Symmetries of Period-Doubling Maps

Linas Vepstas <linas@linas.org>

12 October 2004 (revised 1 September 2006)

Abstract

The concept of self-similarity is central to the notion of a fractal, but the actual symmetry group that generates that self-similarity is rarely named, and it seems fractals are almost never studied according to their symmetries. Yet, in other branches of mathematics and physics, it is well understood that symmetry provides a powerful mechanism for understanding systems.

In this paper, we identify the symmetry group of period-doubling maps as being a monoid (semigroup) of the modular group $PSL(2,\mathbb{Z})$. To anchor this assertion, we work out an explicit, exactly-solvable fractal curve, the Takagi or Blancmange Curve, as transforming under the three-dimensional representation of the (monoid of the) modular group. By replacing the triangular shape that generates the Blancmange curve with a polynomial, we find that the resulting curve transforms under the $n+2$ dimensional representation of the monoid, where n is the degree of the polynomial. We also find that the (ill-defined) derivative of the Blancmange curve is essentially the (inverse of the) Cantor function, thus demonstrating the semigroup symmetry on the Cantor Set as well. In fact, any topologically conjugate map will transform under the three-dimensional representation. We then show how all period-doubling maps can demonstrate the monoid symmetry, which is essentially an outcome of the dyadic representation of the monoid.

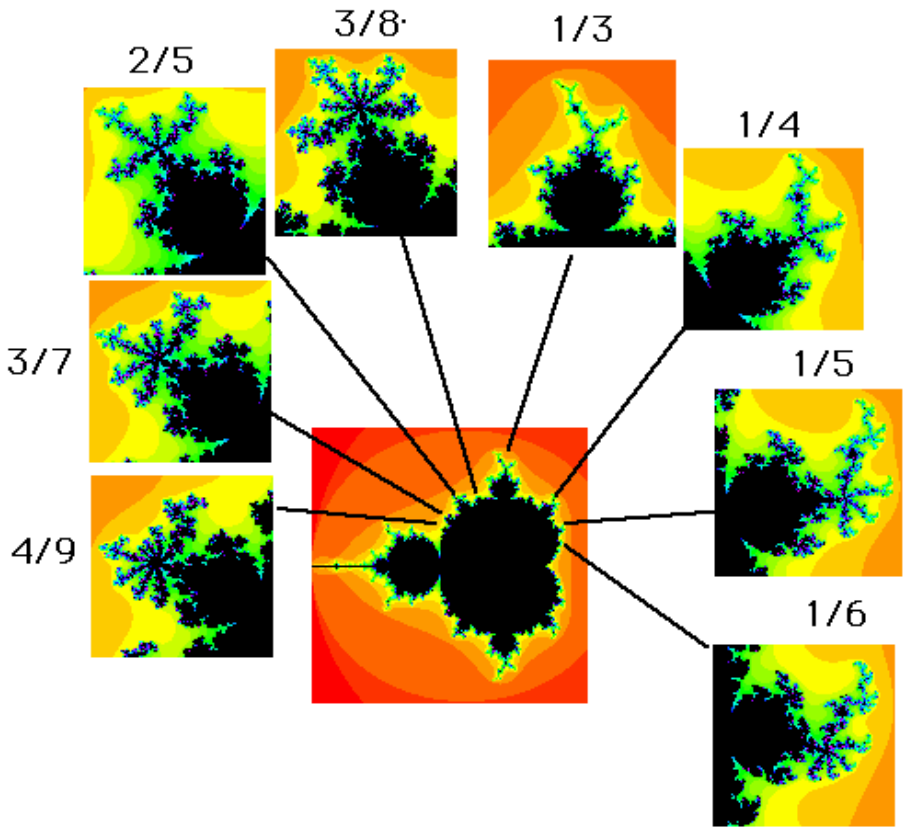
This paper also includes a review of Georges deRham's 1958 construction of the Koch snowflake, the Levy C-curve, the Peano space-filling curve and the Minkowski Question Mark function as special cases of a curve with the monoid symmetry. The left-right symmetric Levy C-curve and the left-right symmetric Koch curve are shown each belong to another, inequivalent three dimensional representation.

This paper is part of a set of chapters that explore the relationship between the real numbers, the modular group, and fractals. Its also a somewhat poorly structured, written at least partly as a diary of research results.

1 Symmetries of Period-Doubling Maps

It has been widely noticed that Farey numbers appear naturally in certain fractals, most famously in the Mandelbrot set. For example figure 1 shows how to count the buds of the Mandelbrot set by the Farey numbers. The reason why the Farey numbers are appropriate for such counting is somewhat more opaque. However, it can be said that many fractal phenomena, and in particular, period-doubling maps, have an

Figure 1: Mandelbrot Farey Numbering



The above picture is stolen from Robert L. Devaney's website The Fractal Geometry of the Mandelbrot Set `http://math.bu.edu/DYSYS/FRACGEOM2/node5.html`. It demonstrates how to label and count buds on the Mandelbrot set using the Farey fractions. That the term "counting" is appropriate follows from the observation that both the sizes and locations of the buds correspond to the location of the fractions on the Farey or Stern-Brocot tree.

---

infinite binary tree structure that appears naturally. By describing the nature of the self-symmetries of the binary tree, one can effectively describe the nature of the fractal in question.

The infinite binary tree is self-similar, in that one can look at the subtree lying under any given node, and see that is isomorphic to the tree as a whole. A given subtree is easily and uniquely located: one may take a walk down the tree, taking either the left *L* or the right *R* branch. Thus, a given node on the tree may be specified by a sequence or string composed of the letters *L* and *R*. When *L* and *R* are taken to be a certain pair of 2×2 matrices, then resulting set of strings turn out to be equivalent to a certain subset of the modular group $PSL(2,\mathbb{Z})$. The modular group is critically important for a broad swath of number theory, from the theory of elliptic functions to the theory of modular forms. It seems remarkable that the symmetry structure of fractals are connected these topics, although, on closer examination, it turns out that what is remarkable is that the popular literature for the most part has not made this connection. The notable exception seems to be the popular book "Indra's Pearls" by David Mumford and Caroline Series (need ref).

More narrowly, the relationship between number theory and hyperbolic geometry meets in an area of mathematics known as ergodic theory. The group $PSL(2,\mathbb{Z})$ is seen to be a special case of a Fuchsian group, which are generally the discrete subgroups of $PSL(2,\mathbb{C})$. These are implicated in a variety of chaotic dynamical systems, most notably in the ergodic or Anosov flow on the tangent space of $PSL(2,\mathbb{C})$. Thus, the relationships between number theory and dynamical systems are known to mathematics; however, this knowledge has not yet escaped fairly narrow confines. In particular, many theorems in the area of dynamical systems are abstractly stated, and make no mention of exactly-solvable cases. Conversely, many areas of concrete number theory fail to emphasize or even point out the highly fractal nature of the relationships. Thus, it seems appropriate to work out in greater detail some concrete connections between fractals, the infinite binary tree, and the modular group.

The focus of this paper is to elucidate how the infinite binary tree, the set of of all strings generated by two letters *L* and *R*, and the set of period doubling fractals are related, and how their self-similarities can be precisely described. This is done in two parts. An earlier paper (need ref to my Minkowski question mark paper) deals with two-dimensional matrix representations, enumerating the elements of these sets, and questions about measure theory and the relationship to the Cantor set. This paper begins with a brief review of those results. It is then followed by a development of the three-dimensional matrix representation for a very specific fractal, the Takagi curve, and then by various generalizations to broader cases.

The first to be explored are the three-dimensional representations. It is shown that there are an uncountably infinite number of these, and that they may be labelled by a real or complex number. They are all inequivalent, in that there is no similarity transform that takes one into another. These occur naturally as the set of self-symmetries of the a family of fractal curves studied by Teiji Takagi in 1903[7], and further developed by G.H. Hardy in 1916 and by others. These now bear the name Takagi curve or blancmange curve.

The Takagi curve is constructed from the iterated tent map. The three-dimensional representation can thus be shown to apply to any iterated map that is topologically con-

jugate to the tent map. The class of such curves is large, and contains many interesting specimens, such as the Logistic map of unit height. It is this observation that leads to a conjecture that essentially all period-doubling maps transform under the three-dimensional representation. In practice, however, there is a big difference between proving that a map is conjugate to the tent map, and giving explicit expression to the conjugating function. As an example, we show how one very simple iterated map is conjugate to the tent map, but the conjugating function is the Minkowski question mark function. Thus, finding the conjugating function is a non-trivial exercise.

The above introduction singled out some special matrices for L and R. One may ask what happens if one picks some arbitrary matrices for L and R, or even a pair of arbitrary functions, and then iterates on them by constructing all possible strings containing L and R. This question was explored by Georges de Rham in 1957[3]. He demonstrates that under certain broad conditions for the fixed points of L and R, any such generalized pair, when iterated, will produce a continuous but non-differentiable curve. A variety of classic fractal curves arise in this construction, including the Koch snowflake, the Levy C-curve, and the Peano space-filling curve. The fact that all of these curves can be identified with the set of strings in two letters implies that the general considerations about the self-similarity of binary trees can be applied to these cases

2 Two-dimensional Representation Basics

This section reviews the basics of the two-dimensional matrix representations of the self-similarities of the infinite binary tree. It is an abbreviated presentation of the material in xxx (ref my paper on Minkowski question mark), and is here only to provide context and notation for the rest of this document.

A critical property of the Farey numbers is that they can be arranged on a binary tree, known as the Farey tree or Stern-Brocot tree, as shown in figure2. One may navigate to a particular location in the binary tree by starting at the root node (located at $1/1$), and making a sequence of left and right moves. The value of the Farey fraction labelling that node can be obtained by performing a simple matrix calculation. Let

$$L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{1}$$

be the "left" and "right" matrices. A sequence of moves, such as *RLLLRRL*, results in a matrix

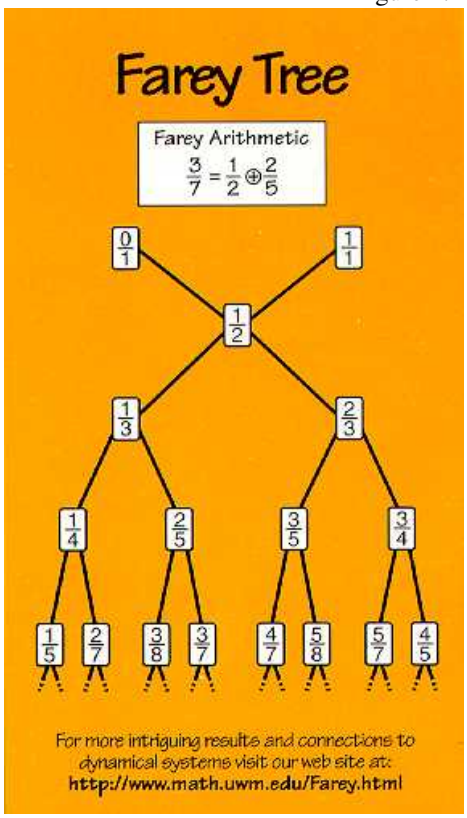
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{2}$$

with integer entries a, b, c, d. The value of the Farey number located at the corresponding node is

$$\frac{a+b}{c+d}$$

This is the simplest example of what I'll call a "matrix representation" associated with a binary tree. This particular representation is two-dimensional (it deals with 2×2

Figure 2: Farey Tree



An illustration of the Farey tree or Stern-Brocot tree, showing the Farey numbers arranged on a binary tree.

matrices), and will be called the "Farey" or "continued fraction" representation, to distinguish it from other possible two-dimensional representations.

