On the Affine Shapes of Quadrilaterals in Self-Projective Carpets

Richard Evan Schwartz *

November 24, 2026

Abstract

This introductory paper studies the affine equivalence classes of shapes of quadrilaterals that appear in the simplest projectively natural subdivision processes. The highlight is a complete understanding of what happens in the one-parameter family of projectively natural subdivision rules with 4-fold projective symmetry. Partly, this paper is an excuse to show off some beautiful pictures.

1 Introduction

This paper is about the simplest projectively natural subdivision processes for quadrilaterals in the projective plane. When these rules are iterated, they produce infinite collections of quadrilaterals which I think of (a bit impressionistically) as *solid self-projective carpets*. They are akin to the *self-affine carpets* studied e.g. in [**McM**]. However, here we are not omitting any quadrilaterals. That is what we mean by *solid*.

As a first pass in understanding the shapes of quadrilaterals that arise in these processes, we will be interested in the *affine* shapes. What we mean to do is to consider convex quadrilaterals up to affine transformations. (We will also discuss the case of similarity classes, though in much less detail.) These self-projective carpets are also related to the iteration of Pappus's theorem that I studied in [**S1**] and [**S2**] and the geometry of the corresponding limit sets in the projective flag space. This connection motivated me to study self-projective carpets. However, I will defer the discussion of how all this relates to Pappus's theorem to a sequel paper.

*Supported by N.S.F. Research Grant DMS-2505281

1.1 Subdivision of Quads

Let $\boldsymbol{P}$ denote the real projective plane. We identify the standard affine patch in $\boldsymbol{P}$ with $\boldsymbol{R}^2$. A *convex quad* is a labeled quadrilateral (both the vertices and edges) which is projectively equivalent to a convex quadrilateral in $\boldsymbol{R}^2$. See §2.1 for more details about all these definitions.

Figure 1.1 shows a subdivision of the labeled square quad into 4 smaller quads. We have added an auxiliary labeling which we will use unless otherwise stated. The subdivision scheme is based on the parameters $(u_0, u_1, u_2, u_3) \in (-1, 1)^4$.

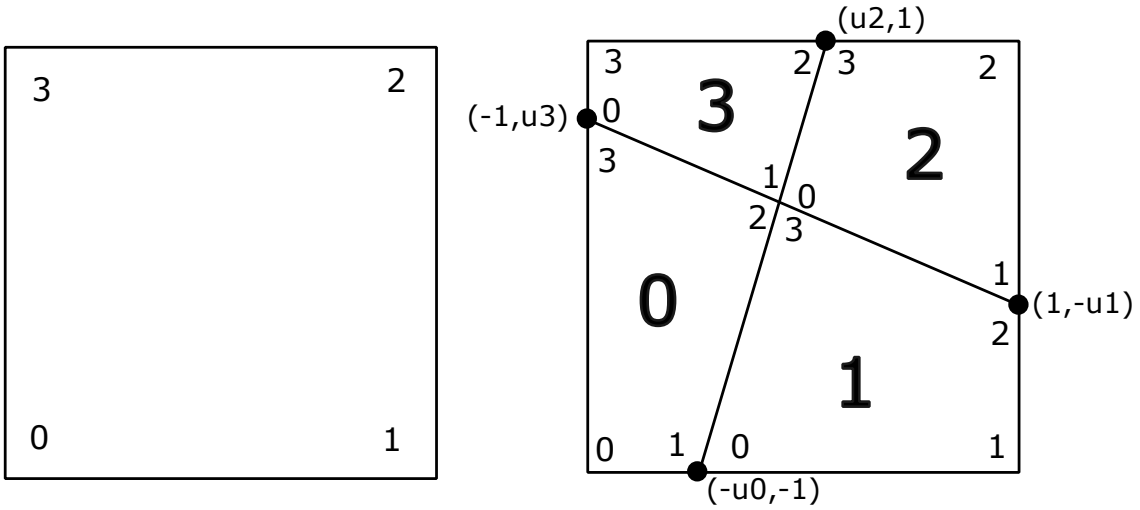


Figure 1.1: The basic subdivision rule

The signs in our parametrization are chosen so that the case $u = (a, a, a, a)$ corresponds to a subdivision rule with (unlabeled) 4-fold rotational symmetry. Figure 1.2 depicts this case. Our subdivision rule, originally just defined for the specific square shown, extends to all labeled convex quads: We map to the square, subdivide, then pull back the subdivision. Thus the 4 parameters (u_0, u_1, u_2, u_3) give a subdivision rule for convex quadrilaterals.

We can now iterate this construction. Each quadrilateral gets a finite 4-ary sequence, called its *address*, which describes the branches of the subdivision that produced it. Figure 1.2 shows the second stage of the subdivision. In a few cases we have indicated how the second (right-most) digit is determined. Each quad shares one vertex with its parent, and the label of this vertex, which is the same for the quad and its parent, determines the last (right-most) digit. These pictures are semi-accurate hand-drawn approximations. Subsequent pictures will be accurate pictures drawn from my computer program.

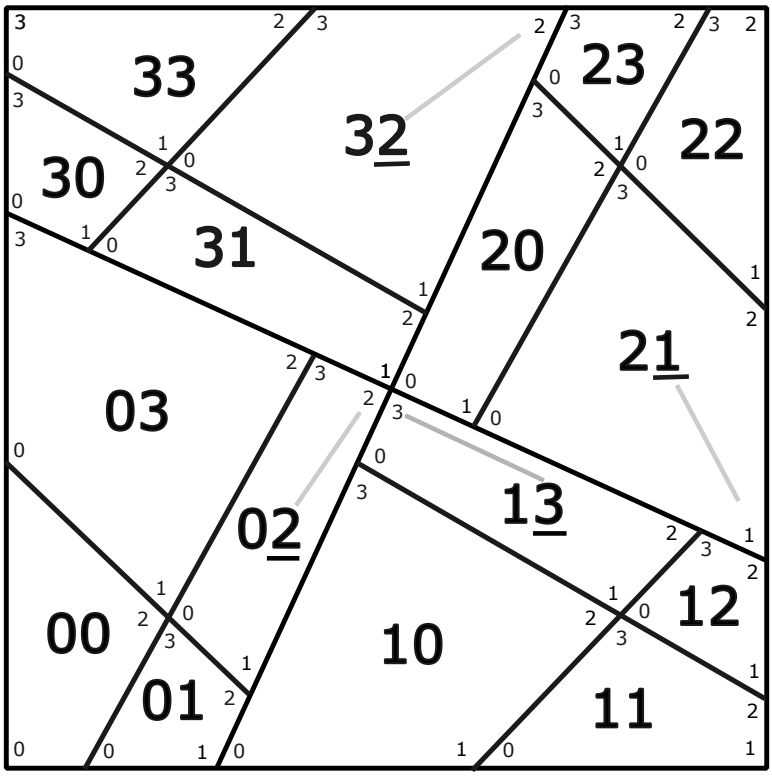


Figure 1.2: A symmetric subdivision rule iterated 2 times.

Figure 1.3 shows what happens when we iterate a symmetric subdivision rule.

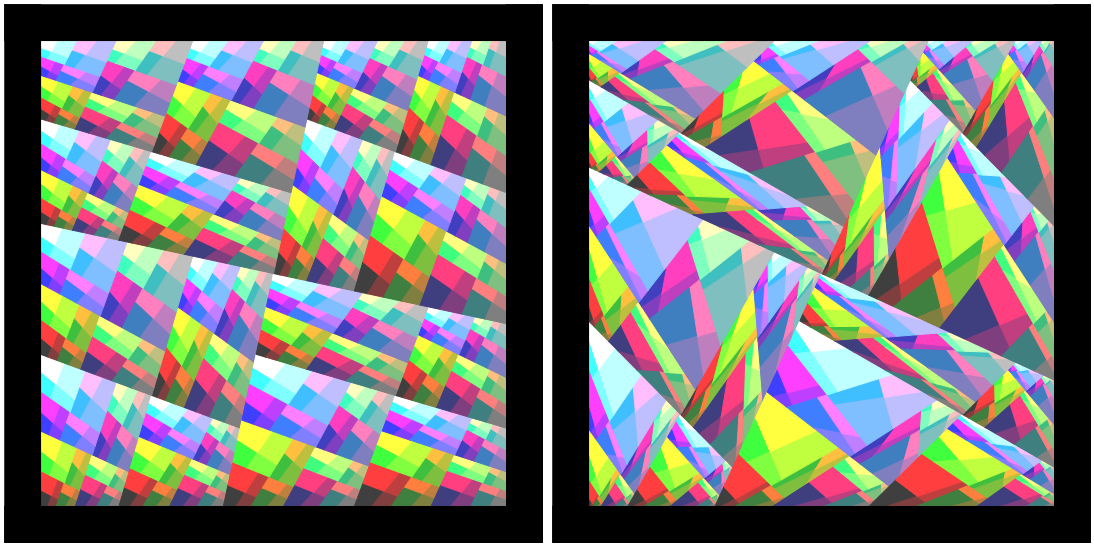


Figure 1.3: 4^5 quads produced by iterated subdivision

In Figure 1.3, we show the process for 2 different parameters. We color the quads according to the last 3 digits of their 4-ary address.

Figure 1.4 shows the iterates of an asymmetric subdivision rule, colored as above. This subdivision rule is an example of what we call a *tame* subdivision rule in the next section. The tame subdivision rules form an open star-shaped subset in $(-1,1)^4$ which contains the set of symmetric subdivision rules. The tame rules turn out to have geometric features which extend some of the properties of the symmetric subdivision rules.

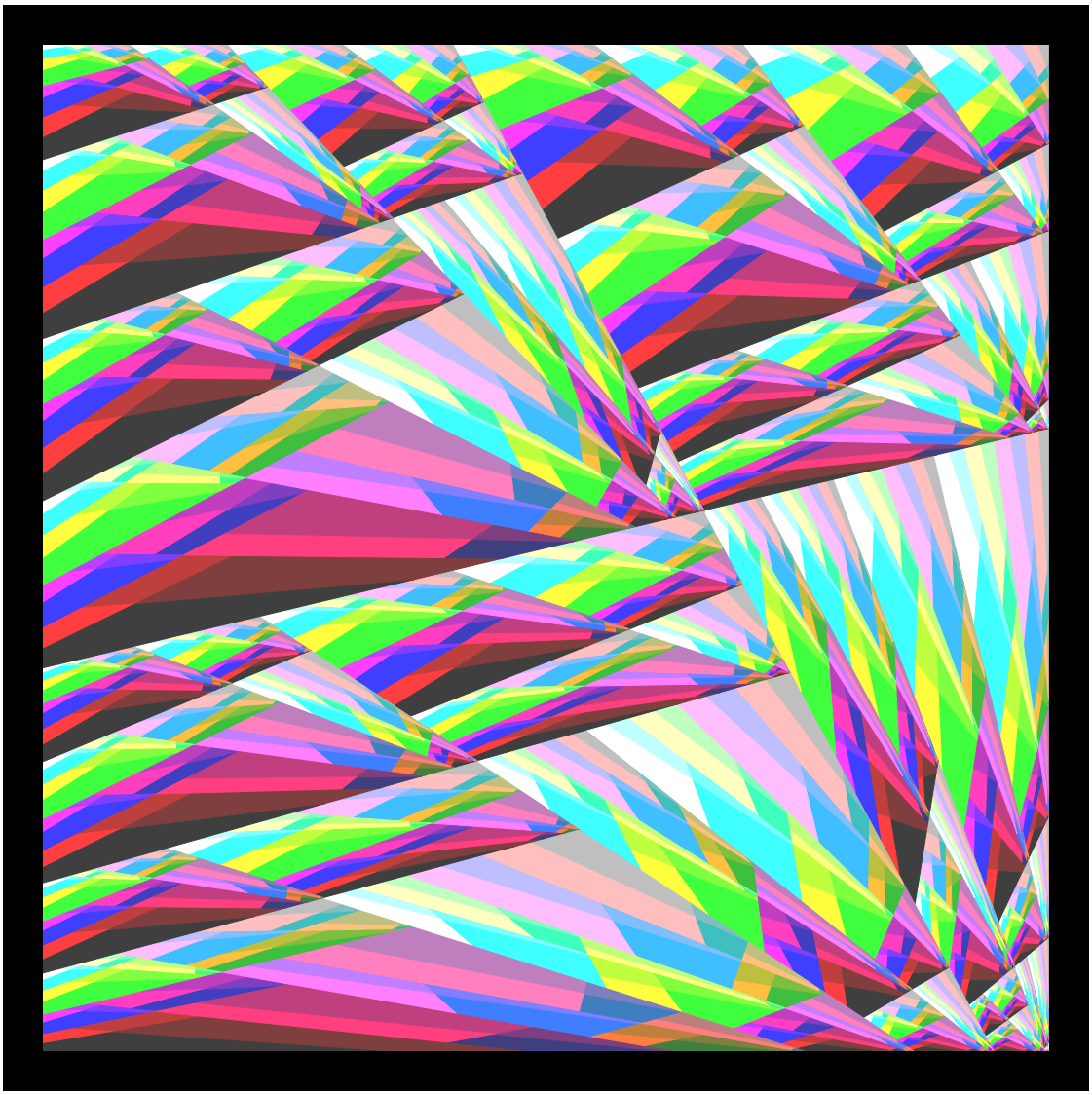


Figure 1.4: An asymmetric subdivision rule

Figures 1.5 and 1.6 show iterations of different asymmetric tame subdivision rules, colored as above. In these two pictures we have zoomed in. I hope that these pictures show some of the beauty and variety of these subdivision processes.

