, the vertices of the approximating polygons become dense in $\Lambda_{ij}$ as the number of sides tends to $\infty$. We passed to the limit of renormalization precisely to arrange these two properties. We want to understand the sets in Equation 56 just from the placement various finite collections of vertices.

Given the properties enjoyed by the vertices, we can establish Equation 56 simply by showing, for $i = 1, 2$ and for all odd $j$ that the vertices of the polygons approximating $\Lambda_{ij}$ are contained in $I_i$ and become dense in $I_i$ as $j \to \infty$.

We will deal with $i = 3$ first. Figure 10.3 shows a piece of the $I_3(3)$, together with a piece of $\Lambda_3(3)$. Notice that all the vertices of $\Lambda_3(3)$ in sight coincide with right-angled vertices of $I_3(3)$. Notice also that there seems to be (most of) a smaller copy of Figure 10.1 sitting near the top of Figure 10.3.
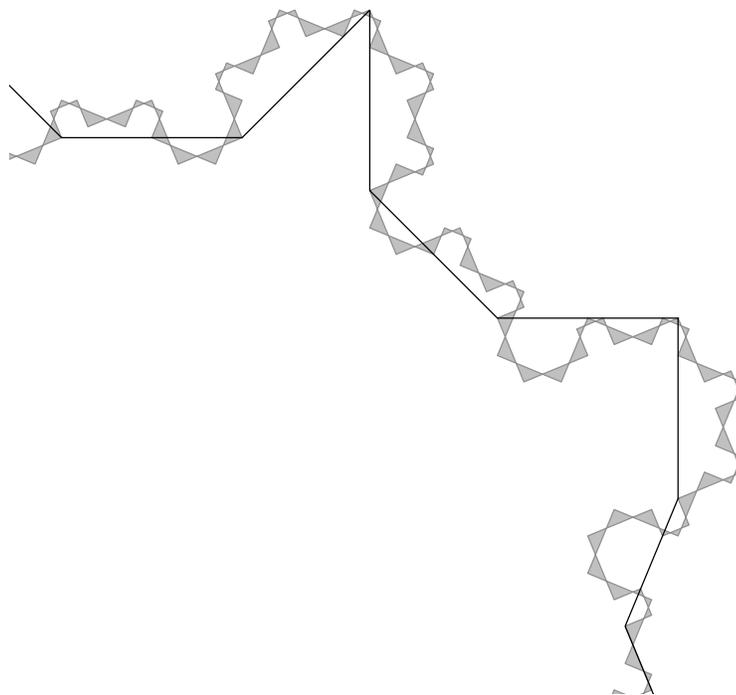


**Figure 10.3:** $I_3(1)$ and $\Lambda_3(1)$.

Say that a *special vertex* of $I_3$ is a vertex of a right-angled triangle of some $I_3(n)$. We check by direct (computer) calculation that every vertex of $\Lambda_3(j)$ is a special vertex of $I_3$ for $j = 1, 3, 5, 7, 9$. This is an easy calculation that takes place entirely inside $\mathbf{Q}[\sqrt{2}]$.

We can't make an infinite number of explicit calculations to deal with each successive polygon, so we describe a method, based on self-similarity, that proves everything we need from a finite number of calculations. For ease of exposition, we will work entirely with the sequence $\Lambda_3(1 + 4k)$. The sequence $\Lambda_3(3 + 4k)$ has a similar treatment. It is convenient to set $\Lambda_3 = \Lambda_{31}$. So, $\Lambda_3$ is the limit of the sequence we are considering.

The polygon $\Lambda_3(1)$ is just a triangle, and too small to be of any use to us (beyond what we have already done with it above.) So, we consider the next polygon in the sequence. Each edge of $\Lambda_3(5)$ is a scaled copy of some $\lambda_3(n)$. We call $n$ the *type* of the edge. The type belongs to the set $S = \{1, ..., 43\} - \{11, 22, 33\}$. Inspecting the numerical sequence associated to $\Lambda_3(5)$ – see §8.3 – we check that $\Lambda_3(5)$ has every possible type edge. We check the same thing for $\Lambda_3(9)$.