The matrix given above is unimodular: that is, its determinant is one. This can be easily seen because the determinant of L and R is one, and the determinant of a product of matrices is equal to the product of the determinants. The entries of the matrix are always integers, and thus its easily seen that the matrix belongs to the group of unimodular 2×2 matrices over the integers $SL(2,\mathbb{Z})$. Given that the Farey fraction is given by the ratio, so that

$$\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

gives the same fraction, we can see that the relevant matrices in fact belong to the *projective* group $PSL(2,\mathbb{Z})$. This group is known as the "Modular Group", and it is central to many areas of number theory and in particular the theory of elliptic integrals and modular forms.

An alternate labelling of an infinite binary tree is with the "dyadic rationals", rational numbers of the form $m/2^n$ with m an odd integer, and n a non-negative integer. This is very simply the tree with rows

$$\begin{array}{cccccccc} & & \frac{0}{1} & & & & \frac{1}{1} & \\ & & & & \frac{1}{2} & & & \\ & \frac{1}{4} & & & & & \frac{3}{4} & \\ \frac{1}{8} & & \frac{3}{8} & & & \frac{5}{8} & & \frac{7}{8} \end{array}$$

This tree, the dyadic rational tree, also has a matrix representation for converting a sequence of left-right moves to the value of the dyadic fraction located at the resulting node. The left and right dyadic matrices are

$$L_D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad R_D = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{3}$$

and these generate what will be called the two-dimensional "dyadic representation" of moves on the binary tree. The value of the dyadic fraction after a sequence of moves is $2c+d$ where c and d are the entries in the matrix shown in 2.

The dyadic matrix that results from a sequence of left and right moves is clearly not unimodular. However, insofar as it labels a set of moves on the binary tree, it is isomorphic to the matrices arising in the Farey representation. There is a (unique) matrix corresponding to each node in the tree, and these can be trivially identified with one another. The representations are not, however "equivalent": there does *not* exist a similarity matrix S that caries the one to the other. That is, the equations

$$SL_D = L_C S$$

and

$$SR_D = R_C S$$

are satisfied only by $S = 0$. Here, we wrote L_C and R_C for the L and R of equation 1, in order to distinguish them from the dyadic L and R.

These two representations are developed and examined in greater detail in another paper (ref my paper on Minkowski question mark), and the details will not be reproduced here; the focus of this paper are the three-dimensional and the higher-dimensional representations. However, there are several important points that must be noted, in order to avoid confusion. Most important is to understand that the moves on the binary tree are *not* isomorphic to the modular group. Although it can be shown that L_C and R_C generate all of $PSL(2,\mathbb{Z})$, the *non-negative* powers do not. The non-negative powers generate only what is called a "monoid" or a "semigroup", and not a full group, because the inverse elements are missing. The monoid does not contain the elements L^{-1} or R^{-1} or any strings containing these elements. Although it may at first seem seem that the negative powers correspond to upward walks on the binary tree, this is not so. The meaning of the negative powers in terms of their action on the tree cannot be made consistent or even complete. Thus, while the monoid can be formally extended to a full group by adjoining the formal inverses, the resulting group no longer describes motion on a binary tree. That is, the monoid acts on the binary tree; but it is impossible to extend this action to an action of the full group.

Another critical aspect make note of is that the action of the monoid is an action on a certain set of intervals of the real number line. When the infinite binary tree is labelled either with the dyadic fractions or with the Farey fractions, it may be shown that any fraction occurring on the left is always strictly less then another fraction on the right. Thus, the binary tree extending under any given node has distinct and unique (Cauchy) limits to the left and to the right, thus defining an interval of the real number line. The subtree under any given node is clearly isomorphic to the tree as a whole. Finally, one may note that the dyadic rationals are dense in the reals, as are the Farey fractions. In fact, every rational shows up somewhere on the Farey tree. These properties make it clear that the act of navigating to any particular sub-node of the binary tree is the same as a map from one interval (say the unit interval) to a certain specific subinterval. It is not possible to navigate to just any interval whatsoever; these are constrained. A detailed discussion of which subintervals are allowed, and how to enumerate them, is given elsewhere (cf. my Minkowski question mark paper). Roughly, though, it may be said that the allowed intervals may be enumerated as $\mathbb{Q} \times \mathbb{N}$. It will become apparent, in the text of this paper, that the "allowed" intervals are precisely those for which a period-doubling fractal is self-similar.

2.1 Numerical representations

The set of strings in two letters is clearly isomorphic to the binary numbers, in that any given string in L and R can be converted to a string on 0 and 1, which may then be taken to be a binary number. Fractions such as 1/3, as well as irrational numbers correspond to strings of infinite length. A convenient alternate notation is in terms of continued fractions. One searches for a matrix r with the properties $r^2 = 1$ and $R = rLr$. The first property suggests that r is a 'reflection'. It follows from these properties that $R^n = rL^n r$ and so any string in R and L can be written as a string in r and L. In what follows, the

letter g is used for L, so that the strings are in g and r. A general string of the form

$$\gamma = L^{a_1} R^{a_2} L^{a_3} \dots$$

may therefore also be written as

$$\gamma = g^{a_1} r g^{a_2} r g^{a_3} r ...$$

While the first form has a natural representation in terms of a binary number, the second has a natural representation in terms of the continued fraction

$$[0; a_1, a_2, a_3, \dots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}$$

The numerical value of these two forms are not the same; the explicit map between these two numerical representations is given by the Minkowski question mark function. For the dyadic representation L_D and R_D of equation 3, one has

$$g_D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad r_D = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \tag{4}$$

while those corresponding to the Farey L_C and $R_C$are

$$g_C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad r_C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5}$$

Interpreted as operators on the binary tree, g corresponds to taking the left branch, while r corresponds to a mirror-reflection of the tree. Expressed in terms of intervals, g maps an interval to the left-half of the interval, while r is a reflection of the interval about its center point. If x is the label on a node of the dyadic tree (that is, if x is a real number), then one has the explicit representation

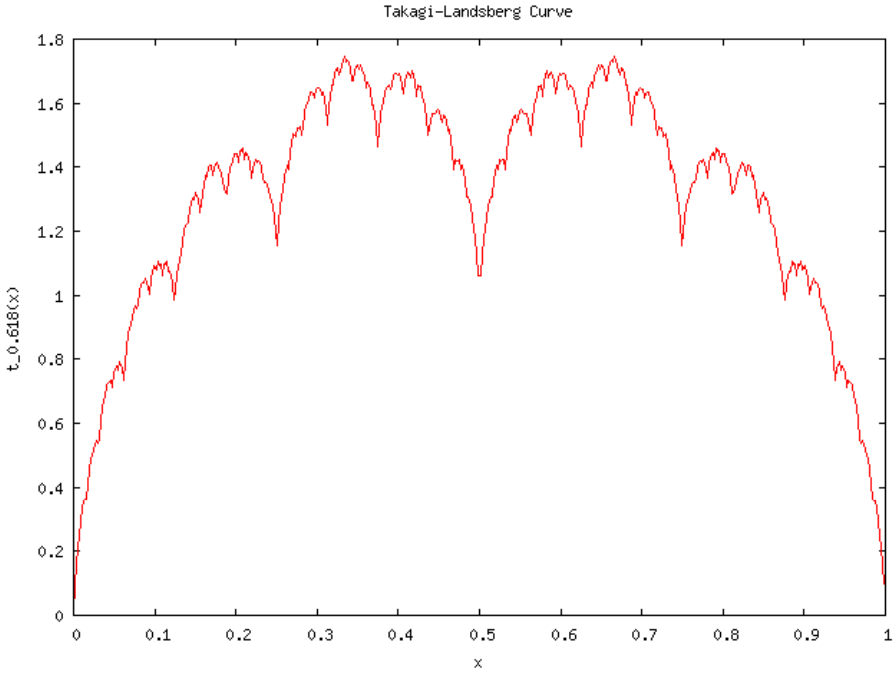
$$g_D(x) = \frac{x}{2} \quad \text{and} \quad r_D(x) = 1 - x$$

for $x \in [0,1]$. The above holds because the distribution of the dyadics is uniform with regards to the canonical topological measure on the real number line, and thus, any subtree will map back to the whole tree linearly.

3 The Takagi-Landsberg Curve

The Takagi-Landsberg curve was explored by Teiji Takagi, circa 1903 as an example of a continuous-everywhere, differentiable-nowhere curve [7, 6]. It is constructed by positive mid-point displacements of straight line segments, that is, by 'fractally' bumping a line segment upwards by an increasingly smaller displacement (xxxx need a diagram of the standard Koch-Curve-like mid-point displacement construction here). Equivalently, one assigns geometrically smaller displacements to each level of the dyadic tree;

Figure 3: The Takagi-Landsberg Curve



This graph shows the Takagi-Landsberg curve $t_w(x)$ as a function of x for a fixed value of $w = 0.618$. The fractal self-similarity of this curve is readily apparent. Other real values of w have a qualitatively similar shape; however, this is not the case when w is complex-valued, as is shown in figure 8.

that is, one picks some $0 \leq w < 1$, and assigns to each dyadic number $x = p/2^n$, for $n > 1$, the value

$$t_w\left(\frac{p}{2^n}\right) = w^{n-1} + \frac{1}{2}\left[t_w\left(\frac{p-1}{2^n}\right) + t_w\left(\frac{p+1}{2^n}\right)\right]$$

where we anchor the recurrence with $t_w(0) = t_w(1) = 0$ and $t_w(1/2) = 1$. One extends this curve to all reals by 'linear interpolation' between the dyadic values, as

$$t_w(x) = \sum_{k=0}^{\infty} w^k \tau\left(2^k x\right)$$

where $\tau(x)$ is the triangle function:

$$\tau(x) = \begin{cases} 2(x - \lfloor x \rfloor) & \text{for} \quad 0 \leq x - \lfloor x \rfloor < 1/2 \\ 2 - 2(x - \lfloor x \rfloor) & \text{for} \quad 1 > x - \lfloor x \rfloor \geq 1/2 \end{cases}$$

The curve is sometimes presented with 2^{-kH} taking the place of w^k in the series summation, where $H = -\log_2 w$ is called the Hurst parameter (need ref). The value $w = 1/4$ is special, in that $t_{1/4}(x) = 4x(1-x)$ is a parabola. This can perhaps best be seen with the mid-point displacement construction of a parabola, which is attributed to Archimedes[6]. A prototypical Takagi-Landsberg curve is shown in figure 3

In a certain technical sense, the curve is not differentiable at any dyadic value $x = p/2^n$, as such values correspond to the tip of the triangle wave, where the derivative is ill-defined. However, one can still perform derivative-like manipulations with the curve, which will be explored in greater detail below. For example, one may choose to define the derivative dt_w/dx by passing the d/dx past the summation, to act on the triangle-wave, replacing it with a square wave. The result is a curve that is discontinuous everywhere, but does have a number of derivative-like properties.

One can generalize this curve in a variety of ways, for example, by replacing the tent $\tau(x)$ with any $f(x)$ on the unit interval. To get a continuous Takagi curve, one must have f continuous and have $f(0) = f(1)$. Discontinuous functions f yield curves that are discontinuous everywhere: for example, the square wave $\sigma(x) = d\tau/dx = 1 - 2\lfloor 2x \rfloor$ gives the self-similar shape shown in the figures4 and 5 below.

Of the many possible different choices for $f(x)$, the triangle wave is special in that it stays self-similar through iteration; that is, $(\tau \diamond \tau)(x) \equiv \tau^2(x) = \tau(2x)$ and generally, iterating on it k times gives $\tau^k(x) = \tau(2^{k-1}x)$. Thus, one may write

$$t_w(x) = \sum_{k=0}^{\infty} w^k \tau^{k+1}(x) \tag{6}$$

Open problem: Develop the curious sum

$$e_w(x) = \sum_{k=0}^{\infty} \frac{w^k \tau^k(x)}{k!} \tag{7}$$

and describe its nascent self-similarity.

Open problem: Describe the fractal structure of the Weierstrass curve [8]:

$$\sum_{k=0}^{\infty} b^k \cos(a^k x\pi) \tag{8}$$

G.H. Hardy (need ref) showed that this function is differentiable nowhere provided that $0 < b < 1$, $a > 1$ and $ab \geq 1$. Despite a not having the value of two, this function still appears to have basic fractal self-similarity, although it is not even about $x = 1/2$ whenever a is not an integer.