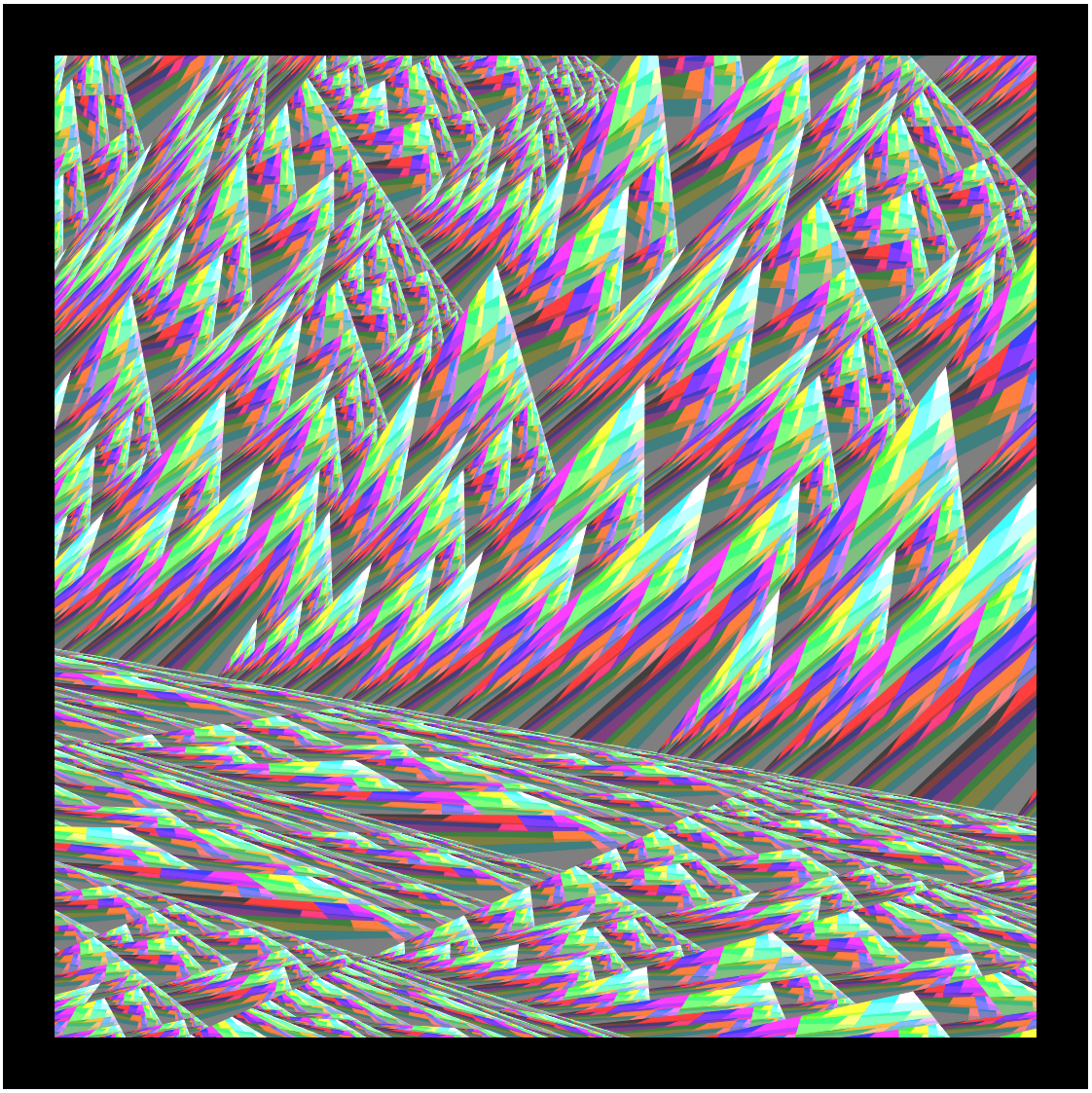


Figure 1.5: Zooming in to an asymmetric subdivision

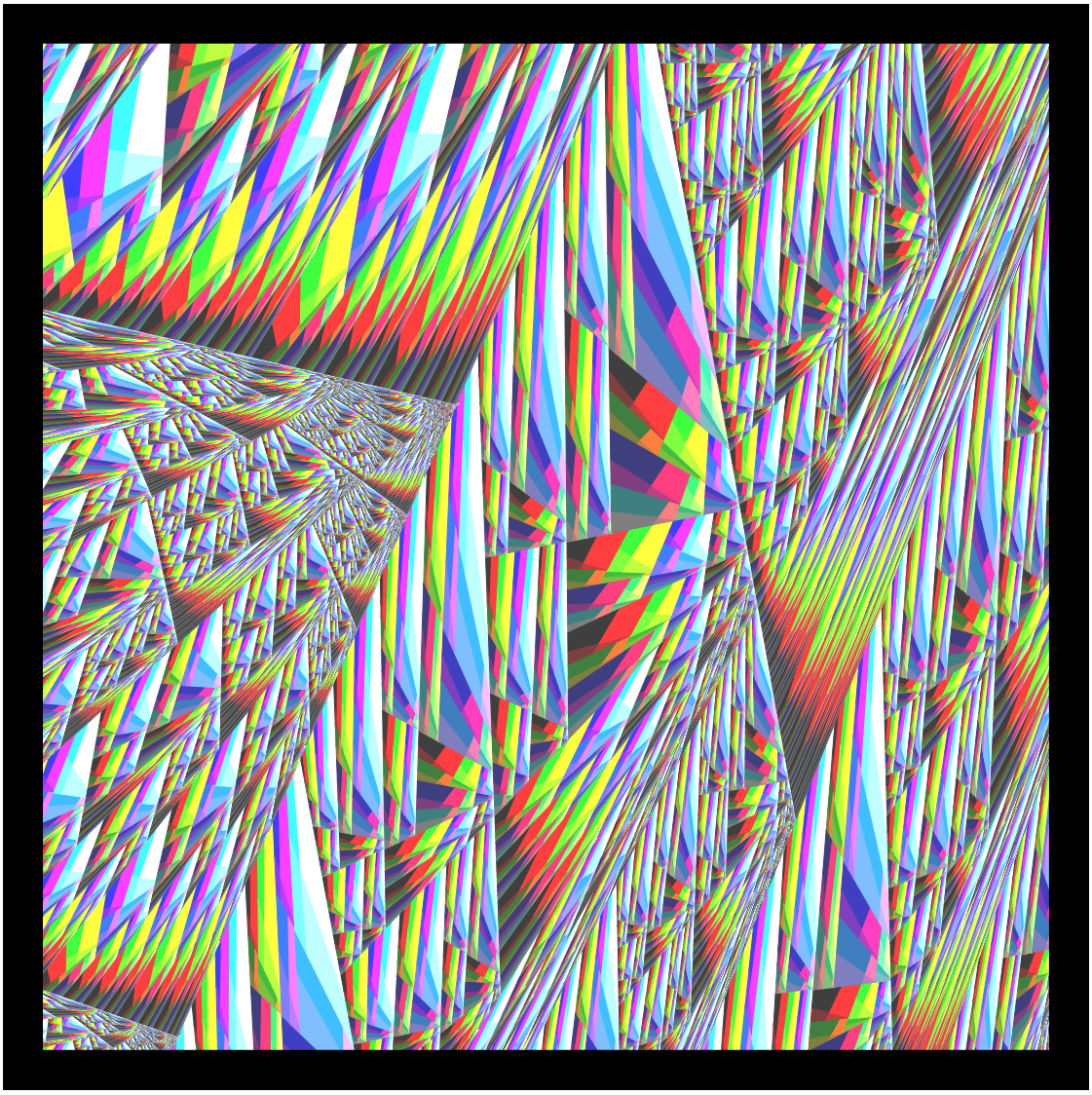


Figure 1.6: Another asymmetric subdivision

The simplest question we address is *Which quads arise under the subdivision rule?* There are many things to say about this. Let us start by fixing some terms. Given a subdivision rule u, we let $\Omega_u(n)$ denote the collection of 4^n quads we get by applying the rule n times to $[-1,1]^2$. We let $\Omega_u = \bigcup_n \Omega_u(n)$. We call Ω_u the *full subdivision* with respect to u. Every quad in the full subdivision has a finite 4-ary sequence associated to it, namely its address. We say that a *nested sequence* in the full subdivision is an infinite sequence $\{Q_n\}$ of quads with $Q_n \in \Omega_u(n)$ and $Q_{n+1} \subset Q_n$. The nested sequences are in bijection with infinite 4-ary sequences.

1.2 Tame Subdivision Rules

Define

$$\begin{aligned}\big[a_1, b, a_2\big] &= 1 + a_1 - b + a_1 b + 2ba_2 \\ &= \big(1 + a_1\big) + \big(-1 + a_1 + 2a_2\big)b \\ &= \big(1 - b\big) + \big(1 + b\big)a_1 + \big(2b\big)a_2.\end{aligned} \tag{1}$$

We have written the same polynomial three ways to highlight its bi-linear structure. This strange looking polynomial turns out to be decisive for many questions we address in this paper. We call a subdivision rule *tame* if

$$\forall k, \qquad \big[+u_{k+0}, +u_{k+1}, +u_{k+2}\big] > 0 \qquad \big[-u_{k-0}, -u_{k-1}, -u_{k-2}\big] > 0. \tag{2}$$

Here indices are taken mod 4. These conditions are satisfied, for instance, when all the u_k have the same sign. The tame locus turns out to be star-shaped with respect to $\big(0,0,0,0\big)$ and also what I call a *bi-convex bi-pentagon*. See §6.4.

We say that the rule u *shrinks quads to points* if every nested sequence shrinks to a point.

Theorem 1.1 *Let u be a tame subdivision rule. Then u shrinks quads to points.*

Remark: When $u = \big(a, -a, a, -a\big)$ Theorem 1.1 says that the rule shrinks quads to points when $|a| < 1/3$. This is sharp in the sense that the rule $\big(a, -a, a, -a\big)$ does not shrink quads to points for $|a| > 1/3$. I did not fully analyze the case when $|a| = 1/3$. In general, I do not know how sharp this result is.

Rather than consider the quads themselves, we can also consider the *shapes* of quads. Let $\mathcal{E}$ denote the space of quadrilaterals modulo similarity and let $\mathcal{B}$ denote the space of quads modulo affine transformations. In §2, we give coordinates on $\mathcal{E}$ so that $\mathcal{B}$ is identified with the open square $\big(-1,1\big)^2$ and $\mathcal{E} = \mathcal{B} \times \boldsymbol{H}^2$, where $\boldsymbol{H}^2$ is the hyperbolic plane. We think of $\mathcal{E}$ as a fiber bundle over $\mathcal{B}$, with fiber $\boldsymbol{H}^2$. We say that an *affine shape* of a quad is its image in $\mathcal{B}$.

We call the subdivision rule u *affinely contained* if the set of affine shapes of quads in Ω_u is contained in a compact subset of $\mathcal{B}$.

Theorem 1.2 *A subdivision rule u is affinely contained if and only if it is tame.*

We define the *affine limit set* Λ_u of the subdivision rule, as follows. A point of $\mathcal{B}$ belongs to Λ_u if and only if every neighborhood of the point intersects an affine shape of a quad in $\Omega_u\big(n\big)$ for arbitrarily large n.

So far we have been studying the affine shapes of quads in the subdivision of the square $[-1,1]^2$ and we defined the limit set relative to this process. But we could iteratively subdivide some other convex quad Q. How would the affine limit set change? In the tame case, not at all! Theorem 1.1 says that the tame subdivision rule shrinks quads to points, and on small scales a projective transformation is nearly affine. So, if we redefined the limit set relative to some other initial convex quad, we would get the same answer.

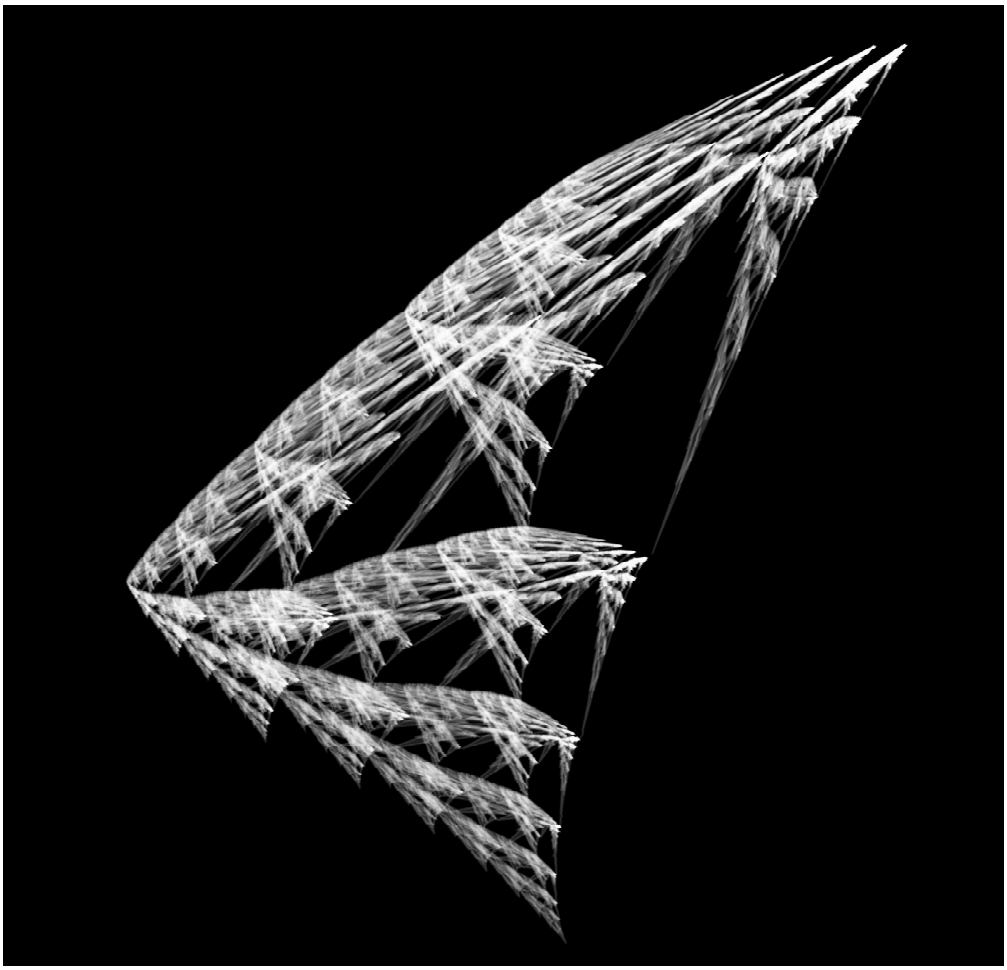


Figure 1.7: The limit set associated to Figure 1.4.