Suppose that $A \subset \Lambda_3(5)$ is an edge of type $n$ and $B \subset \Lambda_3(9)$ is an edge of type $n$. The pair $(A, B)$ determines a homothety

$$H = H(A, B) : A \to B. \tag{58}$$

Given the way $\Lambda_3$ is the limit of our sequence, we have

$$H : \Lambda_3(A) \to \Lambda_3(B). \tag{59}$$

Here $\Lambda_3(A)$ is the portion of $\Lambda_3$ between the endpoints of $A$, and likewise for $\Lambda_3(B)$. So, $H$ is a homothety that carries a small portion of $\Lambda_3$ to an even smaller portion. We call $H$ a *special homothety* associated to the pair $(A, B)$. Every such pair gives rise to a special homothety. We call $H$ *good* if

$$H : I_3(A) \to I_3(B). \tag{60}$$

Here $I_3(A)$ is the portion of $I_3$ bounded by the endpoints of $A$. Likewise we define $I_3(B)$. This definition makes sense because we have shown that the endpoints of $A$ and $B$ both lie in $I_3$. Just to be clear, we mean that $H$ sets up a bijection between $I_3(A)$ and $I_3(B)$ that maps endpoints to endpoints.

Below, we will prove the following result.

**Lemma 10.1 (Good Homothety)** *Every special homothety is good.*

The Good Homothety Lemma involves a finite number of special homotheties. We will show in the next section how one verifies that a special homothety is good using a finite calculation. Indeed, one can practically see this by inspection using our program OctoMap 2.

Before we prove the Good Homothety Lemma, let us explain what it does for us. Suppose we want to show that the vertices of $\Lambda_3(13)$ lie in $I_3$. Let $v$ be a vertex of $\Lambda_3(13)$. If $v$ is also a vertex of $\Lambda_3(9)$, we are already finished. Otherwise, there is some edge $B$ of $\Lambda_3(9)$ such that $v$ is one of the vertices associated to the replacement of $B$. That is, $v$ lies on the polygonal path of $\Lambda_3(13)$ connecting the endpoints of $B$. Choose and edge $A$ of $\Lambda_3(5)$ that has the same type as $B$.

The special homothety $H = H(A, B)$ carries $A$ to $B$. Moreover, $H$ carries $\Lambda_3(9; A)$ to $\Lambda_3(13; B)$. Here $\Lambda_3(9; A)$ denotes the portion of $\Lambda_3(9)$ subtended by $A$. At the same time as this, $H$ carries $I_3(A)$ to $I_3(B)$. But $H^{-1}(v)$ is a vertex of $\Lambda_3(9; A)$. Hence $H^{-1}(v) \in I_3(A)$. Hence $v \in I_3(B)$. This proves that all vertices of $\Lambda_3(13)$ lie in $I_3$.

Continuing in this way, we see that all vertices of $\Lambda_3(1+4k)$ lie in $I_3$ for all positive $k$. To take care of density, we let $d_k$ denote the maximum distance from a point of $I_3$ to a vertex of $\Lambda_3(1 + 4k)$. Since our special homotheties scale distances by a factor of $\sqrt{2} - 1$, we see that

$$d_{k+1} = (\sqrt{2} - 1)d_k. \tag{61}$$

In particular $d_k \to 0$ as $k \to \infty$. This proves that the vertices of $\Lambda_3(1 + 4k)$ becomes dense in $I_3$ as $k \to \infty$.

In short, the Good Homothety Lemma implies Statement 3 of the Main Theorem.

## 10.3   Proof of the Good Homothety Lemma

In §4.2 we described the basic substitution rule for the snowflake, in which each isosceles triangle is replaced by 5 smaller ones. We started with a seed, consisting 8 isosceles triangles arranged in a cyclic pattern, and then iteratively replaced each triangle by the smaller ones.

Here we describe a more conservative approach to the construction. Let $L_1$ and $L_2$ be finite cyclically ordered lists of isosceles triangles. We write $L_1 \to L_2$ if one obtains $L_2$ from $L_1$ by replacing one of the triangles of $L_2$ by the 5 smaller ones from the substition rule and adjusting the cyclic ordering

in the obvious way. We call a list $L$ of triangles *admissible* if we have a finite chain $I_3(0) = L_0 \to ... \to L_n$. The specific lists $I_3(n)$, produced in §4.2 are examples of admissible lists.