3.1 Symmetries of the Takagi Curve

This section demonstrates that the Takagi curve transforms as a three-dimensional representation of the tree-walking monoid. This provides an example where the action of the monoid can be precisely and completely specified.

The interval contraction generator g acts on the Takagi map as

$$[gt_w](x) = (t_w \circ g_D)(x) = t_w\left(\frac{x}{2}\right) = x + wt_w(x)$$

Iterating on this generator gives

$$[g^n t_w](x) = t_w\left(\frac{x}{2^n}\right) = \frac{x}{2^{n-1}}\left(\frac{1-2^n w^n}{1-2w}\right) + w^n t_w(x) \tag{9}$$

The action of r is considerably more trivial:

$$[rt_w](x) = (t_w \circ r_D)(x) = t_w(1-x) = t_w(x)$$

since the Takagi curve is left-right symmetric. Thus, it is easily seen that the actions of g and r on a general expression $a + bx + ct_w(x)$ always returns another expression of the same type; and that the actions are both linear and also closed. That is, a general expression $a + bx + ct_w(x)$ can be interpreted as a vector in a three-dimensional vector space, and the action g and r have representation as operators on that space. We have that the Takagi Curve transforms under a three-dimensional representation of the monoid.

This may be made explicit as follows. Write a general element of the vector space as the dot-product of a row of real numbers times a column of basis vectors

$$a + bx + ct_w(x) = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 1 \\ x \\ t_w(x) \end{pmatrix}$$

Then g and r have the representations

$$g_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 1 & w \end{pmatrix} \text{ and } r_3 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where the subscript 3 is used to denote that this is a 3-D representation. Under the action of this representation, 3-vectors transform as, for example,

$$[gt_w](x) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 1 & w \end{pmatrix} \begin{pmatrix} 1 \\ x \\ t_w(x) \end{pmatrix}$$

This should be compared to the transformation properties of dyadic numbers. By a "dyadic number", we mean a real number represented as a string of binary digits

$$x = \sum_{n=0}^{\infty} b_n 2^{-n}$$

with the binary digits b_n being either 0 or 1. By replacing the number 2 with a different number, it is easily seen that the dyadic representation forms a Cantor Set (discussed in greater detail below), and that it's self-similarity properties are also given by the same semigroup/monoid that describes the Takagi curve. However, the representation of the transformation properties is a two-dimensional representation. That is, the dyadic representation that acted on dyadic numbers arranged in a doublet $(1,x)$ as

$$g_D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \text{ and } r_D = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$$

Alternately, one may write $g_D(x) = x/2$ and $r_D(x) = 1-x$ when acting on real numbers. The 3D representation is then seen to be a an extension of the 2D vector space. The Takagi Curve commutes with the generators as $[g_3 t_w](x) = (t_w \circ g_D)(x)$ and $[r_3 t_w](x) = (t_w \circ r_D)(x)$ where the left hand side is taken as an operator equation, and the right hand side as function composition.

This can be further belabored by examining the formulas for a general contracting group element $\gamma = g^{a_1} r g^{a_2} r g^{a_3} r ... r g^{a_N} \in PSL(2,\mathbb{Z})$. By 'contracting' we mean that we want to take all $a_k > 0$. The contracting elements form the monoid, and we explicitly want to avoid considering the inverses of these group elements. Trying to incorporate the inverses significantly complicates the analysis, and a full inclusion of all inverses cannot be satisfactorily done.

We have that γ acts on the real number x with the dyadic representation: that is,

$$\gamma : x = \gamma_D(x) = \frac{1}{2^{a_1}} - \frac{1}{2^{a_1+a_2}} + ... + (-1)^N \frac{1}{2^{a_1+a_2+\cdots+a_{N-1}}} + (-1)^{N+1} \frac{x}{2^{a_1+a_2+a_3+\cdots+a_N}}$$

which can be quickly computed by multiplying together the representation matrices g_D and r_D. By contrast, on the Takagi curve, γ acts as γ_3, thus:

$$[\gamma_3 t_w](x) = t_w(\gamma_D(x)) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} g_3^{a_1} r_3 g_3^{a_2} r_3 g_3^{a_3} r_3 ... r_3 g_3^{a_N} \begin{pmatrix} 1 \\ x \\ t_w(x) \end{pmatrix}$$

which demonstrates the self-symmetry of the curve for arbitrary (contracting) element $\gamma \in PSL(2,\mathbb{Z})$. That is, we have demonstrated an isomorphism between the set of

contracting elements $\gamma \in PSL(2,\mathbb{Z})$ and a subset of $GL(3,\mathbb{R})$ generated by g_3 and r_3. Note that this isomorphism is only between subsets, and not subgroups: the monoid does not contain its inverses.

Equivalently, this set of contracting elements is isomorphic to the set $\mathbb{Q} \times \mathbb{N} \times \mathbb{Z}_2$: pick any $p/q \in \mathbb{Q}$ and set $p/q = \gamma_D(0)$ which fixes the values of $a_1, ..., a_{N-1}$. Then pick a positive integer $a_N \in \mathbb{N}$ which fixes $\gamma_D(1)$. This fixes the two endpoints of the interval that contains the self-similar portion of the Takagi curve. Then pick from $\mathbb{Z}_2$ to decide if the mapping should invert the interval or not. Thus it is seen just how fine-grained, and also how constrained, the notion of self-similarity is. One may pick an interval one of whose endpoints may start on *any* rational, but once one endpoint has been specified, one only has a limited (albeit countably infinite) choice for the other endpoint.

Caution: The above has only demonstrated an isomorphism for the contracting (and their inverses, the expanding) group elements. General group elements also map intervals to intervals, but these may or may not be intersecting. When a pair of intervals do intersect, then there is a $\gamma_D(x)$ that is valid and defined on the interval, and thus the Takagi curve, restricted to that interval, demonstrates the desired isomorphism. When a pair of intervals do not intersect, then we do not have this demonstration, and so the above arguments fail. A rigorous treatment for general group elements seems to require a discussion of the topology of the modular group, of how to extend the mappings to the remaining group elements from those that can be easily handled. It is not clear if a consistent extension is possible, nor is it clear how to describe the obstruction if its not. This is beyond the level of the current presentation.

Homework: Compute the matrix elements for grg^3 for both the dyadic and the three-dimensional representation. Show that this implies the symmetry $t_w(1/2-x/16) = 1+x(-1/8+w/4+w^2/2+w^3)+w^4 t_w(x)$. Repeat the process for g^3rg and for $g^2rg^7rg^3$.

Homework: Provide a general, explicit proof of the commuting diagram by pulling out one generator at a time out from under the expressions.

What has been done here? Two things have been accomplished. First, the explicit symmetry relationship of the Takagi Curve has been demonstrated. Next, its been shown that the curve, together with 1 and x, belong to the set of basis vectors $\{1, x, t_w\}$ of a three-dimensional representation of the monoid. The matrix representations of g and r make it very easy to compute how the curve transforms under any given element of the monoid.

Corollary: The modular group monoid is a symmetry group of the parabola, where the parabola is given by $t_{1/4}(x) = 4x(1-x)$ and we can trivially validate that $(t_{1/4} \circ g)(x) = t_{1/4}(x/2)$ is equal to $(g \circ t_{1/4})(x) = x + t_{1/4}(x)/4$. The point here is that the monoid does not of necessity act only on fractal structures, but it can also act on differentiable structures as well. This action is seems considerably less exiting; however, it can be visualized as a mechanism for placing a "fractal coordinate system" onto a smooth manifold.

The next theorem shows that these representations, while isomorphic, are not equivalent in the traditional sense.

Theorem: Every real (or complex!) value of w generates a unique, inequivalent representation of the period-doubling monoid.

Proof: If two representations were equivalent, then one could find a matrix U such that $g_v = U^{-1} g_w U$ for $v \neq w$ and $r = U^{-1} r U$. Try to do this by brute force. Let

$$U = \begin{pmatrix} a & b & c \\ d & e & f \\ k & l & m \end{pmatrix}$$

Equating $rU = Ur$ we find immediately that $l = b = 0$. Next, equating $U g_v = g_w U$ one finds that $d = k = c = f = 0$, leaving $a = e = m$. But when $v \neq w$ we find also that $m = 0$, and so all elements of U vanish. There is no equivalency transformation between two different representations. *QED*.

The above shows that we can use the parameter w to label the different 3D representations. Note, however, that the parameter w also has a (one-to-one) correspondence to the fractal dimension of the Takagi curve (at least for $w > 1/4$). Thus, in a certain sense, we can say that the different representations correspond to different fractal dimensions (although the correspondence is not one-to-one for $w < 1/4$).

Homework: Give the fractal dimension of the Takagi Curve, as a function of w.

Open-Question: We have demonstrated a one-dimensional manifold of representations of the modular group. Below, we will show a representation on the complex plane. Are there other 3D representations, or is this it? The answer, given further below, is "yes, there are other 3D reps". These include the left-right symmetric Levy curve, and the left-right symmetric Koch curve, discussed in a later section. While these also transform as 3D reps, they generate plane curves that, unlike the Takagi curve, cannot be written as a real-valued function of the real number line.

3.2 Multi-fractals

The uniqueness of the Takagi curve for each distinct value of w implies that we can use it to define an (incomplete) set of basis vectors for functions on the unit interval. That is, there is a large class of functions on the unit interval $f(x)$ that can be expressed as

$$f(x) = \int_0^\infty c(w)\, t_w(x)\, dw$$

for some function $c : \mathbb{R}^+ \to \mathbb{R}$ or more generally $c : \mathbb{R}^+ \to \mathbb{C}$. Thus, for example, the function of equation 7 be expressed in this form. The resulting function $f(x)$ does not have any particular self-similarity properties, but its decomposition does.

4 The Cantor Polynomials

The action of a group element γ on the Takagi curve can be written as

$$t_w(\gamma_D(x)) = p_\gamma(w) + x\, q_\gamma(w) + w^{M_\gamma} t_w(x)$$

for some polynomials p_γ, q_γ and positive integer M_γ where we use the subscript γ to denote that they depend on the particular γ chosen. A brief bit of scratching shows what the q_γ polynomial is; but p_γ is harder. For $\gamma = g^{a_1} r g^{a_2} r g^{a_3} r ... r g^{a_N}$ as before, one has $M_\gamma = a_1 + ... + a_N$. Writing out the binary expansion for $\gamma_D(1)$ in terms of binary digits $b_k \in \{0,1\}$, one has

$$\begin{aligned} \gamma_D(1) &= \sum_{k=1}^{\infty} \frac{b_k}{2^k} \\ &= 0.\ \underbrace{000...000}_{a_1}\ \underbrace{111...111}_{a_2}\ \underbrace{000...000}_{a_3}\ ... \end{aligned} \tag{10}$$

where, as usual, one takes the expansion such that the final binary digits repeat forever, without upsetting the count a_N of the previous digits; that is,

$$b_k = \begin{cases} 1 & \text{for N odd,} \quad \forall k > M \\ 0 & \text{for N even,} \quad \forall k > M \end{cases}$$

The polynomial q_γ can then be expressed directly in terms of this binary expansion.

$$q_\gamma(w) = \frac{(-)^N}{2^{M-1}} \sum_{k=1}^{M} (2b_k - 1)\ (2w)^{k-1}$$

Note that q_γ is more-or-less equal (up to a factor and a term of $1/(1-2w)$) to the derivative

$$\left. \frac{dt_w(x)}{dx} \right|_{x=\gamma_D(1)}$$

We say "more or less" because some care has to be taken to define which bit expansion defines the derivative. In a certain sense, q_γ is the dis-ambiguated derivative. This is left as an exercise.