Figure 1.7 depicts the limit set associated with the subdivision in Figure 1.4. We use overlays of translucent quadrilaterals to depict it. In §6.3 we discuss this. Figure 1.8 depicts another limit set – just a random one I liked. In Figures 1.7 and 1.8, the black square is a proper subsquare of $[-1,1]^2$. We are zooming in a bit.

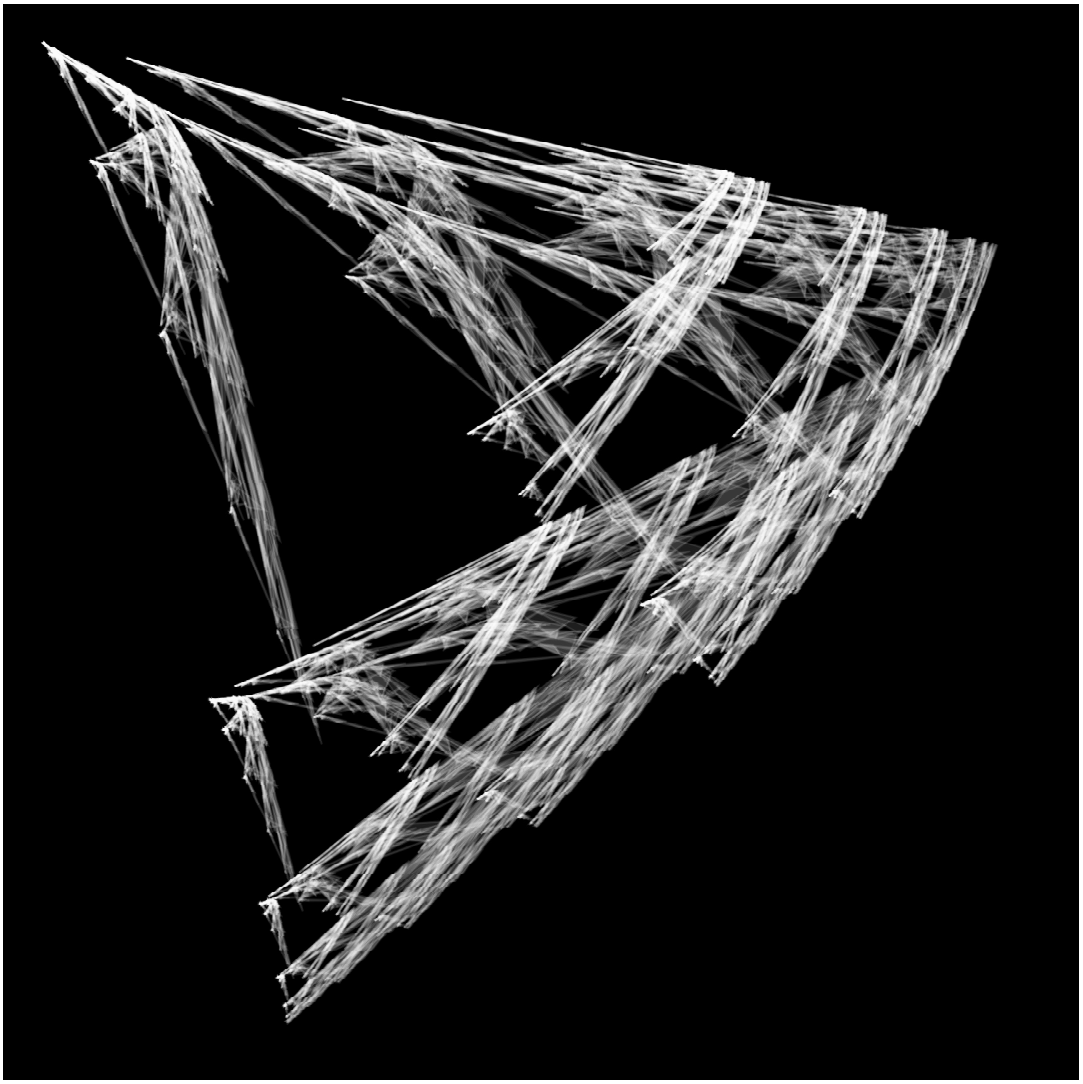


Figure 1.8: Another tame limit set.

For comparison, Figure 6.3 shows a particularly nice limit set that associated to a non-tame subdivision rule. The rule is taken from a more general class of subdivision rules we call *coherent*. In general, the limit set of a subdivision rule is a self-projective fractal produced by a projective iterated function system. Compare [**F**].

1.3 The Reconstruction Theorem

Our most striking result concerns symmetric subdivision rules, namely those of the form $u = (a, a, a, a)$. These are all tame. To avoid trivialities we take $a \neq 0$. We let

$B_u \subset \mathcal{B}$ be the rotationally symmetric square having vertex

$$\left(\frac{2a^2}{1+a^2}, \frac{2a}{1+a^2}\right). \tag{3}$$

The other vertices are obtained from this one by rotating $k\pi/2$ times around the origin for $k = 1, 2, 3$. As a moves from -1 to 1 this vertex traces out an open semicircle on the circle of radius 1 centered at $(1, 0)$.

The notion of a quad makes sense in $\mathcal{B}$ as well, because this is just the open square $(-1, 1)^2$. Though this may seem a strange thing to do, we can carry out the subdivision process in $\mathcal{B}$. Let $\Omega_{-u}(B_u)$ denote the full subdivision of B_u under the subdivision rule $-u = (-a, -a, -a, -a)$.

The *projective center* of a quad is the intersection of its diagonals. This is a projectively invariant notion of a center of a quad.

Theorem 1.3 (Reconstruction) *The affine shape of the quad in Ω_u with address σ is the projective center of the quad in $\Omega_{-u}(B_u)$ whose address is the sequence obtained by putting the digits of σ in reverse order.*

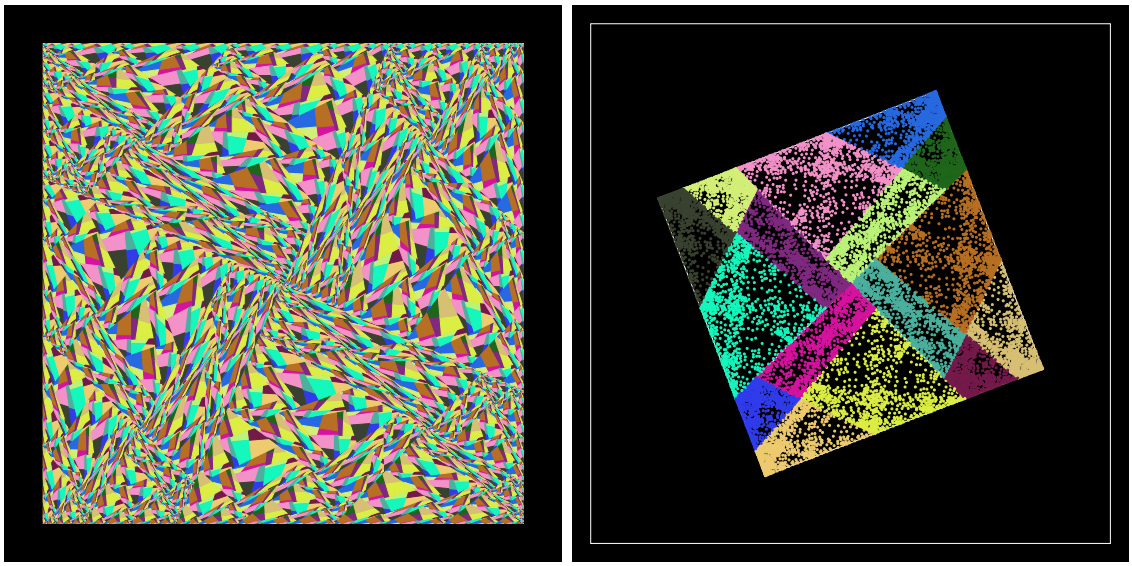


Figure 1.9: The shapes of quads in a symmetric subdivision

Figure 1.9 illustrates Theorem 1.3. On the left we have some symmetric subdivision of the unit square. The quads are colored according to the last two digits of their 4-ary addresses. The right shows their images in $\mathcal{B}$. On the right, the outer white square is $[-1, 1]^2$. The smaller square on the right is B_u.

Note that the projective centers of the quads in the full subdivision are entirely distinct. No center appears twice. (For notational purposes we will explain this for the original subdivision rule, not the parallel process taking place in $\mathcal{B}$.) The reason is that the set of projective centers of quads in $\Omega_u(n)$ is contained in the boundaries of the quads in $\Omega_u(n+k)$ for $k > 0$. Thus, Theorem 1.3 has the following corollary.

Corollary 1.4 *No two quads in Ω_u have the same affine shape.*

Combining Theorems 1.3 and 1.1 we have

Corollary 1.5 *The square B_u is the affine limit set of Ω_u.*

It might seem that the strange behavior described in Theorem 1.3 is just a trick of labeling. This is not the case. To highlight this, we mention one more corollary.

Corollary 1.6 *Almost every infinite nested sequence of quads in the u-subdivision has this property: The set of affine shapes of quads in the sequence is dense in B_u.*

This result holds because almost every infinite 4-ary sequence has the following property: It contains every finite 4-ary string as a consecutive block.

A version of Theorem 1.3 is true for the tame *semi-symmetric* subdivision rules of the form (a, b, a, b). This time B_u is a centrally symmetric parallelogram.

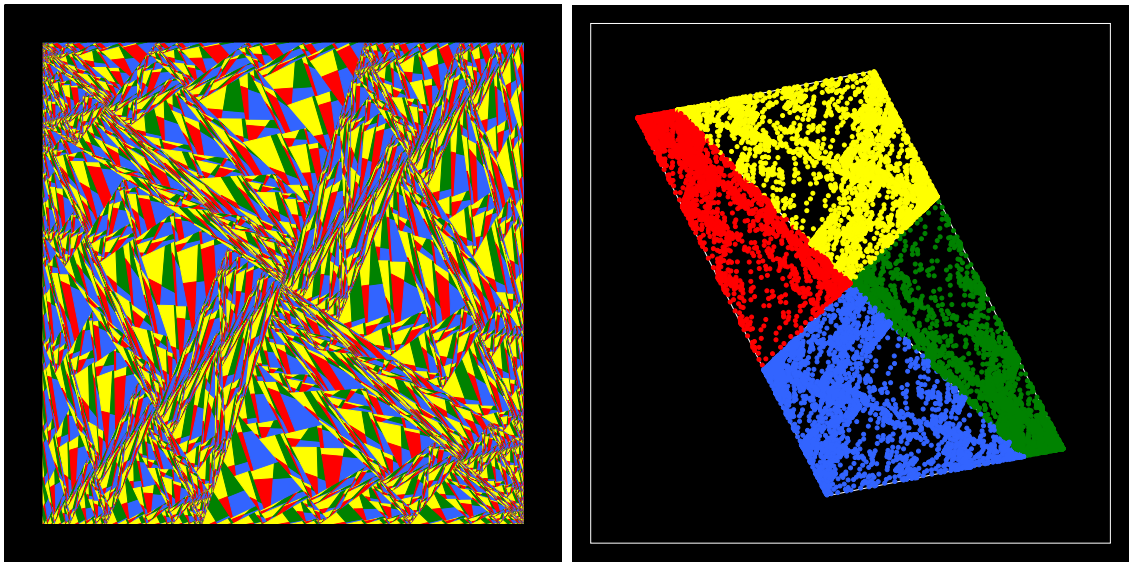


Figure 1.10: The shapes of quads in a semi-symmetric subdivision

The proof in the more general case is essentially the same as what we do in the paper. We omit the details. For completely asymmetric subdivision rules the phenomenon vanishes.

Theorem 1.3 also has a version for the Pappus iteration in [**S1**] and [**S2**]. One can put this iteration inside the framework of self-projective carpets with two enhancements: First, one needs to alternate between two different subdivision rules. Second, one needs to intertwine these rules with a certain cellular automaton that generates the Morse-Thue sequence. I will elaborate the Pappus analog of Theorem 1.3 in a sequel paper.

1.4 Skew Product Structure

Now we go back to the general case, to briefly discuss the main idea that underlies our proofs of the above results. Fixing a subdivision rule (u_0, u_1, u_2, u_3) it turns out that each branch of the subdivision acts as a skew product on $\mathcal{E}$. That is, the kth branch has the form

$$F_k(b, \tau) = (f_k(b), h_{k,b}(\tau)), \tag{4}$$

where f_k is a projective transformation acting on $\mathcal{B}$ and $h_{k,b}$ is a hyperbolic isometry depending on k and b. This is a fusion of a hyperbolic isometry semigroup and a projective transformation semigroup. We work out the structure in §2.

The symmetric case is especially nice. When we use the upper half-plane model for $\boldsymbol{H}^2$, two of the maps $h_{k,b}$ are given by upper triangular matrices and two of them are given by lower triangular matrices.

Remark: (For experts) The projective plane $\boldsymbol{P}$ has its tangent bundle, and also the associated bundle of scale classes of positive definite quadratic forms. One can interpret this bundle as a hyperbolic plane bundle over $\boldsymbol{P}$. Any projective transformation acts differentially on the tangent bundle and induces a skew product on the associated bundle. I tried to reconcile my skew product action with this tautological skew product, which just comes from the base map. They seemed different to me.

In this paper we are really only considering the induced action of the skew product on the base space $\mathcal{B}$. Surely, there is a lot to say about what happens in the fibers. Such an analysis would probably lead to a good understanding of the similarity classes of shapes in the iterated subdivision rule. This project is beyond the scope of our introductory paper. I hope to write one or several sequel papers on this matter.