We say that a *special vertex* of $I_3$ is a right-angled vertex of some $I_3(n)$. An equivalent definition is that a special vertex of $I_3$ is a right-angled vertex of some admissible list. Let $\{v_k\}$ be a finite list of special vertex. We say that an admissible list $L$ is *compatible* with $\{v_k\}$ if each point $v_k$ is a vertex of some triangle in $L$. We say that $L$ is *the minimal pattern* for $\{v_k\}$ if $L$ is compatible with $\{v_k\}$ and there is no $L'$ such that $L' \to L$ and $L'$ is compatible with $\{v_k\}$.

**Remark:** For our proof of the Good Homotopy Lemma, we do not need to know that the minimal pattern associated to a finite list of vertices is unique, but we emphasize that the minimial pattern is unique. The uniqueness comes from the tree-like nature of the subdivision rule. We simply start with the seed and replace triangles only when necessary. After a finite number of steps we arrive at the minimial pattern.

Let $A$ be a line segment whose endpoints are special vertices of $I_3$. We let $J_3'(A)$ be the minimal pattern associated to the endpoints of $A$. We don't care so much about the whole pattern, but only the portion related to $A$. We let $J_3(A)$ denote the portion of $J_3'(A)$ that lies between the two endpoints of $A$. Figure 10.4 shows the minimal pattern $J_3(A)$ when $A$ is a certain type-1 edge of $\Lambda_3(5)$.
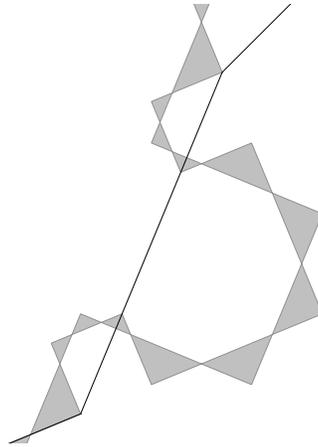


**Figure 10.4:** $J_3(A)$ for $A$ a type 1 edge of $\Lambda_3(5)$.

The figure we have shown is completely representative of the local picture

for the type 1 edges of $\Lambda_3(5)$. The associated minimal patterns are all the same. This is not really so surprising, because these vectors all are translates of each other. We also check that the local picture looks the same for all type 1 edges of $\Lambda_3(9)$. That is, the picture looks the same up to a scale factor of $\sqrt{2}-1$.

So, suppose that $A$ is a type 1 edge of $\Lambda_3(5)$ and $B$ is a type 1 edge of $\Lambda_3(9)$. Then $H(A,B)$ carries $J_3(A)$ to $J_3(B)$ because the patterns are identical up to scale. But $J_3(A)$ and $J_3(B)$ respectively determine $I_3(A)$ and $I_3(B)$. Hence $H$ carries $I_3(A)$ to $I_3(B)$, as desired. This proves the Good Homotopy Lemma for edges of type 1. The key idea is that the minimal pattern $J_3(A)$ only depends on the type of $A$, up to scale.
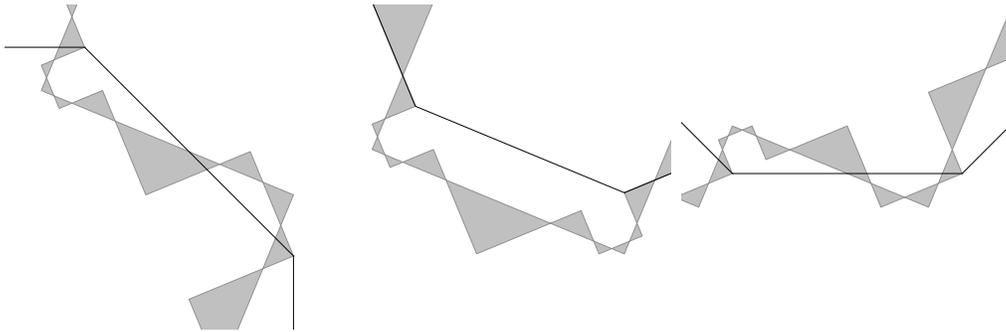