The p_γ polynomial seems to result from some more complex folding that is hard to discern by studying the matrix elements under matrix multiplication. However, determining its value for any given γ is straightforward enough: we merely need to set $x = 0$ to find $p_\gamma(w) = t_w(\gamma_D(0))$.

Homework: Describe $p_\gamma(w)$. Is it possible to give a direct expression for the polynomial?

The polynomials being dealt with here are closely related to the inverse of the Cantor function. This is perhaps best presented visually. Write the binary expansion of a real number x as

$$x = \sum_{k=1}^{\infty} \frac{b_k}{2^k} \tag{11}$$

and then define its Cantorization as

$$c_z(x) = (1-z)\sum_{k=1}^{\infty} b_k z^k$$

Clearly, for $z = 1/2$, we have $c_{1/2}(x) = x$. For $z = 1/3$, we have essentially the inverse of the classic Cantor function, as show in the figure 4.

For larger values of z, one can more clearly see the visual resemblance of the Cantor polynomial to the derivative of the Takagi function. The resemblance is not accidental; they really are the same thing, as long as one takes care to properly define which binary expansion one is working with. The details are not given here, although they are not hard to work out (see homework problem below). The essential point is that every rational number has two possible binary expansions: for example, it is well known that $1.000.. = 0.111...$. But each such expansion yields two inequivalent Cantor polynomials. Cantor's essential insight was that these two inequivalent representations of the real numbers by binary sequences can be re-interpreted as the endpoints of a set of non-overlapping line segments. Converting the binary sequence into a polynomial series in z make these two endpoints explicit by assigning a different polynomial to each.

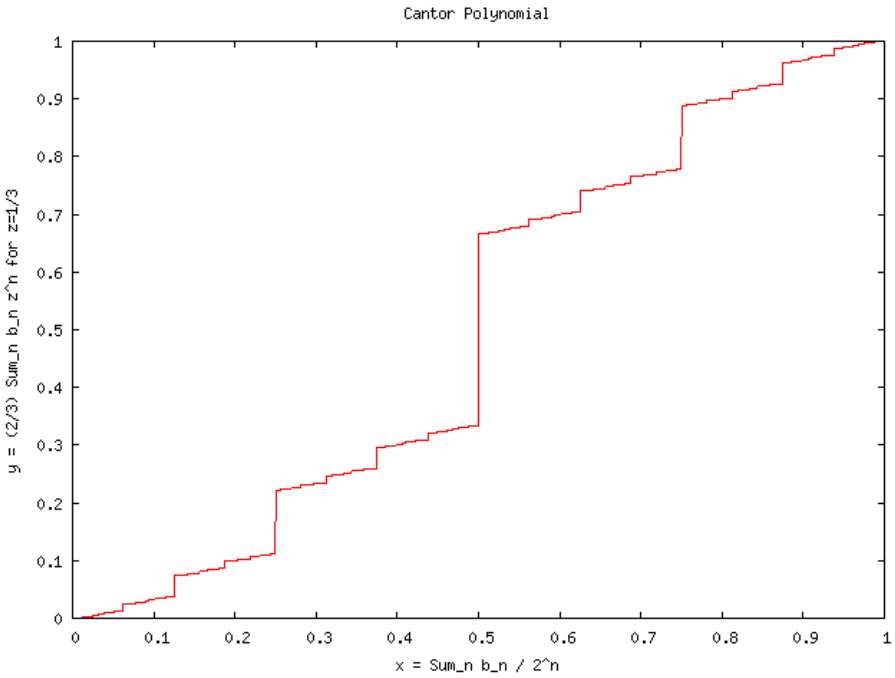
Another language one can use to describe this is to say that there is a "gap" at every dyadic fraction i.e. at every rational of the form $p/2^n$ for p odd and integer $n > 0$. This gap is inherent in the representation of the reals by binary sequences. The gap orders all of the reals above and below it. Note that there is another construction by which one can place gaps at every rational number. This construction is arrived at through the isomorphism between dyadic fractions and Farey Numbers by means of the Minkowski Question Mark function. Essentially, we can represent every rational as a continued fraction, and define gaps analogously. This idea is reviewed in the companion paper.

The main insight provided here is that the two endpoints of Cantor's interval construction can also be mapped to the left and right derivatives of the Takagi function. That is, for any dyadic rational $p/2^n$ for odd integer p, there are two inequivalent binary sequences that represent this rational. Each sequence can be used to build a distinct polynomial. The value of two polynomials at the rational are the two endpoints of an interval in the Cantor Set construction. But these two endpoints are also the left and right derivatives of the Takagi Curve. One sequence maps to the derivative of the Takagi function on one side of this rational, and the other to the other side of this rational.

Homework: Make the above statements firm by providing all the correct factors needed to relate the Cantor polynomial $c_z(x)$ to $q_\gamma(w)$ and to the suitably firmed-up derivative $dt_w(x)/dx$. Start by noticing that the traditional derivative is ill-defined only at dyadic fractions; for all other rationals and irrationals, there is no problem. In particular, the dyadic fractions are a set of measure zero; the derivative is well-defined on a set of measure one. At the dyadic fractions, use the gap to assign left and right derivatives, and give an exact expression for each.

Homework: The classic example of a Cantor Set is built by means of a dyadic subdivision of the unit interval. Because of this, we can apply the monoid symmetry here as well. How do the limit points of the Cantor set (the points that are not the

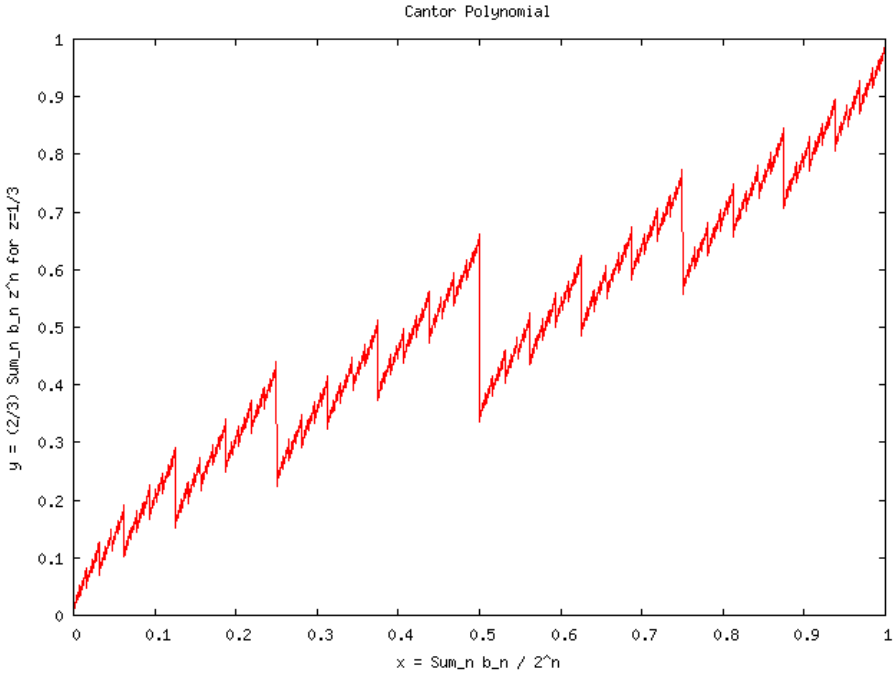
Figure 4: Cantor Polynomial



This figure shows a graph of the function $c_z(x)$ for a value of $z = 1/3$. This graph provides possibly the easiest visual proof that the cardinality of the Cantor Set is equal to the cardinality of the unit interval[2], a result which can sometimes be difficult to visualize. The function $c_{1/3}(x)$ maps the unit interval into the unit interval. It is strictly monotonically increasing, and if $x \neq y$ then $c_{1/3}(x) \neq c_{1/3}(y)$. Thus, this function is one-to-one, and thus the cardinality of the image is equal to the cardinality of the unit interval. But the image of this function is just the canonical Cantor set, constructed by recursively removing the open middle-third interval. The largest "middle-third" corresponds to $x = 1/2$; the next two largest correspond to $x = 1/4$ and $x = 3/4$, and so on: each removed interval corresponds to a rational number, expressed as its binary expansion. It is not hard to see that that the sum of the lengths of the "middle thirds" adds up to one, and thus, the measure of the image of this function is zero. Thus, we have quickly sketched that the cardinality of the Cantor set is that of the real numbers, but the Lebesgue measure of the Cantor Set is zero. It is not hard to see that one gets a similar result using $c_z(x)$ for any value of $0 < z < 1/2$.

There were only two "tricky" parts to this demonstration. One is the assertion that $c_z(x) \neq c_z(y)$ whenever $x \neq y$, but this can be deduced easily enough by contemplating the definition of $c_z(x)$. The other trick we slid by here was that $c_z(x)$ is somewhat vaguely defined for the rationals: every dyadic rational number $x = p/2^n$ has two inequivalent binary expansions. For example, x=1/2=0.1000...=0.0111... and the first binary expansion yields $c_z(1/2) = 1 - z$ while the second yields $c_z(1/2) = z$. But this doesn't affect the proof: however we choose to tighten up the definition of $c_z(x)$, it is still one-to-one and monotonic, and its image still has the cardinality of the unit interval.

Figure 5: Cantor Polynomial



This figure shows a graph of the function $c_z(x)$ for a value of $z = 2/3$.

endpoints of intervals) behave under the dyadic monoid? Given that the Cantor set can be mapped to the real numbers (the cardinality of the Cantor set is the cardinality of the continuum), can we work backwards to deduce the action of the modular group monoid on the real number line?

Presumably, it should be obvious by now that the Cantor Polynomials transform as the two-dimensional (dyadic) representation of the period-doubling monoid. This should be expected, based on the reasoning about the dimension of the representation presented in the next section. XXX These last few paragraphs are a less-than satisfying treatment.

4.0.1 Cantor Measure

The juxtaposition of the two figures 4 and 5 suggests a mathod for defining a measure on Cantor sets. One starts with the ordinary Lebesgue measure on the real number line, and passes it through the Cantor polynomial, defined above, and ask how that measure accumulates on the y-axis. For $z < 1/2$, this mapping defines a distribution that is discontinuous: the measure accumulates on the points in the Cantor set, and is vanishing elsewhere. Essentially, the distribution is that of the membership function of a set. The membership function is a function that is equal to one for points in a set, and zero for points not in the set. For $z > 1/2$, however, the sawtooth graph makes clear that intervals on the x-axis will be projected to a multitude of overlapping intervals on the y-axis. One may ask what the resulting distribution is on the y-axis. The following figures illustrate the answer.

5 The Complex-valued Takagi Curve

Some rather remarkable curves result when one considers the Takagi curve for complex values of the parameter w. The symmetry considerations given above apply equally well to real or complex values of w. What is remarkable is how the self similarity can in fact be so carefully hidden. The phase plot shown in figure 8 visually resembles nothing so much as a Brownian random walk. Some of the mystery of this plot is revealed by examining the real and imaginary parts independently, as shown in figure **??**.