1.5 Organization

In §2 we discuss the skew product action of our subdivision rules. At the end of §2 we prove a case of Theorem 1.2 for what we call *wild* subdivision rules.

In §3 we prove the Shrinking Lemma, a result that explains the connection between Theorems 1.1 and 1.2.

In §4 we prove Theorem 1.1 and the tame case of Theorem 1.2.

In §5 we prove Theorem 1.3.

In §6 we prove Theorem 1.2 for subdivision rules which are neither tame nor wild. This final case completes the proof of Theorem 1.2. We also give more information about the space of all subdivision rules.

1.6 AI Usage

I talked extensively with ChatGPT (GPT-6) about this paper but the ideas for the paper are mine, and came from my own computer experiments using Mathematica [**W**] and Java. Here is what ChatGPT did for me:

- ChatGPT helped typeset the equations in the paper.
- ChatGPT helped extensively with debugging and analyzing my computer code.
- ChatGPT mostly wrote the Java subroutines which exported my pictures to EPS files. This is hard tedious work.
- ChatGPT helped with the proof of Lemma 6.3.

1.7 Acknowledgements

I thank Fanny Kassel, Curtis McMullen, Alexander Nolte, and Pierre Pansu for various discussions about this project. I also thank the I.H.E.S. for their support during the time I did this exploratory research. Their support came in the form of the Gretchen and Barry Mazur Chair.

2 Generalities

2.1 A Primer on Projective Geometry

In this section we give a quick overview of elementary projective geometry.

The Basic Spaces: Let $O \in \mathbf{R}^3$ be the origin. The *real projective plane* $\mathbf{P}$ is the space of lines through O in $\mathbf{R}^3$. We have the *projectivization map* $\pi : \mathbf{R}^3 - O \to \mathbf{P}$ which maps a point $P \in \mathbf{R}^3 - O$ to the point in $\mathbf{P}$ representing the line $\overline{OP}$.

A *line* in $\mathbf{P}$ is the set of points whose preimages under π all lie in the same plane through O. The *flag space* is the space of pairs (p, ℓ) where $p \in \mathbf{P}$ and ℓ is a line in $\mathbf{P}$ and $p \in \ell$.

The *affine patch* of $\mathbf{P}$ is the set of lines through O not contained in the XY-plane. We identify this patch with $\mathbf{R}^2$ by intersecting each such line with the plane $Z = 1$.

Projective Transformations: The group $GL_3(\mathbf{R})$ of invertible linear transformations of $\mathbf{R}^3$ acts on $\mathbf{P}$, with two scalar multiples of the same matrix acting the same way. Thus, the group of projective automorphisms of $\mathbf{P}$ is denoted $PGL_3(\mathbf{R})$. It is the group $GL_3(\mathbf{R})$ modulo scaling.

The Cross Product: The cross product is very useful in projective geometry. If A and B represent points in $\mathbf{P}$ then the vector $L = A \times B$ gives a linear functional which represents the line $\overline{AB}$ in $\mathbf{P}^*$. At the same time, if A and B represent lines in $\mathbf{P}$ (viewed as scale classes of linear functionals) then L represents the intersection point $A \cap B$.

Bare Quads: Four points in $\mathbf{P}$ are in *general position* if no three of them are collinear. We will call such collections *bare quads*. It is a well-known fact that for any two bare quads (p_0, p_1, p_2, p_3) and (q_0, q_1, q_2, q_3) there is a unique projective transformation T such that $T(p_i) = q_i$ for $i = 0, 1, 2, 3$. This situation makes it nice to consider iterative subdivision rules based on bare quads.

Convex Quads: It is even nicer to consider iterated subdivision rules based on *quads*. A *quad* is a bare quad (p_0, p_1, p_2, p_3) together with the extra data specifying, for $i = 0, 1, 2, 3$, one of the two segments connecting p_i to p_{i+1} on the line $\overline{p_i p_{i+1}}$. Here indices are taken mod 4. Thus, each bare quad gives rise to 16 quads.

A quad $Q = (p_0, p_1, p_2, p_3)$ is called *convex* if there is a projective transformation T such that $T(Q) \subset \mathbf{R}^2$, and $T(Q)$ is a convex quadrilateral in the ordinary sense.

This means, in particular, that the points $T(p_0), ..., T(p_3)$ go around $T(Q)$ in cyclic order. For any two convex quads P and Q there is a unique projective transformation which maps P to Q and respects the labeling. Unless otherwise stated, we will work with quads that already live in $\boldsymbol{R}^2$. We always label such quads so that the points go around counter-clockwise.

Every convex quad has a natural *projective center*. This is the intersection of the diagonals of the quad. A projective transformation mapping one quad to another maps the projective center of the first quad to the projective center of the second one.

The cross ratio and the Hilbert metric: The *cross ratio* of 4 points on the line is

$$\chi(a,b,c,d) = \frac{(a-c)(b-d)}{(a-b)(c-d)}.$$

When $a = \infty$ the formula simplifies to $(b-d)/(c-d)$. In particular

$$\chi(\infty, -1, 1+s, 1) = \frac{-2}{s}.$$

A convex domain in $\boldsymbol{P}$ is *bounded* if it is projectively equivalent to a bounded convex domain in $\boldsymbol{R}^2$. Given an open bounded convex domain $X \subset \boldsymbol{P}$, and two points $b, c \in X$ we define

$$d_X(b,c) = \log \chi(a,b,c,d). \tag{5}$$

Here a and d are the points where the line $\overline{bc}$ intersects ∂X. We take a, b, c, d in order. This defines a metric on X called the *Hilbert metric*. This metric is projectively natural: A projective transformation carrying X to Y is an isometry relative to (d_X, d_Y). When X is an open disk, d_X is isometric to the hyperbolic metric, up to scaling.

2.2 The Space of Quads

Now we consider the space of labeled quads mod similarity. We first describe a useful but fairly non-geometric normalization, then we give more geometric coordinates, then we relate the two systems. The first normalization is useful for calculations.

Initial Normalization: Let $\mathcal{Q}$ denote the space of convex labeled quadrilaterals. Let $\mathcal{E}$ denote the quotient of $\mathcal{Q}$ modulo similarity. We take members of $\mathcal{Q}$ to be in $\boldsymbol{R}^2$ and we insist that they are all oriented counter-clockwise. That is, for a quad Q,

the vertices Q_0, Q_1, Q_2, Q_3 go around counter-clockwise. With this convention, we only take orientation-preserving similarities.

Each member $[Q]$ of $\mathcal{E}$ has a unique representation satisfying the following normalization.

$$Q_0 = (x_0, 0), \quad Q_1 = (x_1, y_1), \qquad Q_2 = (x_0 + 1, 0), \qquad Q_3 = (tx_1, ty_1)$$
$$x_0 \in (-1, 0), \qquad x_1 \in \boldsymbol{R}, \qquad y_1 < 0, \qquad t < 0. \tag{6}$$

Geometrically we are scaling so that the diagonals cross at the origin and the diagonal Q_0Q_2 is horizontal and unit length. This shows that $\mathcal{E}$ is homeomorphic to $\boldsymbol{R}^4$.

Geometric Coordinates: Now we explain more geometric coordinates on $\mathcal{E}$. Our construction refers to Figure 2.1.

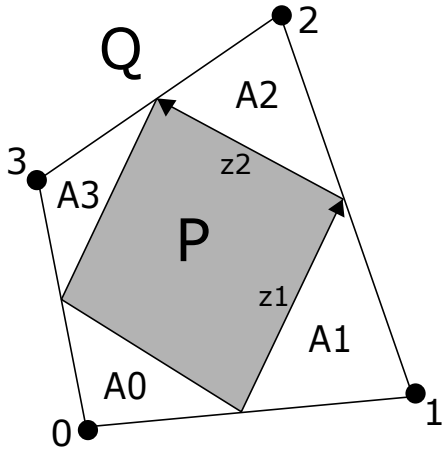


Figure 2.1: The basic construction

Let Q be a labeled convex quadrilateral as in the outermost quad in Figure 2.1. We let P denote the parallelogram obtained by joining the midpoints of the sides of Q. Here P is shaded in Figure 2.1. Let z_1 and z_2 be the two oriented edges of P as shown in Figure 2.1. We interpret these vectors as complex numbers and then define

$$\tau(Q) = \frac{z_2}{z_1}. \tag{7}$$

In other words, we scale P by an orientation-preserving similarity so that $z_1 = 1$ and then we record the value of z_2. We interpret $\tau(Q)$ as a point in the upper half plane model of $\boldsymbol{H}^2$. Let A_0, A_1, A_2, A_3 be the areas of the 4 triangles of $Q - P$. Define

$$\alpha(Q) = (B_2 - B_0, B_3 - B_1) \in (-1, 1)^2, \qquad B_j = \frac{2A_j}{\sum A_k}. \tag{8}$$

Note that $B_0 + B_2 = B_1 + B_3 = 1$. That is why we take the difference. The quantity $\alpha(Q)$ is an affine invariant of Q.

Define

$$\Psi(Q) = (\alpha(Q), \tau(Q)). \tag{9}$$

Relating the two Systems: Let us work with the quads which are normalized as in Equation 6. For such a quad we have

$$\tau = (t-1)(x_1 + iy_1), \qquad \alpha = \left(1 + 2x_0, \frac{t+1}{t-1}\right). \tag{10}$$

Here is a formula for the inverse. We display our coordinates in $\boldsymbol{C}$. If $\alpha = (a, b)$ then a routine calculation shows that

$$Q_0 = \frac{a-1}{2}, \qquad Q_1 = \frac{(b-1)\tau}{2}, \qquad Q_2 = \frac{a+1}{2}, \qquad Q_3 = \frac{(b+1)\tau}{2}. \tag{11}$$

Thus Ψ is a homeomorphism from $\mathcal{E}$ to $(-1,1)^2 \times \boldsymbol{H}^2$.

Mostly we will work with the geometric coordinates. However, we will sometimes use the initial coordinates to compute things.

2.3 Skew Maps

Here we give the basic definitions about the maps comprising projective subdivision rules, and also about skew products.

Skew Maps: A *skew map* on $\mathcal{E}$ is a map of the form

$$F(\alpha, \tau) = (f(\alpha), h_\alpha(\tau)), \tag{12}$$

where f is a projective transformation mapping $(-1,1)^2$ into $(-1,1)^2$ and h_α is an isometry of $\boldsymbol{H}^2$ that depends on α.

Quad Maps: A projective transformation T is *adapted* to $Q \in \mathcal{Q}$ if $T(Q) \in \mathcal{Q}$. A map $F : \mathcal{Q} \to \mathcal{Q}$ is a *quad map* if it satisfies the following two conditions.

1. $F(T(Q)) = T(F(Q))$ for all Q and all projective T adapted to Q.
2. $F(Q) \subset Q$ for all $Q \in \mathcal{Q}$.

To specify a projectively natural map we just have to explain what it does on a square of our choice.

The Method of Computation: We will give a proof below that quad maps define skew maps on $\mathcal{E}$. Here we explain the process we use for computing them. Let P be the quad $[-1,1]^2$ with $P_0 = (-1,-1)$ and $P_1 = (1,-1)$, etc. Suppose we have some quad map F.

1. Find the normalized quad X with invariants $\alpha = (a,b)$ and $\tau = \tau_1 + i\tau_2$.
2. Find the projective transformation T such that $T(P) = X$.
3. Compute the geometric invariants for $T(F(P)) = F(X)$.
4. Extract the matrix representatives for $f \in PGL_3(\boldsymbol{R})$ and $h_\alpha \in PGL_2(\boldsymbol{R})$.

We wrote a Mathematica file to do the calculation. Our logic is a bit roundabout. In the next subsection we use the procedure for a special kind of example. We then use this special example to deduce that the computation will always work. Then we use the calculation several times again later on, knowing in advance that it works.

2.4 Quad Maps as Skew Maps

In this section we prove that every quad map acts on $\mathcal{E}$ as a skew map. We first give two examples which generate all the quad maps, and then we put together the final argument.

Relabeling Example: Cyclically relabeling one click counterclockwise has the action

$$(\alpha,\tau) \to (R\alpha, -1/\tau), \tag{13}$$

where R is counterclockwise rotation by $\pi/2$ radians. This is certainly a skew map. The map sends fibers to fibers and maps each fiber by the same hyperbolic isometry, namely $z \to -1/z$.

Slide Maps: For $s \in (-2,0]$ we consider the map $F_s : \mathcal{Q} \to \mathcal{Q}$ which takes the square P to the quad Q with vertices

$$Q_0 = (-1,-1), \qquad Q_1 = (+1,-1), \qquad Q_2 = (+1,+1+s), \qquad Q_3 = (-1,+1)$$

We extend F_s to a quad map. The map F_0 is the identity. We call F_s a *slide map* because we are essentially sliding one vertex of P along an edge to arrive at Q. There are 8 such 1-parameter families of maps.

Below we will calculate that F_s acts as a skew product, and we will justify the formulas we now give. We suppress the value s from some of our notation. Referring to Equation 12, we give the maps f and h_α which act respectively as a projective transformation of the base and as a linear fractional transformation of the fiber over $\alpha = (a, b)$.

$$f = 4I_3 + s\begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}. \quad h_\alpha = 4I + \frac{s}{1-a}\begin{pmatrix} -2(a+b) & 0 \\ 1-b^2 & (1-a)(1-b) \end{pmatrix}. \tag{14}$$

Here I_k is the $k \times k$ identity matrix.

The first matrix in Equation 14 has a determinant of $16(s+2)^2 \neq 0$ and hence is nonsingular. Let ϕ_s be the projective transformation which maps the quad P to the quad Q. Then ϕ_s and f_s are conjugate by reflection in the line $a = b$. This line goes through the vertex we are "sliding".