**Figure 10.5:** The minimal patterns associated to types $2, 3, 4$.

Figure 10.5 shows the minimal patterns associated to the types 2, 3, and 4. All the edge types except those of the form $6 + 11k$ and $9 + 11k$ have the same property we have already discussed. The local picture is the same, independent of the edge. By inspection, we see that the Good Homotopy Lemma holds for for all edge types except the 8 types we have mentioned.

For each of the remaining edge types, there are two local pictures. up to rotation, these two local pictures are the same for each of the edge types. Figure 10.6 shows the two pictures associated to the edges of type $6 + 11k$. The edges are vertical.
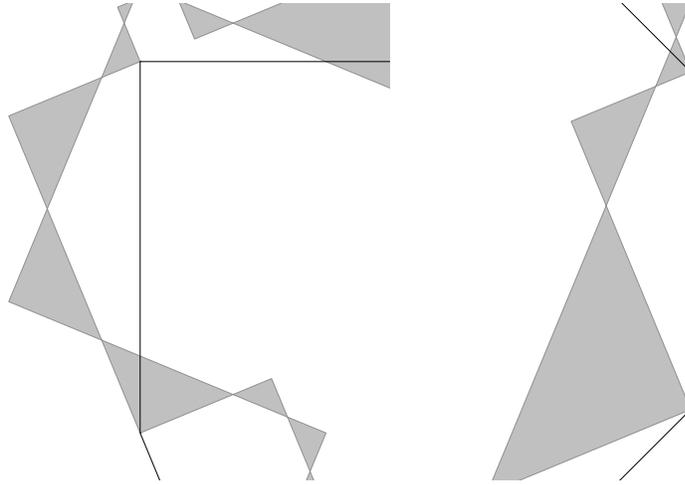
**Figure 10.6:** The minimal patterns associated to the edges of type $6 + 11k$.

Let $A$ and $A'$ be the two type 6 edges shown in Figure 10.6. The pattern $J_3(A)$ consists of 4 triangles $T_1 \cup T_2 \cup T_3 \cup T_4$ and $J_3(A')$ consists of 3 triangles $T_1' \cup T_2' \cup T_3'$. We list these triangles top to bottom. There is a translation $\tau$ such that $\tau(A) = A'$. To finish our proof for edges of type $6 + 11k$, we just have to see that $\tau$ carries $I_3(A)$ to $I_3(A')$. Note that $\tau$ carries $T_j'$ to $T_j$ for $j = 1, 2$. So, things work out as desired for the "top portions" of the sets of interest to us.

Now we consider the bottom portions. Let $S$ be the line segment on the left hand side of Figure 10.6 that connects the bottom vertex of $A$ to the bottom vertex of $T_2$. Note that $S$ is made from the hypotenuse of $T_3$ and a short side of $T_4$. Let $S'$ be the edge of $T_3$ that connects the bottom of $A'$ to a vertex of $T_2'$. We see by direct calculation that $\tau(S) = S'$. The point is that $S$ and $S'$ have the same slope and length. We just have to prove that $\tau$ carries $I_3(S)$ to $I_3(S')$. But we have exactly the situation discussed in §4.3, and the desired result here follows from Lemma 4.1.

it remains to consider The edges of type $9 + 11k$. It turns out that the minimal patterns associated to these edges look exactly like the ones associated to the edges of type $6 + 11k$, except that the picture has been rotated 45 degrees. The same argument as in the previous case works here.

This completes the proof of the Good Homothety Theorem. Again, the Good Homothety Theorem (and our previous work) implies Statement 3 of the Main Theorem

## 10.4   Proof of Statement 4

In considering $I_2$ we make all the same definitions that we made for $I_3$, except that we redefine what we mean by a special point. In the case of $I_3$, the special points are the right-angled vertices of the triangles. In the case of $I_2$, we call a point *special* if it has one of two properties:

- It is the center point of a parallelogram of some admissible list.