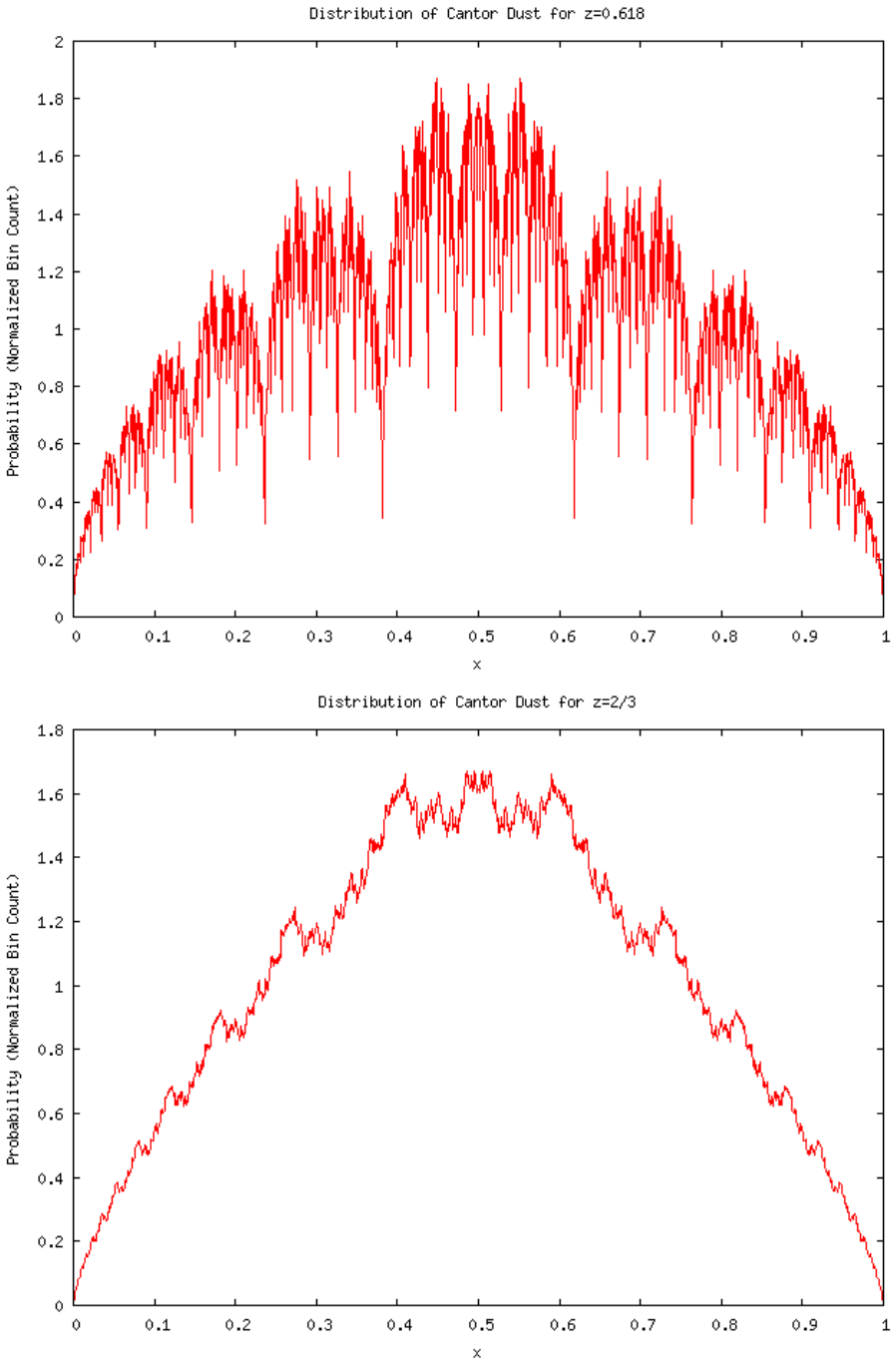
6 The Four-Dimensional Representation

Replacing the sawtooth shape that defines the the Takagi function with a polynomial leads to a curve that transforms under a higher-dimensional representation, specifically, as the degree of the polynomial plus two. This section works out the four-dimensional representation and then presents the general finite-dimensional representation.

Lets replace the sawtooth by a quadratic shape

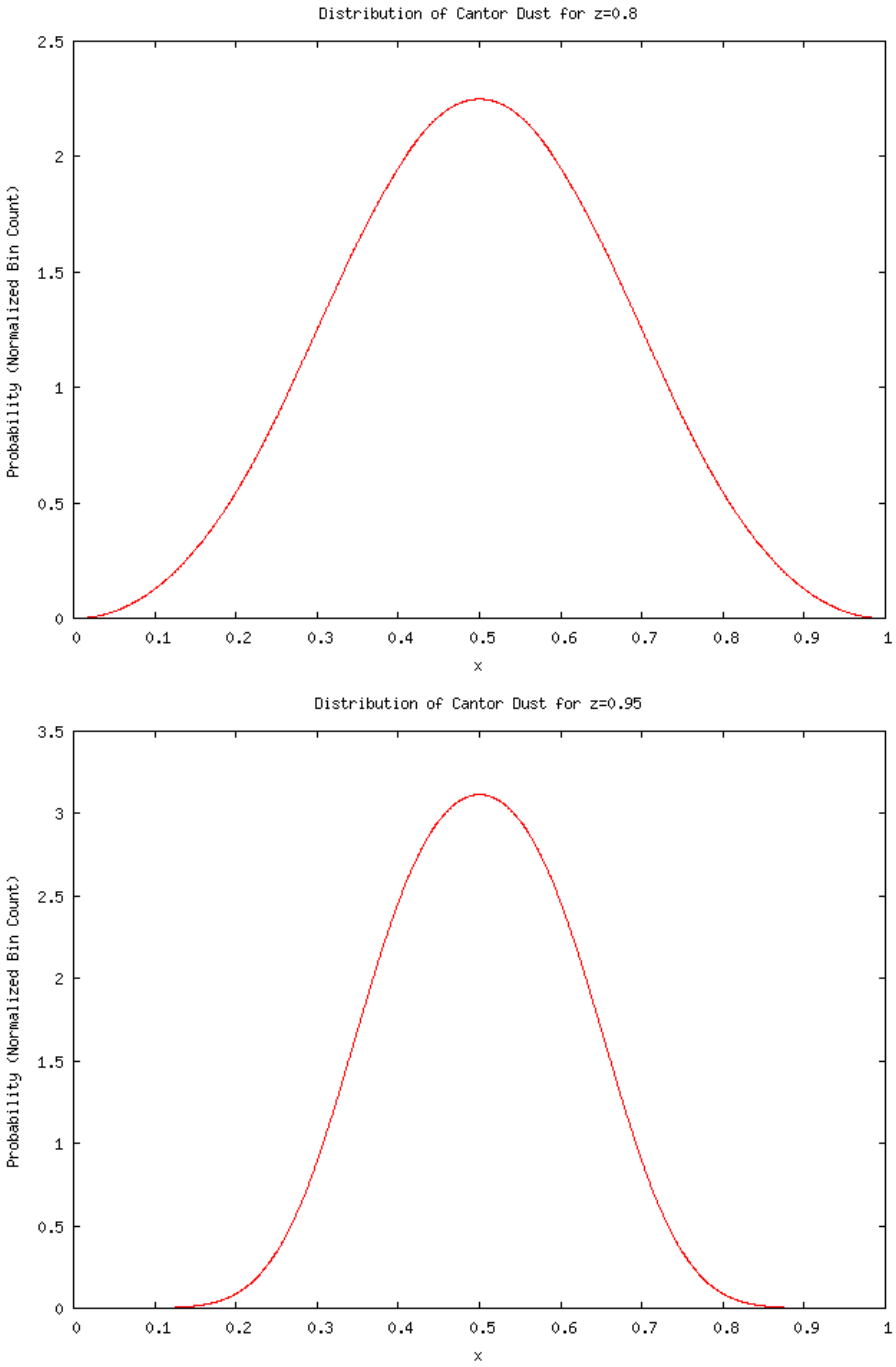
$$\alpha(x) = \begin{cases} 4x^2 & \text{when} \quad 0 \le x \le 1/2 \\ 4(1-x)^2 & \text{when} \quad 1/2 \le x \le 1 \end{cases}$$

Figure 6: Cantor Distributions



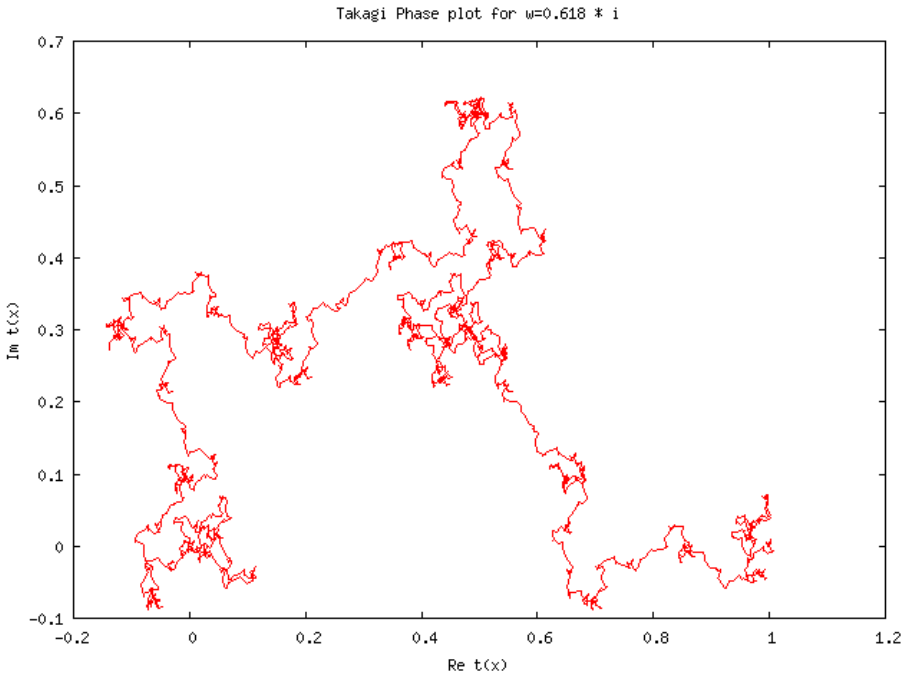
These graphs shows the distribution that results from taking the uniform distribution on the real number line (the Lebesgue measure on the x-axis) and passing it through the polynomial $y = c_z(x)$ and projecting it on the y-axis. The upper graph is for the case of $z = 0.618$, and the lower for $z = 2/3$.

Figure 7: Cantor Distributions

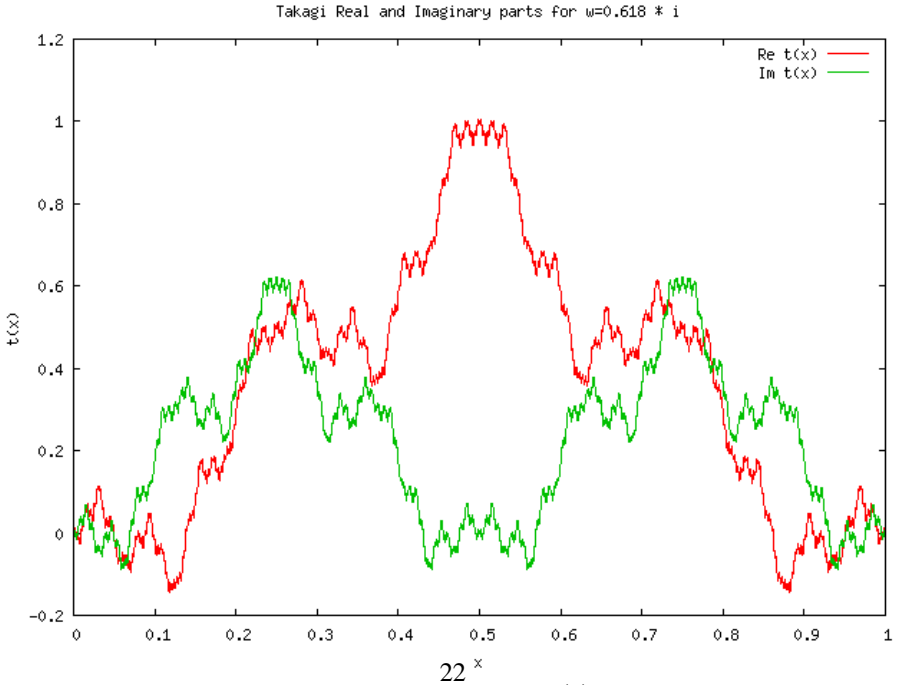


These graphs shows the distribution that results from taking the uniform distribution on the real number line (the Lebesgue measure on the x-axis) and passing it through the polynomial $y = c_z(x)$ and projecting it on the y-axis. The upper graph is for the case of $z = 0.8$, and the lower for $z = 0.95$. Both graphs appear smooth, but are presumably fractal at microscopic scales.