The second matrix, the hyperbolic part, has determinant

$$\det(h_\alpha) = \frac{2cd}{1-a}, \quad c = (2+s)(1-a) - s(1+b), \quad d = (2+s)(1-b) + 2(1+b).$$

Both c and d are positive because $s \in (-2, 0]$ and $a, b \in (-1, 1)$. Hence the second determinant is positive as well. Indeed, for $s < 0$ we compute that

$$\frac{\mathrm{tr}(h_\alpha)^2}{\det(h_\alpha)} = 4 + \frac{s^2(1+a)^2(1+b)^2}{2(1-a)cd} > 4.$$

Thus h_α is a loxodromic isometry in every fiber when $s \in (-2, 0)$. When $s = 0$, of course, the map is the identity.

We also remark that the family $\{F_s\}$ is a 1-parameter semigroup of skew maps. We will not do anything with this information in the paper, but it seems nice to know.

Composition: We leave to the reader the fun exercise that any quad map is the composition of cyclic relabeling and some finite number of slide maps. (Hint: draw a kind of pinwheel picture.) The idea is that you can modify the vertices one at a time until you reach the desired target map. Given the skew product nature of the slide maps and cyclic relabeling, all quad maps act as skew products on $\mathcal{E}$.

2.5 Wild Subdivision Rules

We say that a subdivision rule $u = (u_0, u_1, u_2, u_3)$ is *wild* if at least one of the inequalities in Equation 2 changes from $>$ to $<$. The subdivision rules which are neither tame nor wild are a boundary case, where at least one of the inequalities in Equation 2 is an equality. In this section we will prove Theorem 1.2 in the wild case.

Lemma 2.1 *A wild subdivision rule u is not affinely contained.*

Proof: Recall that $\mathcal{B} = (-1,1)^2$. Let $f_k : \mathcal{B} \to \mathcal{B}$ be the projective transformation coming from the skew product structure of the kth branch of the subdivision. We compute f_k using the method described above. We will concentrate on f_0.

We check that the linear transformation representing f_0 has $(-1,-1,1)$ and $(-1,1,1)$ as eigenvectors, and also a third eigenvector $\widehat{b}$. When we normalize so that the eigenvalue corresponding to $\widehat{b}$ equals 1, we find that the other two eigenvalues are

$$\lambda_1 = \frac{(1-u_0)(1+u_3)}{\langle u_2, u_3, u_0\rangle} > 0. \qquad \lambda_2 = \frac{(1-u_0)(1+u_3)}{\langle u_3, u_0, u_1\rangle} > 0. \tag{15}$$

Here we have set

$$\langle a,b,c\rangle = 2 + a - c + ab + bc = (1+a)(1+b) + (1-b)(1-c) > 0. \tag{16}$$

From these equations we see that $\lambda_1, \lambda_2 > 0$. We compute

$$1 - \lambda_1 = \frac{[u_2, u_3, u_0]}{\langle u_2, u_3, u_0\rangle}, \qquad 1 - \lambda_2 = \frac{[-u_1, -u_0, -u_3]}{\langle u_3, u_0, u_1\rangle}. \tag{17}$$

In the tame case everything in sight is positive, and so these two eigenvalues are in $(0,1)$ and the eigenvalue corresponding to $\widehat{b}$ is largest. Dynamically, the corresponding point $b \in \boldsymbol{R}^2$ is an attracting point for the base map. We call b the *apex* of the system. We set $b_0 = b$ and then let b_1, b_2, b_3 be the remaining apices. So, in the tame case, all the apices are attracting fixed points.

In the wild case, we can cyclically relabel so that at least one of the expressions in Equation 17 is negative. Hence at least one of λ_1 or λ_2 is greater than 1. If $\lambda_1 = \lambda_2$ then the corresponding edge of $[-1,1]^2$ is an attracting set for the corresponding base map. If $\lambda_1 \neq \lambda_2$ then the vertex of $[-1,1]^2$ corresponding to the larger eigenvalue is attracting. In either case, if we start with a point in $\mathcal{B}$ and iterate the base map, the orbit exits every compact subset of $\mathcal{B}$. If we start with the point $(0,0)$, which is the affine shape of the square, then the orbit consists entirely of affine shapes in the u-subdivision. Hence this orbit exits every compact subset of $\mathcal{B}$. ♠

3 The Shrinking Lemma

3.1 The Main Result

In this chapter we give sufficient conditions for the quads in an iterated subdivision to shrink to points. This result will feed into the proof of Theorem 1.1.

Let $u = (u_0, u_1, u_2, u_3)$ be a subdivision rule, as in the introduction. We have four projective transformations $\phi_0, \phi_1, \phi_2, \phi_3$ associated to the subdivision rule. The map ϕ_k maps $[-1,1]^2$ to the quad we get by taking the kth branch of the subdivision. We call these maps the *associated projective transformations*.

We define the *full subdivision* relative to u to be the set of all quads which arise in the iterated u-subdivision of $[-1,1]^2$. We say that a *nested sequence* in the full subdivision is an infinite sequence $Q_0 \supset Q_1 \supset Q_2...$ of quads such that Q_{n+1} is one of the 4-quads in the u-subdivision of Q_n for all $n = 0, 1, 2, ...$ Recall that Ω_u is the collection of all quads in the full u-subdivision of $[-1,1]^2$. In this section we prove the following result, which aids in the proof of Theorem 1.1.

Lemma 3.1 (Shrinking) *Suppose that*

1. *The set of affine shapes of quads in Ω_u is contained in a compact subset of $\mathcal{B}$.*
2. *Each associated projective transformation has no fixed points in $[-1,1]^2$ except at vertices.*

Then every nested sequence of quads shrinks to a point.

Here we outline the proof. We will argue by contradiction. Suppose that we have a nested sequence that does not shrink to a point. Corresponding to our nested sequence, there is an infinite 4-ary sequence which says which branch of the subdivision is used at each stage. Write this sequence as $s = s_0 s_1 s_2....$ We will consider 3 cases.

1. There are infinitely many indices k such that $s_{k+1} = s_k + 2 \bmod 4$.
2. The 4-ary sequence has infinitely many pairs s_k, s_{k+1} consisting of cyclically adjacent but unequal indices.
3. The 4-ary sequence is eventually constant.

3.2 Case 1

Let $\{Q_j\}$ be our nested sequence and suppose that $s_k = 0$ and $s_{k+1} = 2$. In this situation, Q_k and Q_{k+2} have disjoint boundaries. The second quad is contained in the interior of the first quad. The left side of Figure 3.1 shows what this looks like.

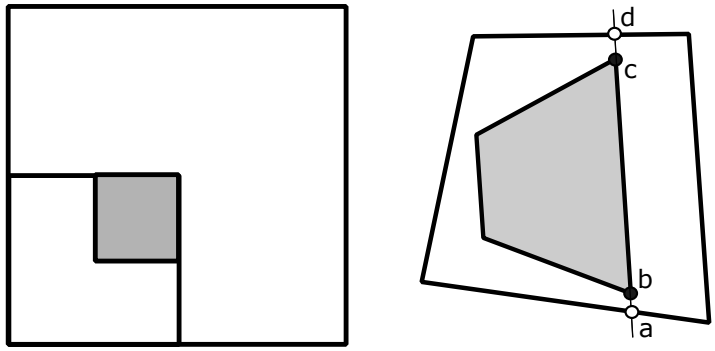


Figure 3.1: The picture at a nice index

There are only finitely many pairs (Q_k, Q_{k+2}) like this up to projective transformation. Therefore there is a finite upper bound to the Hilbert diameter of Q_{k+2} relative to the Hilbert metric on Q_k.

On the other hand, if our quads are not shrinking to points, the ratio

$$\frac{\text{diam}(Q_{k+2})}{\text{diam}(Q_k)}$$

tends to 1 as we take increasingly large good indices k. But then, using basic properties of the cross ratio, the Hilbert distance between two diameter-realizing points $b, c \in Q_{k+2}$ tends to ∞. This is a contradiction. The right side of Figure 3.1 shows what is going on geometrically.

3.3 Case 2

Say that a *model* is a pair (Q, Q') of quadrilaterals where one edge of Q' lies in the interior of an edge of Q, and the remaining edges of Q' are disjoint from the remaining edges of Q. Call the model *normalized* if the diagonals of Q are perpendicular and have unit length, and one vertex of Q is the origin. Every model is affinely equivalent to a normalized model.

Again by finiteness, we can find infinitely many indices where (Q_k, Q_{k+2}) is a model, and all models we consider are projectively equivalent. The bounded image hypothesis implies that the set of normalized models (Q, Q') affinely equivalent to

(Q_k, Q_{k+2}) is such that Q lies in a compact set. But because all these models are projectively equivalent, Q' also lies within a compact set. In other words, there is a compact set of normalized models such that every model (Q_k, Q_{k+2}) is affinely equivalent to one of them.

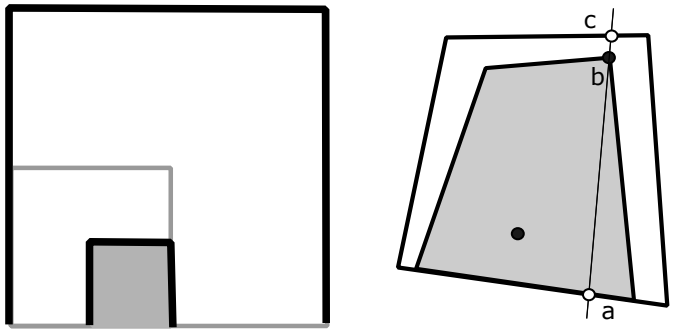


Figure 3.2: A model

If our sequence does not shrink to a point then we can find a sequence of examples, as on the right side of Figure 3.2, where the ratio of distances $d(a,b)/d(a,c)$ converges to 1. But ratios of distances for collinear points are preserved by affine transformations. This contradicts the compactness of our set of normalized models.

3.4 Case 3

In this case we can chop off the beginning of the sequence, cyclically relabel, and normalize by a projective transformation to reduce to the case that the sequence is 000.... In this case we have

$$Q_n = \phi_0^n(Q_0).$$

If $\{Q_n\}$ does not shrink to a point then $\Delta = \bigcap Q_n$ is a convex polygon with at most 4 vertices. By construction, ϕ_0 permutes these vertices. Since ϕ_0 fixes the vertex $(-1,-1)$ of Δ and preserves the cyclic order of the vertices, we see that ϕ_0 fixes each vertex. In particular, $\phi_0(v) = v$ for some $v \neq (-1,-1)$. Note that ϕ_0 fixes $(-1,-1)$ and maps the other three vertices of $[-1,1]^2$ to non-vertices of $[-1,1]^2$. Hence $v \in [-1,1]^2$ is not a vertex. This is a contradiction.

4 Tame Subdivision Rules

4.1 Path Connectedness

This chapter is devoted to establishing that the conditions in the Shrinking Lemma are satisfied for tame subdivisions. In particular, we prove Theorem 1.2 for tame subdivision rules: Every tame subdivision rule is affinely contained. The Shrinking Lemma then implies Theorem 1.1. All subdivision rules in this chapter are tame.

Let $S \subset (-1,1)^4$ denote the set of tame subdivision rules. These are the ones which satisfy Equation 2. In this section we prove that S is star shaped and in particular path connected.

Lemma 4.1 *If* $(u_0,u_1,u_2,u_3) \in S$ *and* $s,t \in [0,1]$ *then* $(su_0,tu_1,su_2,tu_3) \in S$.

Proof: A moment's thought – think cyclic relabeling to reverse the roles of s and t – will convince the reader that it suffices to prove this for $s = 1$. Let Π be the plane on which we fix the values of u_0, u_2. We identify Π with the (u_1,u_3) plane. Note that $(u_0,0,u_2,0) \in S$ because $[a,0,c] = 1 + a > 0$ and $[0,b,0] = 1 - b > 0$ for all $a,b,c \in (-1,1)$. Hence $(0,0) \in \Pi \cap S$. On Π, the constraints in Equation 2 are linear in u_1 and u_3. Hence $\Pi \cap S$ is an open convex polygon, and in particular star-shaped with respect to the origin in Π. ♠

4.2 Location of the Fixed Points

We fix a tame subdivision rule u. Let Σ_k denote the kth subdivision operation relative to u. Let ϕ_k be the projective transformation that maps $[-1,1]^2$ to $\Sigma_k([-1,1]^2)$. We will analyze the situation for ϕ_0. The other fixed points for ϕ_0 are

$$p_1 = \left(-1, \frac{2 - u_0 + u_2 - 2u_3 + 3u_0u_3 + u_2u_3}{(u_0 + u_2)(1 + u_3)}\right). \tag{18}$$

$$p_2 = \left(-\frac{2 + 2u_0 - u_1 + u_0u_1 + u_3 + 3u_0u_3}{(1 - u_0)(u_1 + u_3)}, -1\right). \tag{19}$$

Lemma 4.2 *There is no fixed point in* $[-1,1]^2$ *other than* $(-1,-1)$.

Proof: When $|u_k|$ is very small for $k = 0,1,2,3$ we have $u \in S$ and moreover p_1, p_2 are nearly at ∞. The line $L = \overline{p_1p_2}$ misses $[-1,1]^2$. Given the action of ϕ_0, neither

the bottom edge nor the left edge is fixed pointwise. Hence, the fixed set is either $\{(-1,-1), p_1, p_2\}$ or $\{(-1,-1)\} \cup L$.

Suppose that the fixed set intersects $\big[-1,1\big]^2$ at some point other than $(-1,-1)$ for some $u' \in S$. Since S is path connected we can take a path in S from our small u to our failing u'. We think of this path as being parametrized by time. The points p_1 and p_2 vary continuously with time, on the projective lines that contain them.

Let us first rule out the possibility that $p_1 = (-1,-1)$ along our path. Solving $p_1 = (-1,-1)$ gives

$$\frac{2\big[u_2, u_3, u_0\big]}{(u_0 + u_2)(1 + u_3)} = 0. \tag{20}$$

The numerator cannot vanish because this is exactly one of the expressions in the tameness criterion. If $u_0 + u_2 = 0$ then the fixed point is on the line at infinity. So, there is no way to arrange that $p_1 = (-1,-1)$ for $u \in S$. A similar calculation rules out the possibility that $p_2 = (-1,-1)$.