- It is the center of the long edge of a trapezoid of some admissible list.

We check directly that every vertex of $I_2(j)$ is a special point, for the initial values $j = 1, 3, 5, 7, 9$. This check entirely involves calculations with rational points. We want to extend this result for all odd $j$. As in the $\Lambda_3$ case, we ignore the sequence $\Lambda(4k + 3)$ and concentrate on the sequence $\Lambda(4k + 1)$. The other sequence has essentially the same treatment.

To finish the proof, we just have to prove the analogue of the Good Homothety Lemma for $I_2$. Given the combinatorially identical subdivision rules that produce $I_2$ and $I_3$, it makes sense to say that a minimal pattern associated to $I_2$ is isomorphic to a minimal pattern associated to $I_3$: The canonical bijection between the $I_3$-shapes and the $I_3$-shapes carries the one pattern to the other. For $j = 1, 3, 5, 7, 9$ we make the following observation: The minimial pattern associated to some edge of $\Lambda_2(j)$ is isomorphic to the minimal pattern associated to the corresponding edge of $\Lambda_3(j)$. There is a perfect correspondence between the two kinds of patterns.

Given the combinatorial isomorphism between the $I_3$-patterns and the $I_2$-patterns, the proof of the Good Homothety Lemma goes through, word for word for all the edges except those of type $6 + 11k$ and $9 + 11k$. In these cases, there are two combinatorial types of minimial pattern, and we must make a separarate analysis. Figure 10.7 shows the relevant portions of the minimal patterns associated to edges of type $6 + 11k$.
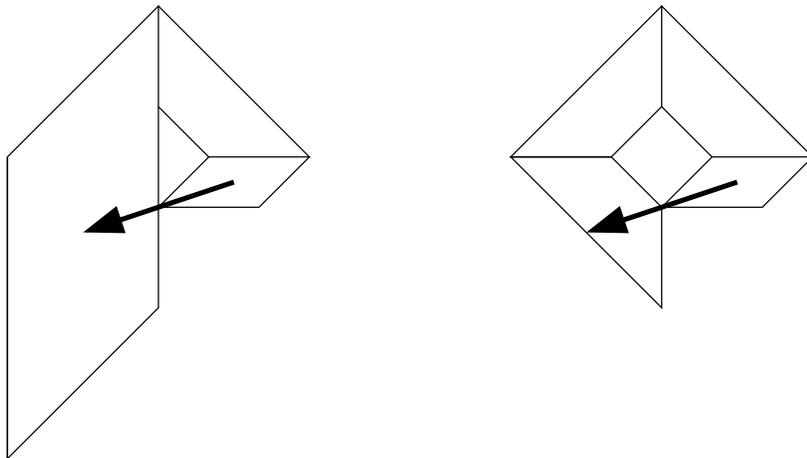
**Figure 10.7:** The minimal patterns associated to the edges of type $6 + 11k$.

On the left hand side we have a chain of 3 quadrilaterals and on the right hand side we have a chain of 4 quadrilaterals. The first two pieces agree on both sides. When we chop two pieces off, and superimpose what remains of the left and right sides, we get exactly the left hand side of Figure 4.7, the figure we used to illustrate the hidden symmetry of the carpet. The rest of the proof is just like what we did for the $\Lambda_3$ case, except that we use Lemma 4.6in place of Lemma 4.1.

The edges of type $9 + 11k$ have the same treatment. This completes the proof of the Good Homotopy Lemma in the $I_2$ case, and thereby completes the proof of Statement 4 of the Main Theorem.

## 10.5   Proof of Statement 5

Statement 5 takes a bit of unwrapping. We start with an arbitrary sequence $\{x_n\}$ of good points in the half-strip $\Sigma_0^1$, and we let $A_n$ be the arithmetic graph corresponding to $x_n$. We want to understand the map

$$\phi_n : S_{3,n}\pi_3(A_n) \to S_{2,n}\pi_2(A_n) \tag{62}$$

Here $S_{2,n}$ and $S_{3,n}$ are the dilations discussed in the Main Theorem.