Figure 8: Takagi Phase Plot



This figure shows a phase plot of $t_w(x)$ as a function of x for a fixed value of $w = 0.618i$. This curve has exactly the same fractal self-similarity as the curve shown in figure 3, but here, the self-similarity of this curve is far from apparent.



This figure shows the real and imaginary parts of $t_w(x)$ as a function of x for a fixed value of $w = 0.618i$. The general symmetry is more readily appearent here.

so that

$$a_w(x) = \sum_{k=0}^{\infty} w^k \alpha\left(2^k x - \left\lfloor 2^k x \right\rfloor\right)$$

Its not hard to find that

$$[g_4 a_w](x) \equiv (a_w \circ g_D)(x) = a_w\left(\frac{x}{2}\right) = x^2 + w a_w(x)$$

Combining this with the reflection $r(x) = 1 - x$, it is seen again that the transformations are linear and closed, and thus again a vector space, this time four-dimensional. The generator $g_4$can be represented as the matrix

$$g_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & w \end{pmatrix}$$

and the reflection $r(x) = 1 - x$ as

$$r_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with the constructions of the isomorphisms proceeding as before.

In general, if $\alpha(x)$ is replaced with another quadratic construction that is even about $x = 1/2$, then the corresponding generator g will be conjugate to g_4. For example, if we were to construct the curve from

$$\beta(x) = 4x(1-x)$$

so that

$$b_w(x) = \sum_{k=0}^{\infty} w^k \beta\left(2^k x - \left\lfloor 2^k x \right\rfloor\right)$$

then we'd find that

$$g_b = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 2 & -1 & w \end{pmatrix}$$

is conjugate to g_4 through $U g_4 = g_b U$ with U given by

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \frac{4}{1-2w} & \frac{-4}{1-2w} & \frac{-2w}{1-2w} \end{pmatrix}$$

Note that this conjugacy breaks down when $w = 1/2$. This is because the curve generated by $\alpha(x)$ becomes a parabola at this point; that is, $a_{1/2}(x) = 4x(1-x)$, which transforms under the three-dimensional, not the four-dimensional representation.

We can also construct a function that is odd about $x = 1/2$:

$$\beta_o(x) = \begin{cases} 16x(1/2-x) & \text{when} \quad 0 \le x \le 1/2 \\ 16(1-x)(1/2-x) & \text{when} \quad 1/2 \le x \le 1 \end{cases}$$

which is a pair of oppositely oriented parabolas. We use the subscript "o" to denote odd parity. A graph looks like a rude, crude version of the sine function. We then have $\beta_o(1-x) = -\beta_o(x)$ and so $b_{o,w}(1-x) = -b_{o,w}(x)$. The representation for r for this function has an extra minus sign:

$$r_{4o} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and cannot be made conjugate to r_4; that is, there does not exist a U with $\det U \neq 0$ such that $U r_{4o} U^{-1} = r_4$. The rep for g is

$$g_{4o} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 4 & -4 & w \end{pmatrix}$$

because $b_{o,w}(g(x)) = 4x(1-x) + w b_{o,w}(x)$. So in fact, for a fixed w, we have two 4D representations, of even and odd parity, that aren't conjugate. Working backwards, we realize that there is actually another 3D representation, of odd parity, that is not conjugate to the even-parity representation. It is the one that is built out of the shifted sawtooth

$$\sigma(x) = \begin{cases} 4x & \text{for} \quad 0 \le x \le 1/4 \\ 2-4x & \text{for} \quad 1/4 \le x \le 3/4 \\ 4x-4 & \text{for} \quad 3/4 \le x \le 1 \end{cases}$$

and give an odd-parity representation because $\sigma(1-x) = -\sigma(x)$. Under g, we have $\sigma(g(x)) = \tau(x) + w\sigma(x)$. However, if we try to work out the matrix representation g_{3o} for this function, we find that the bottom row of the matrix has elements that aren't constants, but are rather piece-wise assemblies that depend on whether x is greater or less than 1/2. Strictly speaking, then, this isn't a linear representation. We can try to save the day by building a different odd-parity sawtooth; but then we find that such a sawtooth will be discontinuous at $x = 0$ and/or at $x = 1/2$. In other words, there doesn't seem to be a continuous odd-parity representation in $GL(3,\mathbb{R})$. There is, however, an odd-parity representation in $GL(3,\mathbb{R})$ that is not continuous: it is the one built from $\sigma(x) = x - 1/2$.

For the most part, we've tried to be careful to choose functions that yield a continuous Takagi curve. One can build discontinuous curves by working with the square-wave. The square-wave, in a certain sense, transforms as the 2D rep. Curiously, the odd-parity quadratic above has a continuous first derivative. We can in fact build an even-parity quadratic with a continuous first derivative, by following the same recipe:

$$\beta_e(x) = \begin{cases} 1-16x^2 & \text{for} \quad 0 \le x \le 1/4 \\ 16(x-1/2)^2 - 1 & \text{for} \quad 1/4 \le x \le 3/4 \\ 1-16(1-x)^2 & \text{for} \quad 3/4 \le x \le 1 \end{cases}$$

This is explicitly even: $\beta_e(1-x) = \beta_e(x)$. A graph of this looks like a rude, crude version of the cosine. If we try to work out the matrix representation g_{4e} for this function, we run into difficulty again: the bottom row of the matrix has elements that aren't constants, but are rather piece-wise assemblies that depend on whether x is greater or less than 1/2. It seems that there isn't any even-parity quadratic function with continuous first derivatives that has a matrix representation in $GL(4,\mathbb{R})$. This seems to mirror the odd-parity problem for the 3D rep.

7 Higher-Dimensional Representations

It should now be apparent how to build arbitrary higher-dimensional representations. One can, for instance, start with

$$\alpha_n(x) = \begin{cases} 2^n x^n & \text{when} \quad 0 \le x \le 1/2 \\ 2^n(1-x)^n & \text{when} \quad 1/2 \le x \le 1 \end{cases}$$

and discover that

$$[g_{n+2}a_{n,w}](x) \equiv (a_{n,w} \circ g_D)(x) = a_{n,w}\left(\frac{x}{2}\right) = x^n + w a_{n,w}(x)$$

The matrix representation of g has inverse powers of 2 along the diagonal, except for the last row, and the reflection matrix is lower-triangular, with binomial coefficients making up the triangle (excluding the last row). If one attempts different constructions by replacing the monomial $2^n x^n$ with a polynomial, but keeping α_n even about 1/2, then one finds that all these representations are conjugate to each other. Alternately, we can construct α_n so that it is odd about 1/2; in this case, we again find that all the representations are equivalent up to conjugacy. Thus, we conclude that for any given dimension, there are two distinct representations, which we shall call even and odd, according to the symmetry of the Takagi curve that forms the basis vector.

The above $\alpha_n(x)$ is explicitly even-parity, but its first derivative is discontinuous at $x = 1/2$. One can then try to build even and odd-parity functions of a given degree, with continuous $(n-1)$'th derivatives. The corresponding representations then succeed or founder as in the 3D/4D cases above, with even and odd order functions alternately causing the piecewise-trouble for even and odd representations. That is, one of the two representations will allow Takagi curves that are $n-1$ times differentiable, while the other can work only up to $n-2$. The obvious choice for the differentiable-curve construction are the Bernoulli polynomials. These have several appealing properties for this task: they alternate in parity as they alternate in order, and, in higher orders, approximate sine and cosine. That is, $B_{2n}(x) \to \pm A_{2n}\cos(2\pi x)$ and $B_{2n+1}(x) \to \pm A_{2n+1}\sin(2\pi x)$ as $n \to \infty$ for some constants A_n. That is, we assert that an appropriate and natural choice for the basis vectors for the higher-dimensional representations are given by Bernoulli polynomials. We thus define the basis vectors to be

$$T_{n+}(x;w) = \sum_{k=0}^{\infty} w^k B_n\left(2^k x - \left\lfloor 2^k x \right\rfloor\right)$$

for the one parity, and, for the opposite parity,

$$T_{n-}(x;w) = \sum_{k=0}^{\infty} w^k \check{B}_n \left(2^k x - \left\lfloor 2^k x \right\rfloor\right)$$

where we define

$$\check{B}_n(x) = \begin{cases} B_n(2x) - B_n(1) & \text{when} \quad 0 \le x \le 1/2 \\ (-)^n \left(B_n(1) - B_n(2(1-x))\right) & \text{when} \quad 1/2 \le x \le 1 \end{cases}$$

The next chapter will explore these expressions in greater detail, xxxxxxxx fix the above to match basis used in other chapters. xxxxx

Given that higher and higher degree polynomials require higher and higher-dimensional representations, this implies that a Takagi curve built out of a general function will in some way transform as an 'infinite dimensional' representation. We find it curious, though, that alternating even and odd-parity representations seem to allow sine and cosine to be built up in a natural fashion, thus opening the door for Fourier analysis

The next section takes a quick look at one example of an "infinite-dimensional" curve, the fractal eigenfunctions of the Bernoulli map transfer operator.

8 Bernoulli Map Symmetry

The eigenfunctions of the Bernoulli map [4] bear a close resemblance to the Takagi curve, and one might say they are generalization of the Takagi Curve to the complex plane. These are

$$\psi_{z,l}(x) = \sum_{n=0}^{\infty} z^n \exp\left(2\pi i \, 2^n \left(2l+1\right) x\right)$$

Insofar as the exponential can be considered to be a polynomial of infinite degree, or to be constructed as the limit of a series of polynomials, then one might expect these functions to transform under some 'infinite-dimensional' representation, taken as a limit of the above development for polynomials.

The two figures show two of these curves, for different values of z.

It can be readily worked out that the factor of $2l+1$, which is needed to enumerate the eigenfunctions, does not add any new or interesting complications to the overall symmetry. With a minimum of effort, one can quickly see that these curves transform in a fashion similar to the Takagi Curve. Starting with g, we have

$$\left[g\psi_{z,l}\right](x) = \left[\psi_{z,l} g\right](x) = \psi_{z,l}(\frac{x}{2}) = \exp\left((2l+1)\pi i x\right) + z\psi_{z,l}(x)$$

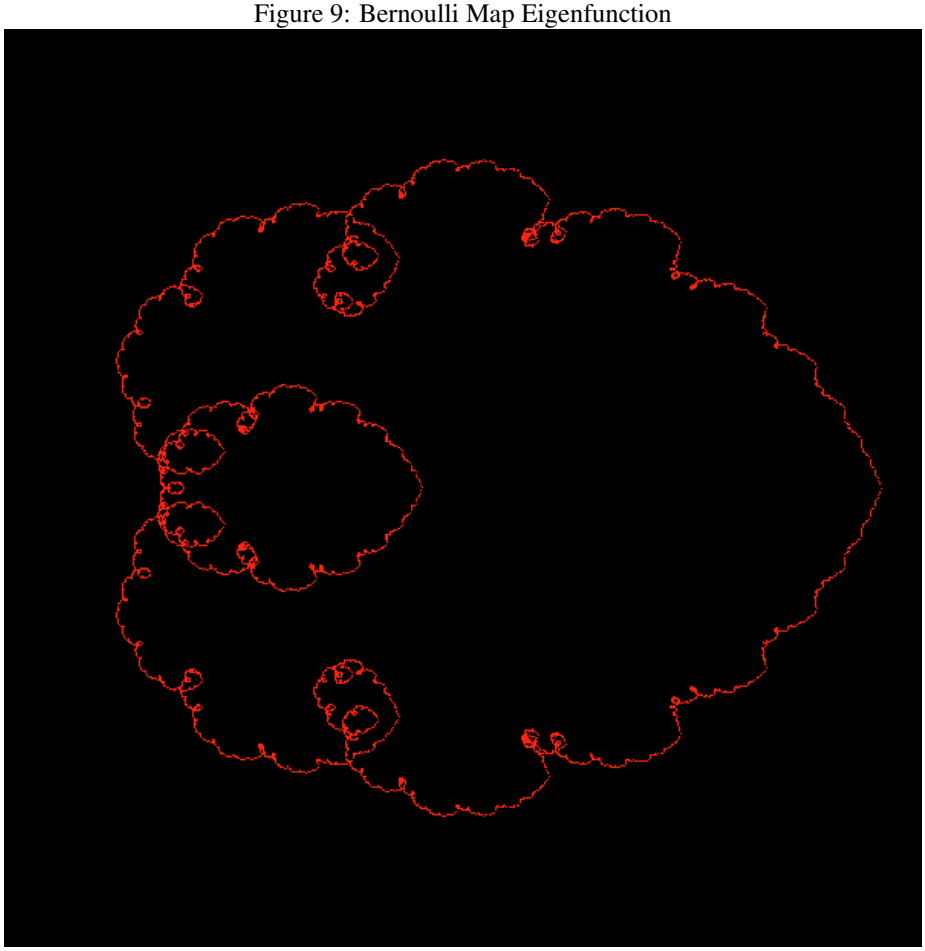
and for reflection,

$$\left[r\psi_{z,l}\right](x) = \left[\psi_{z,l} r\right](x) = \psi_{z,l}(1-x) = \psi_{z,l}(-x)$$