At the same time, we cannot have $p_1 = (-1,1)$ or $p_2 = (1,-1)$ because ϕ_0 does not fix these points. We conclude that at our failed parameter u', the points p_1 and p_2 are respectively disjoint from the left and bottom edges of $\big[-1,1\big]^2$. So, ϕ_0 must fix all of L, and L must intersect either the top or the right edge of $\big[-1,1\big]^2$ or both. Figure 4.1 shows the various possibilities, and only the thickly shaded one is relevant.

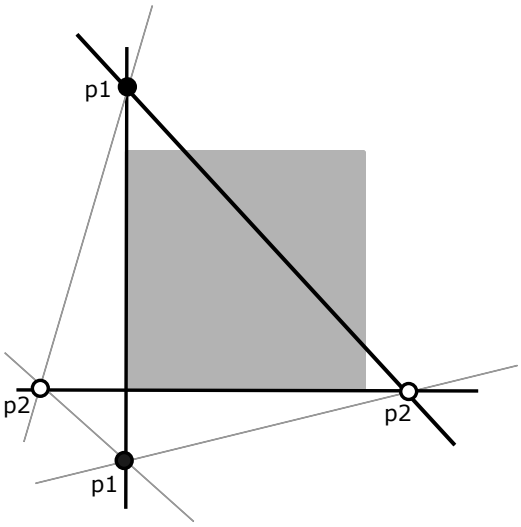


Figure 4.1: The line L and the square $\big[-1,1\big]^2$.

But this is impossible because ϕ_0 fixes no point in these two edges. ♠

4.3 The Remaining Goal

We will produce a closed quadrilateral $\Psi \subset (-1,1)^2$ such that $(0,0) \in \Psi$ and

$$f_k(\Psi) \subset \Psi, \qquad k = 0,1,2,3. \tag{21}$$

Here f_k is the projective transformation on $\mathcal{B}$ realizing the action of the branch k subdivision rule on affine shapes.

By Equation 21, every orbit of $(0,0)$ under the semigroup $\langle f_0, f_1, f_2, f_3 \rangle$ lies in Ψ. The union of these orbits coincides with the set of affine shapes of quads in the full subdivision. Hence the subdivision rule has bounded image in $\mathcal{B}$. This reduces the bounded image property to the existence of Ψ. The rest of the section is devoted to constructing Ψ.

The quad Ψ will be just a slight shrinkage of $[-1,1]^2$. The next two sections discuss how the base maps act on $[-1,1]^2$ and then the final section puts it together and produces Ψ. There is a crude part and a subtle part.

4.4 The Crude Part

Using the method described in the previous chapter, we compute the maps f_k for $k = 0,1,2,3$. We will really only need to consider f_0 and f_1.

- The map f_0 fixes the points $v_3 = (-1,1)$ and $v_0 = (-1,-1)$.
- The map f_1 fixes the points $v_0 = (-1,-1)$ and $v_1 = (1,-1)$.

Thus both f_0 and f_1 fix v_0.

Lemma 4.3 *f_0 maps the vertices $(1,-1)$ and $(1,1)$ into the interior of $[-1,1]^2$.*

Proof: This proof works for all subdivision rules. The map f_0 is a well-defined projective transformation that maps $\mathcal{B} = (-1,1)^2$ into itself. Since f_0 fixes the left two vertices of $[-1,1]^2$ the only way our lemma could fail is if $f_0(1,1)$ lies in the top or right edge of $[-1,1]^2$ or $f_0(1,-1)$ lies in the bottom or right edge.

To rule out the right edge case for $(1,1)$ we compute

$$[f_0(1,1)]_1 - 1 = \frac{2(u_0-1)(u_3+1)\langle u_1,u_2,u_3\rangle}{D_{01}}.$$

Here we are using the notation from Equation 16. Also $[...]_1$ denotes the first coordinate. So the numerator is negative. Since $f_0(1,1)$ is a well-defined point D_{01} never

vanishes, no matter what the choice of u. We conclude that the first coordinate of $f_0(1,1)$ is never 1. Since $f_0(1,1) \in [-1,1]^2$, the first coordinate is less than 1. Next, we have:

$$[f_0(1,1)]_2 - 1 = -\frac{2(1-u_0)(1-u_3)\langle u_3, u_0, u_1\rangle}{D_{02}}.$$

The rest of the argument is the same in this case. The other cases are similar, and actually follow from relabeling and dihedral symmetry. ♠

4.5 The Subtle Part

Each base map violently pushes in two of the corners of $[-1,1]^2$. Now we will explain how they push in the other two corners in a more subtle way. This part uses the tameness property. Let v be a vertex of $[-1,1]^2$. We will first explain the concept in general and then we will specialize to the case at hand. Let $\llcorner$ denote a quadrant in $\boldsymbol{R}^2$. Each vector $V \in \llcorner$ defines a cone $\llcorner_V$ based at the endpoint of V and parallel to $\llcorner$. We simply translate $\llcorner$ to V. Suppose T is a linear transformation so that $T(\llcorner) \subset \llcorner$. We say that V is an *expanding vector* for T if $T(\llcorner_V) \subset \text{interior}(\llcorner_V)$. We insist that V points strictly into $\llcorner$. Figure 4.2 shows what we have in mind.

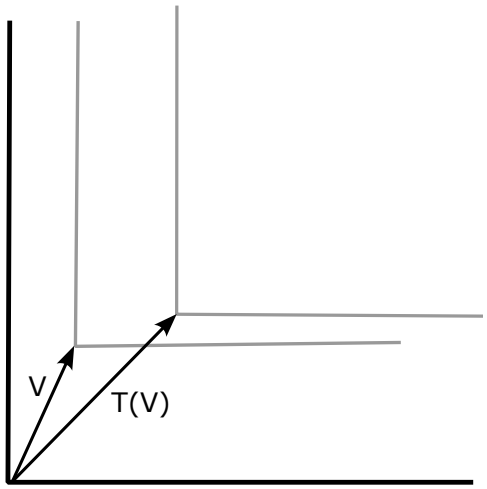


Figure 4.2: An Expanding vector

What we are saying is that T pushes $\llcorner_V$ strictly inside itself. When $\llcorner$ is the positive quadrant, we are saying that each coordinate of $T(V)$ is strictly larger than the corresponding coordinate of V. This notion is a generalization of the idea of an expanding eigenvector: V and $T(V)$ need not be parallel. Note that V is an expanding vector for T if and only if λV is for $\lambda > 0$.

Lemma 4.4 *Let $v_0 = (-1,-1)$. Let $\llcorner$ denote the positive quadrant, considered as a quadrant in the tangent plane $T_{v_0}(\boldsymbol{R}^2)$. Then $\llcorner$ has a simultaneous expanding vector for both differentials $df_0|_{v_0}$ and $df_1|_{v_0}$.*

Proof: We will start with a vector $(1,s)$. The conditions that $(1,s)$ be an expanding vector for each differential will place a condition on s. We will show that we can choose $s > 0$ to satisfy both conditions. We first define

$$\langle a,b,c\rangle = 2 + a - c + ab + bc = (1+a)(1+b) + (1-b)(1-c) > 0. \tag{22}$$

This positivity only requires $a,b,c \in (-1,1)$. Next, we define:

$$A_0 = (u_3 - u_1) + (u_0 - u_2) + (u_0u_1 - u_2u_3).$$

$$A_1 = (u_3 - u_1) - (u_0 - u_2) + (u_0u_3 - u_1u_2).$$

$$C_0 = \left(\frac{1+u_0}{1-u_0}\right)\left(\frac{\langle u_2,u_3,u_0\rangle}{A_0}\right) > 0.$$

$$C_1 = \left(\frac{1+u_0}{1-u_0}\right)\left(\frac{A_1}{\langle u_0,u_1,u_2\rangle}\right) > 0.$$

We mean to define C_j only when $A_j > 0$. We compute that

$$C_0 - C_1 = \left(\frac{1+u_0}{1-u_0}\right)\left(\frac{\langle u_3,u_0,u_1\rangle\langle u_1,u_2,u_3\rangle}{A_0\langle u_0,u_1,u_2\rangle}\right) > 0. \tag{23}$$

There are 4 cases.

1. If $A_0 \le 0$ and $A_1 \le 0$, we choose any $s > 0$.
2. If $A_0 > 0$ and $A_1 \le 0$ we choose any $s \in (0, C_0)$.
3. If $A_0 \le 0$ and $A_1 > 0$ we choose any $s \in (C_1, \infty)$.
4. If $A_0 > 0$ and $A_1 > 0$ we choose any $s \in (C_1, C_0)$.

The various positivity results above show that each case allows us a nonempty set of choices for s. What remains is to show in each case that our choices lead to expanding vectors.

We consider the differentials in turn. First set $T = df_0|_{(-1,-1)}$. We compute

$$T = \begin{pmatrix} \frac{\langle u_2,u_3,u_0\rangle}{(1-u_0)(1+u_3)} & 0 \\ \frac{(1+u_0)\langle u_2,u_3,u_0\rangle}{(1-u_0)\langle u_3,u_0,u_1\rangle} & \frac{\langle u_2,u_3,u_0\rangle}{\langle u_3,u_0,u_1\rangle} \end{pmatrix}.$$

We compute

$$[T(1,s)-(1,s)]_1 = \frac{[u_2,u_3,u_0]}{(1-u_0)(1+u_3)}.$$

$$[T(1,s)-(1,s)]_2 = \frac{(1+u_0)\langle u_2,u_3,u_0\rangle}{(1-u_0)\langle u_3,u_0,u_1\rangle} - \frac{A_0 s}{\langle u_3,u_0,u_1\rangle} =^* \frac{A_0}{\langle u_3,u_0,u_1\rangle}(C_0 - s).$$

The first expression is positive by the tameness condition. When $A_0 \le 0$ any choice of $s > 0$ makes the second equation positive. The starred equality makes sense when $A_0 > 0$, and shows, in this case, that any $s \in (0, C_0)$ makes the expression positive.

Now we set $T = df_1|_{(-1,-1)}$.

$$T = \begin{pmatrix} \frac{\langle u_0,u_1,u_2\rangle}{\langle u_3,u_0,u_1\rangle} & \frac{(1-u_0)\langle u_0,u_1,u_2\rangle}{(1+u_0)\langle u_3,u_0,u_1\rangle} \\ 0 & \frac{\langle u_0,u_1,u_2\rangle}{(1+u_0)(1-u_1)} \end{pmatrix}.$$

We compute

$$[T(1,s)-(1,s)]_2 = \frac{s[-u_2,-u_1,-u_0]}{(1+u_0)(1-u_1)}.$$

$$[T(1,s)-(1,s)]_1 = \frac{(1-u_0)\langle u_0,u_1,u_2\rangle s}{(1+u_0)\langle u_3,u_0,u_1\rangle} - \frac{A_1}{\langle u_3,u_0,u_1\rangle} =^* \frac{(1-u_0)\langle u_0,u_1,u_2\rangle}{(1+u_0)\langle u_3,u_0,u_1\rangle}(s - C_1)$$

The first expression is positive for any $s > 0$ by the tameness condition. When $A_1 \le 0$ any choice of $s > 0$ makes the second expression positive. The starred equality makes sense when $A_1 > 0$, and shows, in this case, that any $s \in (C_1, \infty)$ makes the expression positive.

This analysis shows that all of our cases above produce expanding vectors. ♠

4.6 The End of the Proof

Now we prove Theorem 1.2 for tame subdivision rules. The only thing we have left to check is that there exists a quad $\Psi \subset (-1,1)^2$ such that $f_k(\Psi) \subset \Psi$ for $k = 0,1,2,3$. This gives us the bounded image result for tame subdivisions.

Let us first recast Lemma 4.4 for the other indices. The *tangent cone* to a quad at a vertex is just the cone obtained by extending the sides of the quad that are incident to the vertex.

Lemma 4.5 *Let $k \in \{0,1,2,3\}$. Let v_k be the kth vertex of $[-1,1]^2$. Then the tangent cone to $[-1,1]^2$ at v_k has a vector ζ_k for which $df_k|_{v_k}$ and $df_{k+1}|_{v_k}$ are simultaneously expanding. Indices are taken mod 4.*

Proof: This follows from Lemma 4.4 together with rotational symmetry and cyclic relabeling. ♠

Now for the main construction. We fix $\epsilon > 0$. We choose ζ_k to be a unit vector for all k. We then let q_k be such that $q_k - v_k = \epsilon\zeta_k$. In other words, we push v_k by ϵ units along the direction determined by ζ_k. We let Ψ be the quadrilateral whose vertices are q_k. We claim that $f_k(\Psi) \subset \Psi$ for ϵ sufficiently small. Recall that each base map f_j is a projective transformation. By convexity we just have to check that $f_j(q_k) \in \Psi$ for $j,k = 0,1,2,3$. We will show this for f_0. The cases $j = 1,2,3$ are the same, and indeed follow from the case $j = 0$ and rotation and cyclic relabeling.

The vertices q_1 and q_2 are very near the vertices of $[-1,1]^2$ on the right edge of $[-1,1]^2$. By Lemma 4.3, the map f_0 carries v_1 and v_2 into $(-1,1)^2$. So, there is a single compact region $K \subset (-1,1)^2$ such that $f_0(q_k) \in K$ for $k = 1,2$ once ϵ is sufficiently small. If we then further shrink ϵ so that $K \subset \Psi$, then we get $f_0(q_k) \in \Psi$ for $k = 1,2$.

Now we show that $f_0(q_0) \in \Psi$ for $\epsilon > 0$ sufficiently small. We will assume that this fails for a sequence $\{\epsilon_n\}$ converging to 0 and derive a contradiction. We will just write ϵ in place of ϵ_n. Now we write Ψ_ϵ to denote the dependence on ϵ.