In light of the analysis done in the previous chapter, we can assume that

$$S_{k,n}\pi_k(A_n) = \Gamma_k(n). \tag{63}$$

Here $n$ is odd, and $\Gamma_k(n)$ is as in §9.1. For ease of exposition, we now make the blanket assumption that $n \equiv 1 \bmod 4$. The case when $n \equiv 3 \bmod 4$ has the same treatment.

The two paths $\Gamma_2(n)$ and $\Gamma_3(n)$ have the same number of sides, and so there is a canonical map

$$\phi_n : \Gamma_3(n) \to \Gamma_2(n). \tag{64}$$

There is a canonical map from $\Gamma_k(n)$ to $\Lambda_k(n)$, and this gives us a map

$$\psi_n : \Lambda_3(n) \to \Lambda_2(n). \tag{65}$$

Our strategy is to first analyze the sequence $\{\psi_n\}$ of maps and then use Lemma 9.5 to get the desired information about $\{\phi_n\}$.

There is a sense in which $\psi_n$ and $\phi$ are compatible, we we now explain. The map $\phi$ is the limit of a sequence of canonical bijections between the admissible $I_3$-patterns of triangles and the admissible $I_2$-patterns of quadrilaterals. We call this bijection the *shape bijection*. We denote it by $\Phi$. Note that $\Phi$ is canonical only up to composition with isometries of $I_2$ and $I_3$. This accounts for the isometry $F$ that appears in Statement 5 of the Main Theorem.

Given an edge $e_3$ of $\Lambda_3(n)$, there are two triangles $T_1$ and $T_2$ in the minimal pattern $\Lambda_3(e_3)$ whose right-angled vertices are the endpoints of $e_3$. Let

$$e_2 = \psi_n(e_3)$$

be the edge of $\Lambda_2$ corresponding to $e_3$. We say that the pair $(e_3, e_2)$ is *compatible* if the endpoints of $e_2$ are special points of the two shapes $\Phi(T_1)$ and $\Phi(T_2)$. In case $\Phi(T_1)$ is a parallelogram, the relevant endpoint of $e_2$ should be the center point of this parallelogram. In case $\Phi(T_1)$ is a trapezoid, the relevant endpoint of $e_2$ should be the midpoint of the long side of this trapezoid.

We check that all relevant edge-pairs are compatible for $n = 1, 3, 5, 7, 9$. But our proofs of the Good Homotopy Lemma, in the two cases, involved combinatorially identical sets in each case. That is, whatever we did in the $I_3$ case carries over to the $I_2$ case *via* the shape bijection. From this we see, by induction and symmetry, that all relevant edge-pairs are compatible for all odd positive integers $n$.

Given this notation of compatibility, we have

$$\lim_{n \to \infty} \psi_n = F \circ \phi, \tag{66}$$

for some isometry $F$. The convergence takes place in the sense that the graph of $\psi_n$ converges, to the graph of $\phi$ in the Hausdorff topology. In particular, the left hand side of Equation 66 does have a limit. But Lemma 9.6 now says that

$$\lim_{n \to \infty} \phi_n = \lim_{n \to \infty} \psi_n. \tag{67}$$

Putting the two equations together gives us Statement 5 of the Main Theorem.

This completes our proof of the Main Theorem.

# 11　References

[**BC**] N. Bedaride, J. Cassaigne, *Outer Billiards outside Regular Polygons*, preprint (2009)

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

[**G**] A. Graham, *Nonnegative matrices and Applicable Topics in Linear Algebra*, John Wiley and Sons, New York (1987).

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

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

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

[**M2**] J. Moser, *Stable and random motions in dynamical systems, with special emphasis on celestial mechanics*, Annals of Math Studies 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.

[**S1**] R. E. Schwartz, *Unbounded Orbits for Outer Billiards*, J. Mod. Dyn. **3**:371–424 (2007).

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

[**S3**] R. E. Schwartz, *Outer Billiards and the Pinwheel Lemma*, preprint (2009)

[**T1**] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library 30, Amer. Math. Soc. (2005).

[**T2**] S. Tabachnikov, *On the Dual Billiard Problem*, Advances in Mathematics **115** pp 221-249 (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).