Concatenating group elements proceeds in a fashion similar to before. For example,

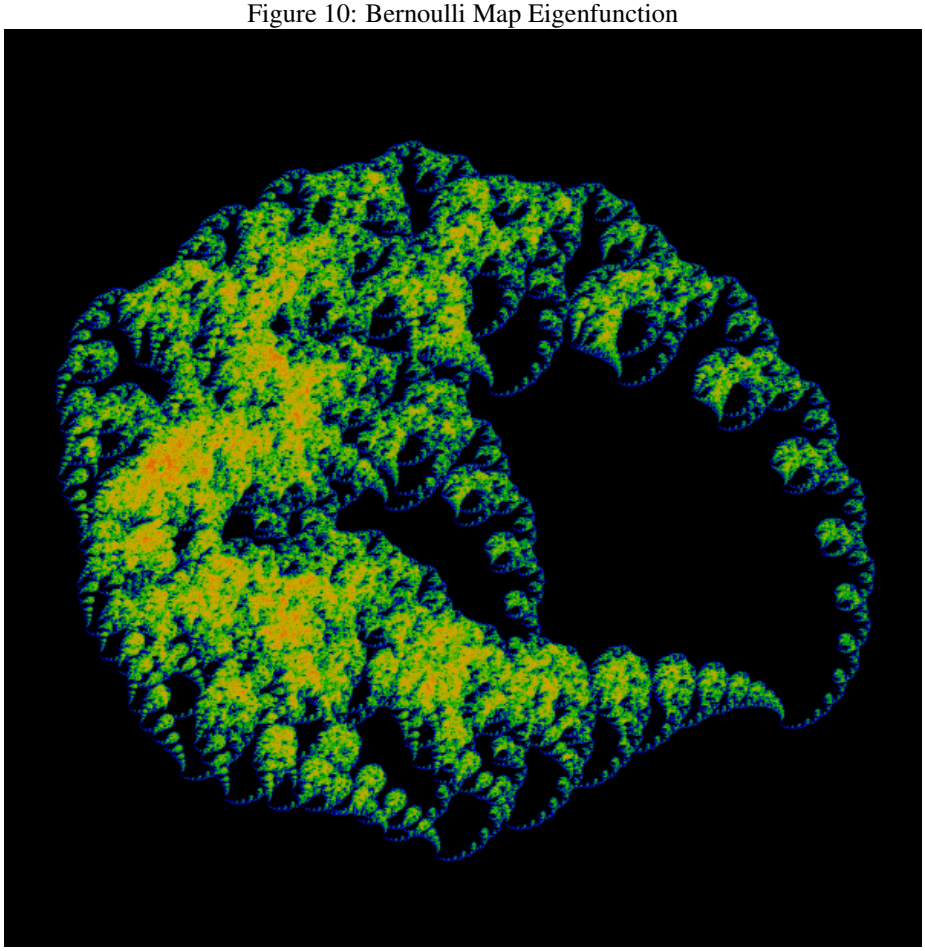
$$g_D^3 r_D g_D^2(x) = \frac{1}{8} - \frac{x}{32}$$

Figure 9: Bernoulli Map Eigenfunction



Above is shown a polar plot of the Bernoulli Map eigenfunction $\psi_{zl}(x)$ for $l = 0$ and $z = 0.55$. The x and y axis are respectively the real and imaginary parts of $\psi_{zl}(x)$. For smaller $|z|$, the loops shrink in size and quickly disappear. Adding imaginary components breaks the mirror symmetry of the curve. Larger values of $|z|$ makes the loops increasingly convoluted, as the next figure shows.

Figure 10: Bernoulli Map Eigenfunction



Above is shown a polar plot of the Bernoulli Map eigenfunction $\psi_{zl}(x)$ for $l = 0$ and $z = 0.7 + 0.16i$. The x and y axis are respectively the real and imaginary parts of $\psi_{zl}(x)$. Because the curlicues of the figure overlap enough to make a big mess, we've colored the image so that green and yellow indicate areas where the curve overlaps many, many times, and the blue areas where the curve is visited only a few times.

and so

$$\left[g^3rg^2\psi_z\right](x) = \psi_z\left(\frac{1}{8}-\frac{x}{32}\right) = \exp\left(i\pi\left(\frac{1}{4}-\frac{x}{16}\right)\right) + z\psi_z\left(\frac{1}{4}-\frac{x}{16}\right)$$

where we've taken, for convenience, $l = 0$. Peeling off one group element at a time, we finally get

$$\begin{aligned}\psi_z\left(\frac{1}{8}-\frac{x}{32}\right) &= \exp\left(i\pi\left(\frac{1}{4}-\frac{x}{16}\right)\right) + z\exp\left(i\pi\left(\frac{1}{2}-\frac{x}{8}\right)\right) + \\ &+ z^2\exp\left(i\pi\left(1-\frac{x}{4}\right)\right) + z^3\exp\left(-i\pi\frac{x}{2}\right) + \\ &+ z^4\exp\left(-i\pi x\right) + z^5\psi_z\left(-x\right) \qquad (12)\end{aligned}$$

However, this obscures the true form of the transformation, which is exhibited more clearly if we write the above as

$$\begin{aligned}g^3rg^2\psi_z &= \psi_z g^3rg^2 \\ &= Eg^3rg^2 + zEg^2rg^2 + z^2Egrg^2 + z^3Erg^2 + z^4Erg + z^5\psi_z r \qquad (13)\end{aligned}$$

where we defined $E(x) \equiv \exp(2\pi ix) = \psi_{z=0}(x)$. Just to be perfectly clear, in the above expression, we mean functional composition whenever two elements are next to each other, with the exception of z, which is just plain multiplication. The operator E has the effect of peeling off one group element at a time, and converting it to a z. The result is a funny polynomial which captures the action of the group element. The funny polynomial is the orbit of ψ_z under the action of a general group element. We abstain from trying to write the corresponding polynomial for the general element $\gamma = g^{a_1}rg^{a_2}rg^{a_3}r...rg^{a_N}$ mostly because it would be quite ungainly. We note only that the highest power of z will be $a_1 + a_2 + ... + a_N$ and also that r can take the place of a minus sign, in that $(Er)(x) = E(r(x)) = E(1-x) = E(-x)$, which remains in place, until the next r cancels it out. None-the-less, this polynomial is curious, and calls for further development.

9 Topologically Conjugate Maps

The Modular Group Symmetry describes not just this peculiar handful of curves, but in fact a large class of curves. Any monotonically increasing function $\phi(x)$ with $\phi(0) = 0$ and $\phi(1) = 1$ can be used to build topologically conjugate maps. For the tent map, we write $\sigma(x) = (\phi \circ \tau \circ \phi^{-1})(x)$ so that

$$s_w(x) = \sum_{k=0}^{\infty} w^k(\phi \circ \tau^{k+1} \circ \phi^{-1})(x)$$

which transforms under the action of

$$g_\phi(x) = (\phi \circ g_D \circ \phi^{-1})(x)$$

as

$$[gs_w](x) \equiv s_w(g_\phi(x)) = x + ws_w(x)$$

This transformation law follows rather trivially, because we had $\tau^{n+1} g_D = \tau^n$ for $n > 0$ and $(\tau g_D)(x) = (1)(x) = x$. In the same vein, we define

$$r_\phi(x) = (\phi \circ r_D \circ \phi^{-1})(x)$$

for $r_D(x) = 1 - x$. Working with these two generators, we can construct the generic element very easily, as $\gamma_\phi = \phi \circ \gamma_D \circ \phi^{-1}$, and discover, of course, that s_w transforms under exactly the same three-dimensional representation that t_w does.

Lets quickly examine a specific example, the Logistic map of unit height. This map is conjugate to the tent map, and the conjugating functions are given by

$$\phi(x) = \frac{1}{2} - \frac{1}{2}\cos(\pi x)$$

and

$$\phi^{-1}(x) = \frac{1}{\pi}\arccos(1 - 2x)$$

so that we get

$$\lambda(x) \equiv (\phi \circ \tau \circ \phi^{-1})(x) = 4x(x-1)$$

Then, iterating on this map, we get

$$L_w(x) = \sum_{k=0}^{\infty} w^k \lambda^{k+1}(x)$$

The generators of the symmetry group for this conjugating function are

$$g_\phi(x) = \frac{1}{2} - \frac{1}{2}\sqrt{1-x}$$

and $r_\phi(x) = r_D(x) = 1 - x$. The graph of this curve is not that interesting and generally resembles that of the Takagi Curve shown in 3, and thus we do not graph it here.

Although constructed from a parabola, do not confuse this curve with the curve that transforms under the 4D rep: this is because $\lambda^{k+1}(x) \neq \lambda(2^k x)$: in one case, we are iterating, in the other, we are making copies. In the above discussion, we've conflated two different concepts: the act of iteration, and the construction of the Takagi curve. We can do this because iterating the tent map just makes copies of the map: $\tau^{k+1}(x) = \tau(2^k x)$. This is generally not the case for other maps, and the tent map is unique in this way (XXX??? right? or is there some other bizarre map that behaves like this?) Thus, we have an important corollary here: **curves generated by iterated maps always transform under the three-dimensional representation, and not the higher-dimensional representations**. The higher-dimensional representations cannot be made conjugate to iterated maps, because if they could be, they'd transform under the three-dimensional rep, a contradiction. We've subtly failed to mention one limitation: not all iterated maps can be made conjugate to the tent map. It is not clear to this author, at this time, which maps can and cannot be made conjugate to the tent map, and answering this question should provide insight into the nature of iteration.

The rest of this section is devoted to examining some more examples of iterated maps that can be made conjugate.

A tad more interesting is the Isola Map[5], which provides a connection to Farey Numbers and continued fractions:

$$F(x) = \begin{cases} x/(1-x) & \text{if } 0 \le x \le 1/2 \\ (1-x)/x & \text{if } 1/2 \le x \le 1 \end{cases}$$

This map is conjugate to the tent map, although the conjugating function is not analytic: it is the Minkowski Question Mark:

$$F(x) = (?^{-1} \circ \tau \circ ?)(x)$$

Iterating on this map give curve shown in the figure 11 below. The symmetry generators are

$$g_{?^{-1}}(x) = (?^{-1} \circ g_D \circ ?)(x) = g_C(x) = \frac{x}{1+x}$$

and $r_{?^{-1}}(x) = r_C(x) = r_D(x) = 1-x$. These are the generators explored in a previous chapter. Notice that the local minima in the graph correspond to the Farey Fractions, of course. Of interest here is that these arise out of iteration: the Farey Fractions are the "pre-images" of the fixed point at zero. Given any Farey Fraction, the Isola map will iterate them to zero after a finite number of steps. Insofar as the Farey Fractions are dense on the unit interval, then the Julia Set of the iterated Isola Map is the whole unit interval.