Let

$$\Delta_\epsilon(z) = v_0 + \epsilon z. \tag{24}$$

This map shrinks distances by a factor of ϵ and carries the origin to v_0. Since f_0 is differentiable (and in fact analytic) we have

$$\lim_{\epsilon \to 0} \Delta_\epsilon^{-1} \circ f_0 \circ \Delta_\epsilon = df_0|_{v_0}. \tag{25}$$

It is convenient to let $f_{0,\epsilon}$ be the composition on the left side of Equation 25. Then

$$f_{0,\epsilon}(\zeta_0) \notin \text{interior}(\Delta_\epsilon^{-1}(\Psi_\epsilon)), \tag{26}$$

for all $\epsilon = \epsilon_n$ sufficiently small. This is a closed condition. Moreover, the sets on the right hand side of this equation converge, uniformly on compact sets, to the translated sector $\llcorner_{\zeta_0}$. Taking a limit as $\epsilon_n \to 0$ we see that

$$df_0|_{v_0}(\zeta_0) \notin \text{interior}(\llcorner_{\zeta_0}). \tag{27}$$

But this contradicts the fact that ζ_0 is expanding for $df_0|_{v_0}$. This contradiction proves what we want.

Finally, for v_3 we run the same argument, using the fact that ζ_3 is expanding for $df_0|_{v_3}$.

This completes the proof that $f_0(\Psi) \subset \Psi$ for $\epsilon > 0$ sufficiently small.

This proves that every tame subdivision rule is affinely contained.

5 The Reconstruction Theorem

5.1 The Order of Operations

In this chapter we prove the Reconstruction Theorem. The most mysterious statement in the theorem is the reversal of the 4-ary sequence indexing the quads in the iterated subdivision. Here we discuss this phenomenon in a toy case to see where it comes from.

Consider the situation where we iteratively divide the interval $I = [0,1]$ in half. There are two ways to think about doing this. Let S_0 and S_1 be the two subdivision operations which cut an interval in half and replace it respectively by the left and right halves. Let T_0 and T_1 be the two maps defined by $T_0(x) = x/2$ and $T_1(x) = 1/2+x/2$. Note that $T_j(I) = S_j(I)$ for $j = 0, 1$.

As an experiment, imagine that we compare the compositions $S_1 \circ S_0$ and $T_1 \circ T_0$. In both these cases we mean "first apply X_0 and then apply X_1", where $X \in \{S, T\}$. When we do $S_1 \circ S_0$ we first get $[0, 1/2]$ and then we get $[1/4, 1/2]$. When we do $T_1 \circ T_0$ we first get $[0, 1/2]$ and then we get $[1/2, 3/4]$. We do not end up with the same intervals. On the other hand, if we do $S_1 \circ S_0$ we get the same result as $T_0 \circ T_1$. In general, the subdivision method of generating the intervals is the reverse of the mapping method.

5.2 Formulas for the Symmetric Subdivisions

Here we consider the symmetric subdivision based on the parameter (u, u, u, u). We always take $u \in (-1, 1) - \{0\}$. The cases $+u$ and $-u$ are essentially the same. We just present the formulas for the 0th branch. This skew product is given by matrices f and $h_{a,b}$.

$$f = \begin{pmatrix} 1+u^2 & 0 & 2u^2 \\ 2u & 1-u^2 & 2u \\ 1+u^2 & 0 & 2 \end{pmatrix}. \tag{28}$$

$$h_{a,b} = \begin{pmatrix} -2(b^2-1)(2+a+au^2) & 2(a^2-1)(b-2u+bu^2) \\ 0 & (a-1)(2+a+b+au-bu)(-2-a+b+au+bu) \end{pmatrix}. \tag{29}$$

The other branches are conjugate to this one *via* the map in Equation 13. In particular the fiber maps for branches 0 and 2 are always given by upper triangular matrices and the fiber maps for branches 1 and 3 are given by lower triangular matrices.

The matrix f is always nonsingular: We have

$$\det(f) = 2(u^2-1)^2(u^2+1) > 0.$$

Here is the eigensystem for f.

$$\begin{array}{c|c} 1-u^2 & (-1,-1,1) \\ 1-u^2 & (-1,+1,1) \\ 2(1+u^2) & (2u^2,2u,1+u^2) \end{array}$$

Acting as a projective transformation, f maps $[-1,1]^2$ into itself and pointwise fixes the left edge $\{-1\} \times [-1,1]$.

We consider (x,y,z) as a point in the affine patch by taking $(x/z, y/z)$ as usual. Considered as a point in the affine patch, the third eigenvector travels from $(0,0)$ to $(1,1)$ along the circle of radius 1 centered at $(1,0)$ as u moves from 0 to 1. The third eigenvalue is largest and so f is repelling on the left edge of $[-1,1]^2$. This fact will become important for Theorem 1.1 in the next chapter.

For Theorem 1.3 we will not need to analyze $h_{a,b}$. The entire skew product, however, is probably quite beautiful.

5.3 The Geometric Coincidence

Now we augment our notation and let $f_{k,u}$ be the projective transformation associated to Branch k for the parameter u. Let $b_{k,u} \in (-1,1)^2$ be the point represented by the third eigenvector of $f_{k,u}$, as above. Let B_u be the square whose vertices are $b_{k,u}$ for $k = 0,1,2,3$. This is the square shown on the right side of Figure 1.9. We call B_u the *pinwheel square*.

Let $\phi_{k,u}$ be the projective transformation that carries $[-1,1]^2$ to the kth quad in its u-subdivision. We produce our initial quads by iterating the semigroup generated by the ϕ maps. We produce the locations of these quads in $\mathcal{B}$ by iterating the f maps. Here is the algebraic coincidence that powers Theorem 1.3.

Lemma 5.1 (Parallel) *The label-preserving similarity Φ that maps B_u to $[-1,1]^2$ conjugates $f_{k,u}$ to $\phi_{k,-u}$.*

Proof: Note the minus sign. This is deliberate. Our result is a direct calculation, which we do in Mathematica. ♠

5.4 The End of the Proof

The rest of the proof of the Reconstruction Theorem is just bookkeeping. We combine the Parallel Lemma with the reversal phenomenon discussed above. We suppress the dependence on the parameter u. Our proof is meant to work for any $u \neq 0$ in $(-1,1)$. Recall that $\alpha(Q) \in \mathcal{B}$ is the affine shape of the quad Q.

Let σ be some finite 4-ary sequence. Let Q_σ be the quad whose address is σ. Writing $\sigma = \sigma_0 ... \sigma_k$ we have

$$Q_\sigma = \Sigma_{\sigma_k} \circ ... \circ \Sigma_{\sigma_0}([-1,1]^2). \tag{30}$$

The rightmost digit of the address records the last subdivision operation. By our skew product result,

$$\alpha(Q_\sigma) = f_{\sigma_k} \circ ... \circ f_{\sigma_0}(0,0). \tag{31}$$

Here $(0,0) = \alpha([-1,1]^2)$ is the affine shape of the square and f_j is the base map associated to Branch j of the subdivision process.

Let $\tilde{\phi}_j$ be the projective transformation that maps $[-1,1]^2$ to the jth quad in the $(-u)$ subdivision. By the Parallel Lemma,

$$f_j = \Phi^{-1} \circ \tilde{\phi}_j \circ \Phi. \tag{32}$$

Since $\Phi((0,0)) = (0,0)$, we have

$$\alpha(Q_\sigma) = \Phi^{-1} \circ \tilde{\phi}_{\sigma_k} \circ ... \circ \tilde{\phi}_{\sigma_0}(0,0). \tag{33}$$

By projective naturality, the map

$$\Phi^{-1} \circ \tilde{\phi}_j \circ \Phi$$

is precisely the projective transformation associated to Branch j of the $(-u)$ subdivision of the pinwheel square B_u.

Now comes the reversal. The composition

$$(\Phi^{-1} \circ \tilde{\phi}_{\sigma_k} \circ \Phi) \circ ... \circ (\Phi^{-1} \circ \tilde{\phi}_{\sigma_0} \circ \Phi)$$

maps B_u to the quad $B_{\overline{\sigma}}$ whose address is the reverse of σ. This is the reversal phenomenon discussed in §5.1.

Finally, $(0,0)$ is the projective center of B_u, and projective transformations preserve projective centers. Hence $\alpha(Q_\sigma)$ is the projective center of $B_{\overline{\sigma}}$. This proves Theorem 1.3.

6 Some Structure of the Parameter Space

6.1 The Apices

Generically, the three fixed points of each base map f_k form a triangle, with two of the vertices being consecutive corners of $[-1,1]^2$. We call the third fixed point a_k, when it exists, the *apex*. Referring to Theorem 1.3, the pinwheel square B_u is the quadrilateral in $\mathcal{B}$ obtained by connecting the apices in cyclic order. The 4 images of B_u under the base maps turn out to exactly tile the pinwheel square, and their shapes are the same as in the subdivision rule associated to the parameter $-u$. This is an algebraic coincidence having to do with the action of the base maps on the apices. Most notably,

$$f_1(a_3) = f_3(a_1) = f_0(a_2) = f_2(a_0).$$

For a general subdivision rule, it turns out that these coincidences fail, but nevertheless there is an analog of the pinwheel square.

Lemma 6.1 *A subdivision rule u is tame iff the four apices exist and lie in $(-1,1)^2$.*

Proof: In the tame case, the formulas in §2.5 say that the apices are attracting fixed points of the base map. In particular, they exist. Hence the limit set Λ_u contains these apices. But $\Lambda_u \subset (-1,1)^2$ because, as we have already established, tame subdivision rules are affinely contained.

Now consider the case when u is not tame. Then one of the 8 inequalities in Equation 2 fails. After dihedrally relabeling we can assume that $[u_2, u_3, u_0] \le 0$. We consider the base map f_0. Let $a_0 = (a_{01}, a_{02})$ be the apex for f_0. We have

$$a_{01} + 1 = \frac{2[-u_1, -u_0, -u_3][u_2, u_3, u_0]}{D}, \tag{34}$$

$$1 - a_{02} = \frac{2(1-u_0)(1-u_3)[-u_1, -u_0, -u_3]}{D}, \tag{35}$$

$$a_{02} + 1 = \frac{2(1+u_0)(1+u_3)[u_2, u_3, u_0]}{D}. \tag{36}$$

Here D is a polynomial which one can deduce from the fact that the last two equations sum to 2. These are valid as long as D is nonzero. If $D > 0$ then $a_{02} \le -1$ and $a_0 \notin (-1,1)^2$. If $D < 0$ then the product of the three expressions is non-positive, and this would be impossible if $a_0 \in (-1,1)^2$. If $D = 0$ then the vector representing a_0 in homogeneous coordinates has third coordinate 0. This is not possible if a_0 is a well-defined point in $(-1,1)^2$. ♠

6.2 The Boundary Case of Theorem 1.2

In this section we finish the proof of Theorem 1.2 by considering subdivision rules which are neither wild nor tame. These are limits of tame subdivision rules.

Let us study the transition from the tame case to the non-tame case. One of the apices must exit every compact subset of $(-1,1)^2$ as our subdivision rule exits the tame locus. Suppose the apex a_0 has this property. If the polynomial D does not vanish in the limit, then at least two of the expressions in the lemma vanish. But this means that a_0 converges either to $(-1,-1)$ or to $(-1,1)$. On the other hand, if D vanishes in the limit then we must have

$$[u_2, u_3, u_0] = [-u_1, -u_0, -u_3] = D = 0.$$

In this case, f_0 simplifies to

$$\begin{pmatrix} 1-u_1+2u_3 & 0 & -1-u_1 \\ -1+u_1+2u_3 & 2(1+u_3) & -1+u_1+2u_3 \\ 1+u_1 & 0 & 3+u_1+2u_3 \end{pmatrix} \tag{37}$$

This matrix is not diagonalizable, and indeed has no fixed points off the line $x = -1$. In this case, a_0 ceases to exist and is "absorbed" into the left edge of $[-1,1]^2$. So in all cases, what is happening is that a_0 accumulates on the left edge of $[-1,1]^2$. The triangle of fixed points of f_0 collapses into a line segment.

The degenerate case just described happens when both constraints above vanish. My interpretation is like this: Each of the vanishing constraints pushes a_0 towards one of the two left vertices of $[-1,1]^2$. If both constraints vanish then a_0 is somehow just pushed into the edge. This is the classic situation one sees in algebraic geometry, when a blow-up is useful. We will not pursue this direction here, but let us use the analysis above to show that the remaining non-tame subdivision rules are not affinely contained.

After cyclically relabeling we can assume f_0 has no apex in $(-1,1)^2$. But then there are three cases.

- $(-1,-1)$ is an attracting fixed point for f_0.
- $(-1,1)$ is an attracting fixed point for f_0.
- f_0 is a non-identity parabolic map which fixes the line $x = -1$ pointwise.

In the first two cases, the f_0-iterate of any point in $(-1,1)^2$ converges to the attracting fixed point. In the third case, the f_0 iterate of any point in $(-1,1)^2$ converges to the left edge of $[-1,1]^2$. So, in all cases, the subdivision rule is not affinely contained.

6.3 The Eigenquad

Now we consider the tame case again. Given a tame subdivision rule u we define the *eigenquad* B_u to be the quadrilateral obtained by connecting the apices in cyclic order. This is a generalization of the pinwheel square.

Lemma 6.2 *Let $u = (u_0, u_1, u_2, u_3)$ be a tame subdivision rule. The following is true.*

- *If $u_0 + u_2 \neq 0$ and $u_1 + u_3 \neq 0$ then B_u is embedded.*
- *If $u_0 + u_2 = 0$ or $u_1 + u_3 = 0$, but not both, then B_u is a line segment.*
- *If $u_0 + u_2 = 0$ and $u_1 + u_3 = 0$ then B_u is a point.*

Proof: The main diagonals of $(-1,1)^2$ divide it into four open triangles. We call these triangles *sectors*.

Let a_0, a_1, a_2, a_3 be the apices. We compute

$$\det\big(a_0, (-1,-1,1), (1,1,1)\big) = \frac{4(u_0-1)(u_1+u_3)[u_2,u_3,u_0]}{D}, \tag{38}$$

$$\det\big(a_0, (-1,1,1), (1,-1,1)\big) = \frac{4(u_0+u_2)(1+u_3)[-u_1,-u_0,-u_3]}{D}. \tag{39}$$

$$D = (1+u_0)(1+u_3)[u_2,u_3,u_0] + (1-u_0)(1-u_3)[-u_1,-u_0,-u_3]. \tag{40}$$

These equations show in the tame case that a_0 lies in a sector as long as $u_0+u_2 \neq 0$ and $u_1 + u_3 \neq 0$. Similar equations hold for the other apices, and follow from cyclic reordering. Moreover, by looking carefully at the signs (or checking in representative cases) we see that the sectors containing a_0, a_1, a_2, a_3 are ordered counter-clockwise. This guarantees that B_u is embedded when $u_0 + u_2 \neq 0$ and $u_1 + u_3 \neq 0$.

When $u_0 + u_2 = 0$ the equations above, and their cyclic permutations, say that all four apices lie in one of the diagonals. Hence B_u is a line segment in this case. The same goes when $u_1 + u_3 = 0$. ♠

The four base map images $f_j(B_u)$, in general, do not quite fit together, and they also are not necessarily contained in the original eigenquad. Figure 6.1 shows an example which is fairly near the symmetric case.

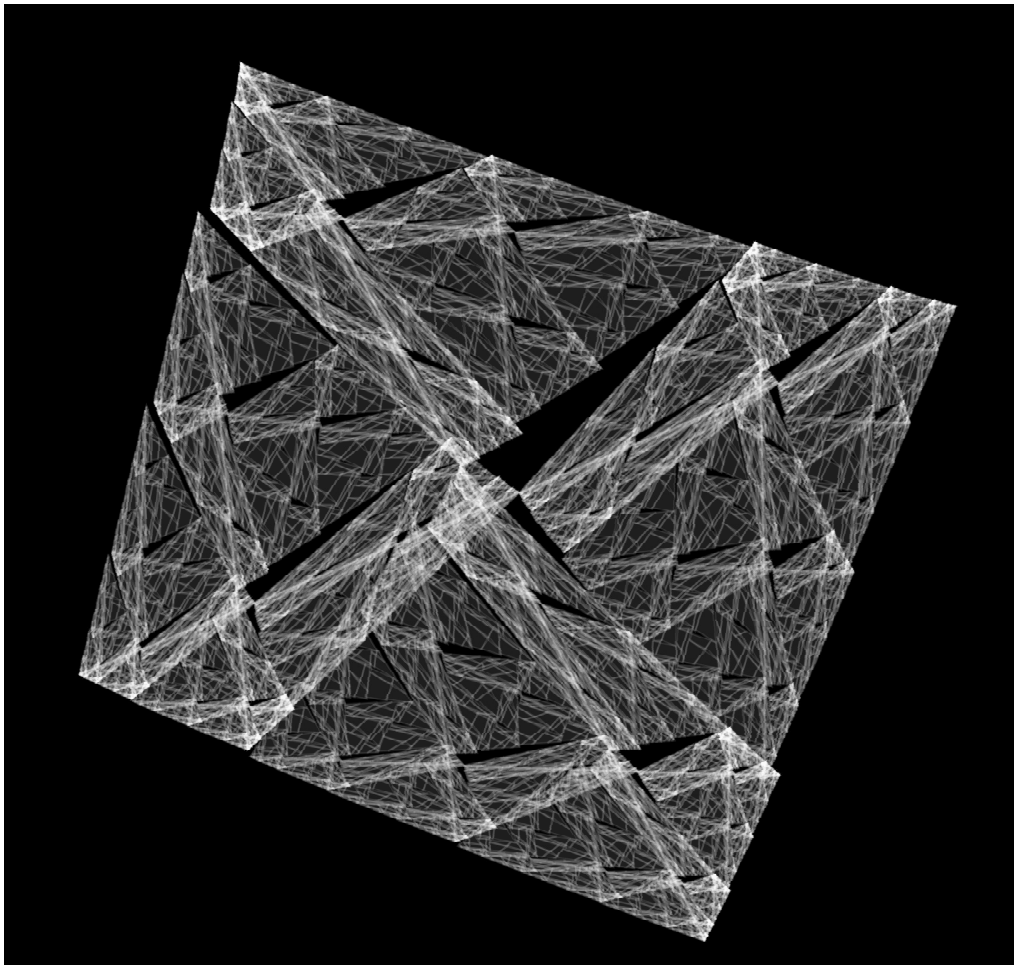


Figure 6.1: A near tiling

Nevertheless they give a good sense of the limit set. If we take a large number of eigenquad images under the action of the semigroup and then give them a nearly translucent color, then the places which have the most color to them – i.e. the points contained in many eigenquad images – give a good approximation of the limit set. This is what Figures 1.7, 1.8, 6.1, and 6.3 show. I think that the complicated pattern of overlaps between the first 4 eigenquad images ought to force the limit set to be connected in the tame case, but I have not thought this through.

6.4 The Coherent Locus

The tame locus S is cut out by quadratic inequalities. It is neither convex nor a polytope. However, there is a lot of structure. Consider the splitting $\boldsymbol{R}^4 = \boldsymbol{R}^2 \times \boldsymbol{R}^2$

corresponding to the pairs of variables (u_0, u_2) and (u_1, u_3). Every planar slice obtained by fixing either pair intersects S in an open convex polygon.

To explain the picture in more detail, we "prefer" one factor in the splitting: We think of $(-1,1)^4$ as a fiber bundle over the $(-1,1)^2$, with the base space given by the variables (u_0, u_2) and the fiber variables given by (u_1, u_3). In each fiber, the constraints all have the form $A + Bu_1 + Cu_3 = 0$. Thus they define lines in each fiber. The tame locus is the component of the complement of these lines which contains the origin in each fiber.

Lemma 6.3 *The two constraints* $[u_{k-1}, u_k, u_{k+1}] = 0$ *and* $[-u_{k+1}, -u_k, -u_{k-1}] = 0$ *never intersect the same fiber.*

Proof: By relabeling symmetry it suffices to check two constraint pairs in a row. Equivalently, it suffices to consider both ways of fibering $(-1,1)^4$ and then check one constraint pair on both fibrations. We prefer this approach. We change variables so as to work with a, b, c, d. We look at the pair $[a, b, c]$ and $[-c, -b, -a]$. We compute

$$[a, b, c] = (1 - b)(1 - c) + (1 + b)(a + c) > a + c$$

Thus $[a, b, c] > 0$ when $a + c \geq 0$. Likewise $[-c, -b, -a] > 0$ when $a + c \leq 0$. Hence, one or the other constraint line avoids any (a, c) fiber.

Now suppose we hold (b, d) fixed and vary a and c. Note first that $[a, 0, c] = 1 + a > 0$. When $b < 0$ we have

$$[a, b, c] > 1 + a - b + ab + 2b = (1 + a)(1 + b) > 0.$$

Similarly when $b > 0$ we have $[-c, -b, -a] > 0$. Hence, one or the other constraint line avoids the (b, d) fiber. ♠

From Lemma 6.3, we see that at most 4 constraint lines intersect each fiber. Now we make an experimental observation that comes from plotting everything in sight. The base space is divided into 8 open triangles by the lines of slope 0, ± 1, and ∞. In each open triangle, the intersection pattern of the constraint lines does not change, and the tame locus S intersects each fiber in a convex pentagon.

How does this happen from at most 4 constraint lines? Well, we get additional *side conditions* from the sides of the square fiber: $u_1, u_3 \in [-1, 1]$. Three or four constraint lines combine with one or two side conditions to give the pentagon in all 8 cases. See Figure 6.2 for three representative cases. The tame locus is the darkly shaded pentagon. The black dot indicates the point in the base space over which the fiber lives. Our pictures are doing double duty here, showing structures both in the base and the fiber.

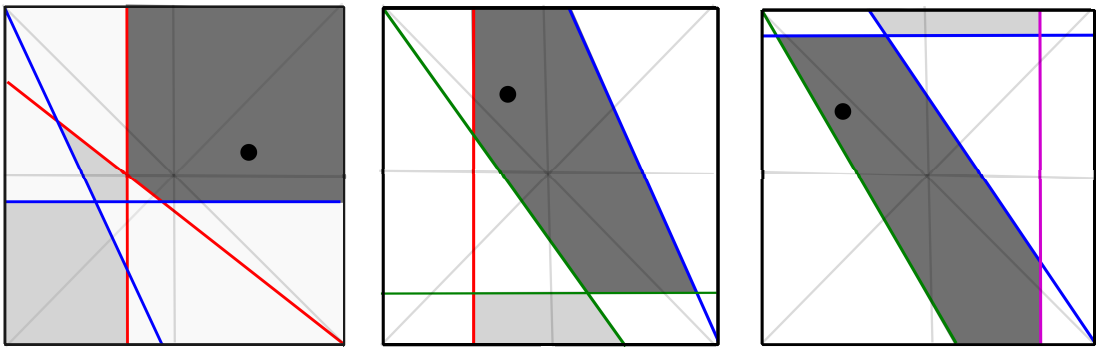


Figure 6.2: The constraint lines and the domains they define.

The edge cases along the lines of slope ± 1 are typically quadrilaterals, and the edge cases along the lines of slope 0 and ∞ are pentagons. All of this structure inspired me to call S a *bi-convex bi-pentagon*.

In view of the discussion above, each constraint forces one of the apices into a corner as it vanishes. So, we can color the 8 constraint lines according to which apex it is associated to. Guided by this combinatorics, we can multiply together the constraints which correspond to the same vertex.

$$[u_{k+0}, u_{k+1}, u_{k+2}][-u_{k-1}, -u_{k-2}, -u_{k-3}] > 0. \tag{41}$$

Here $k = 0, 1, 2, 3$ with indices taken mod 4.

In the generic fiber, the locus of points satisfying these constraints seems to be a collection of 2 or 3 convex polygons one of which is the tameness polygon. The closures of these polygons are either disjoint or meet "kitty corner" across two crossing lines of the same color. Let us call this set the *coherent locus*. So, the coherent locus contains the tame locus. In Figure 6.2, the shaded regions together comprise the intersection of the coherent locus with the given fiber. The lightly shaded regions are coherent but not tame.

The limit sets of subdivision rules in the coherent locus seem to be quite nice. Figure 6.3 shows an example.

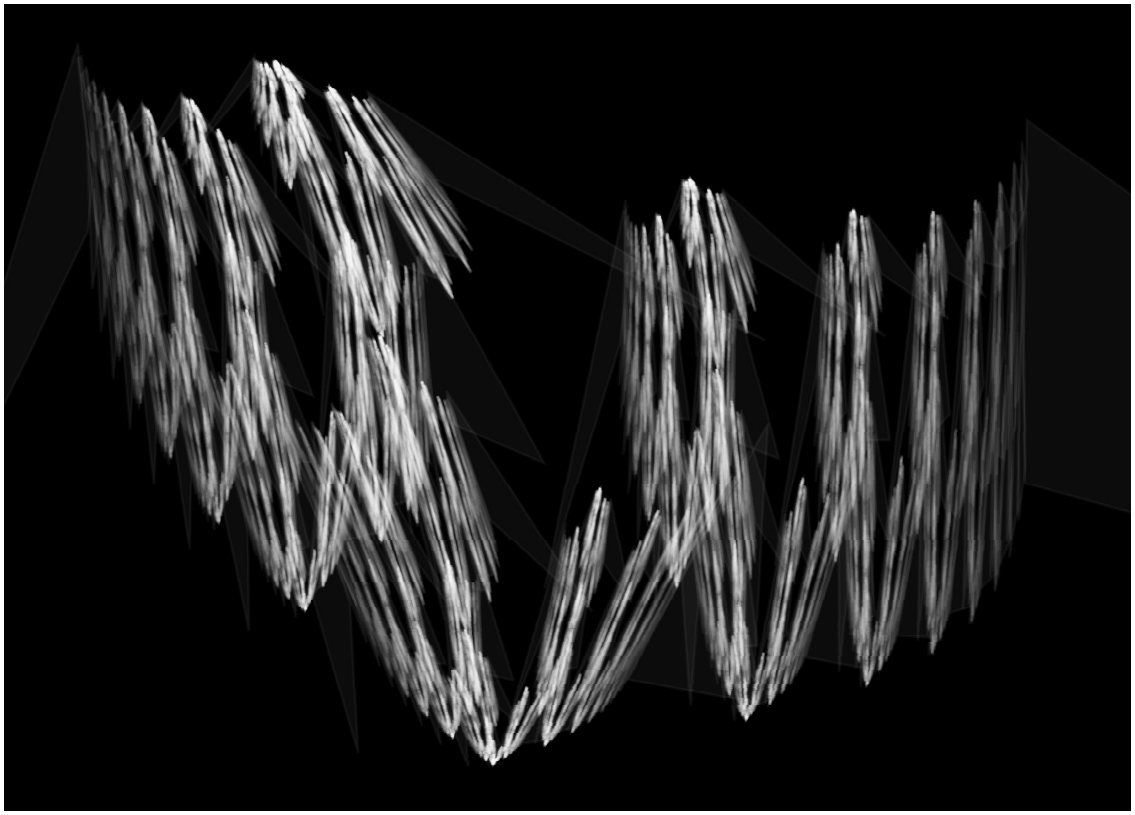


Figure 6.3: A non-tame limit set from the coherent locus

This example has two opposite apices outside the base space. Other examples have just one apex outside the base space. The interested reader can see these limit sets with my software, or else draw them on their own.

This is all I can say about these limit sets, tame or otherwise. I don't really have a good understanding of their structure yet.

7 References

[**F**] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd edition, John Wiley & Sons, 2014.

[**McM**] C. McMullen, *The Hausdorff dimension of general Sierpinski carpets*, Nagoya Math. J. **96** (1984), 1–9.

[**S1**] R. E. Schwartz, *Pappus's Theorem and the Modular Group*, Publ. Math. IHES **78** (1993), 187–206.

[**S2**] R. E. Schwartz, *Le Retour de Pappus*, KIAS–Springer Series in Mathematics (2025). See also arXiv:2412.02417.

[**W**] Wolfram Research, Inc., *Mathematica*, Version 15.0, Champaign, Illinois, 2026.