As the above example shows, it is remarkably difficult to construct a conjugating function even when it is "obvious" that a map should be conjugate to the Tent map. We were lucky to know that the Minkowski Question Mark was what was needed: it would be very difficult to construct from first principles. The Isola map is deceptively simple: it has a simple analytic form, yet the conjugating function is highly singular. In view of this difficulty, it is of some merit to create a dictionary of conjugating functions and describe the structure of the space of functions that are conjugate to the Tent Map.

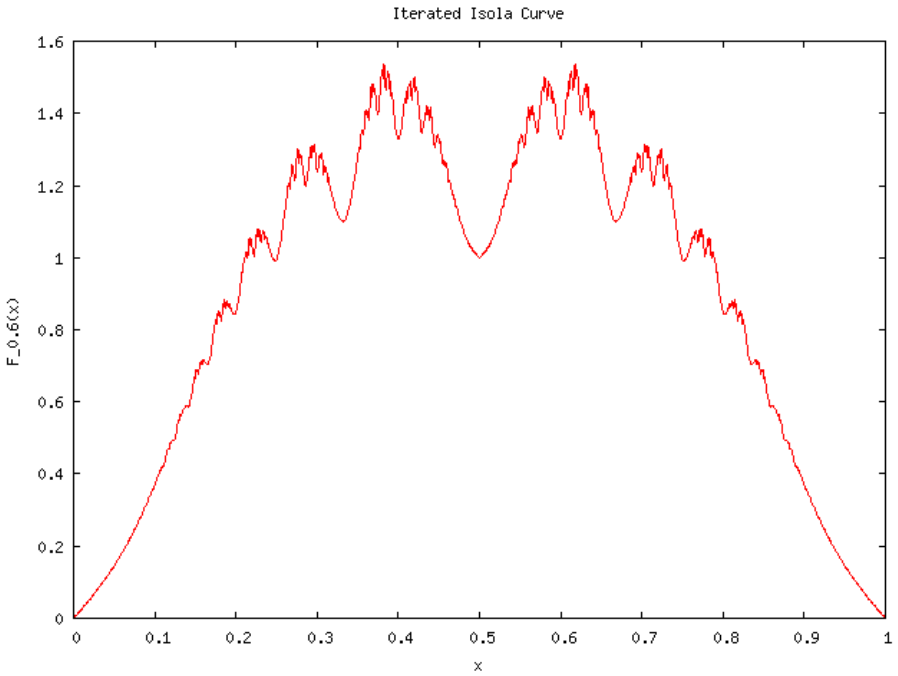
10 The de Rham Construction

In a classic 1957 paper[3], Georges de Rham constructs a class of curves, and proves that these curves are everywhere continuous but are nowhere differentiable (more precisely, are not differentiable at the rationals). In addition, he shows how the curves may be parameterized by a real number in the unit interval. The construction is simple. This section illustrates some of these curves.

Consider a pair of contracting maps of the plane $d_0 : \mathbb{R}^2 \to \mathbb{R}^2$ and $d_1 : \mathbb{R}^2 \to \mathbb{R}^2$. By the Banach fixed point theorem, such contracting maps should have fixed points p_0 and p_1. Assume that each fixed point lies in the basin of attraction of the other map, and furthermore, that the one applied to the fixed point of the other yields the same point, that is,

$$d_1(p_0) = d_0(p_1) \tag{15}$$

Figure 11: Iterated Isola Map



This figure shows the result of constructing the Takagi Curve for the iterated Isola Map. That is, the graph shows

$$F_w(x) = \sum_{k=0}^{\infty} w^k F^{k+1}(x) \tag{14}$$

with $F = ?^{-1} \circ \tau \circ ?$ and w=0.6.

Then these maps can be used to construct a certain continuous curve between $p_0$and p_1. Consider next the expansion in binary digits of a real number x, as given previously in eqn. 11:

$$x = \sum_{k=1}^{\infty} \frac{b_k}{2^k}$$

The de Rham curve is then a map characterized by the continuous parameter x:

$$d_x = d_{b_1} \circ d_{b_2} \circ \ldots \circ d_{b_k} \circ \ldots \tag{16}$$

The above map will map points in the common basin of attraction of the two maps to a single point. De Rham provides a simple proof that the resulting set of points form a continuous curve as a function of x, and that furthermore, this function is not differentiable in any conventional sense.

De Rham provides several examples. Let $z = u + iv$ and $a \in \mathbb{C}$ be a constant such that $|a| < 1$ and $|a-1| < 1$. Then consider the maps

$$d_0(z) = az$$

and

$$d_1(z) = a + (1-a)z$$

These two maps clearly have fixed points at $z = 0$ and $z = 1$, respectively. The generated curve is is the non-differentiable curve of Cesàro and Faber, now known more generally as the Lévy C-curve, especially when $a = 0.5 + i0.5$. See figures 12 and **??**.

Written as affine transformations, the two transforms can be expressed as

$$d_0(u,v) = \begin{pmatrix} 1 \\ u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & -\beta \\ 0 & \beta & \alpha \end{pmatrix} \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}$$

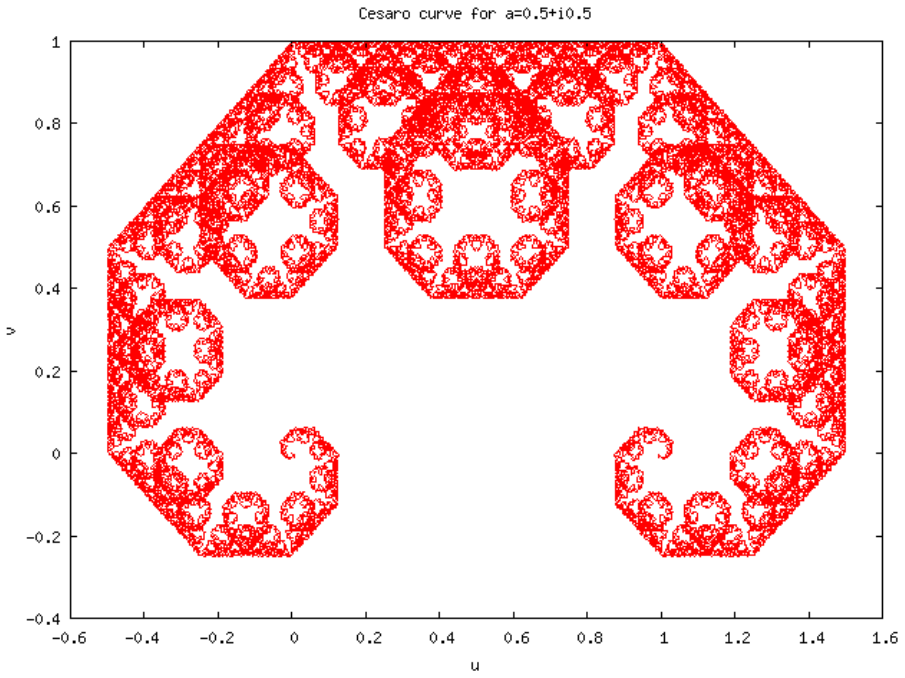
and

$$d_1(u,v) = \begin{pmatrix} 1 \\ u' \\ v' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1-\alpha & \beta \\ \beta & -\beta & 1-\alpha \end{pmatrix} \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}$$

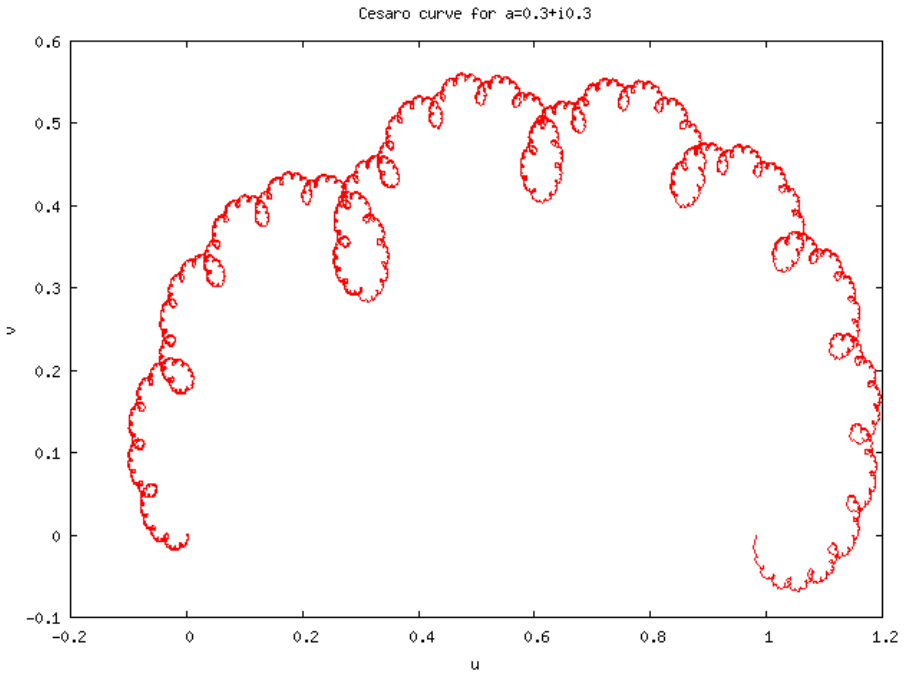
where $z = u + iv$ and $a = \alpha + i\beta$. If these look odd, it is because they are written "upside-down" from the way that affine transforms are commonly written. The reason for this will become clear below: it is to connect with the previous sections on the symmetries of the Takagi curve.

The d_0 and the d_1 transformations are essentially the L and R operators from the introduction of this paper. Although these are represented by three-dimensional linear transformations, there is no equivalent to the transformations g and r. This is because for every case, with the exception of $\alpha = 1/2$, the curves are not left-right symmetric, and so there is no matrix r such that $r^2 = 1$ and $L = rRr$. Furthermore, there is not even an s such that $L = s^{-1}Rs$. This last result is perhaps a bit unexpected, and to be

Figure 12: Cesàro Curve



The Cesàro curve, graphed for the value of $a = (1+i)/2$. Note that this figure is more commonly known as the Lévy C-curve.



The Cesàro curve, graphed for the value of $a = 0.3 + i0.3$. The real parameter shifts the symmetry point: thus the biggest loop is located at 0.3 in this picture, instead of being located at 0.5 as in the Lévy curve. The imaginary parameter provides a "strength" of the non-differentiability, playing a role similar to the w parameter in the Blancmange curve.

explored further. When $\alpha = 1/2$, the curves are left-right symmetric, and one has as a solution

$$r_{LC} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

the subscript *LC* denoting "Lévy C-curve". Notice that it is identical to the r_3 of equation **??**. The corresponding g_{LC} is of course

$$g_{LC} = d_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & -\beta \\ 0 & \beta & \frac{1}{2} \end{pmatrix}$$

Thus, there is a 1-parameter class of left-right symmetric Lévy C-curves. This is another, distinct three-dimensional representation for the period-doubling monoid.

The Takagi curve analyzed in the previous section can be generated in the same way, using

$$d_0 = L_3 = g_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 1 & w \end{pmatrix} \text{ and } d_1 = R_3 = r_3 g_3 r_3 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & -1 & w \end{pmatrix}$$

The Koch and Peano curves are similarly obtained, by introducing a mirror reflection through the complex conjugate:

$$d_0(z) = a\overline{z}$$

and

$$d_1(z) = a + (1-a)\overline{z}$$

Expressed in terms of affine left and right matrices, these are:

$$d_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \beta & -\alpha \end{pmatrix} \text{ and } d_1 = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1-\alpha & -\beta \\ \beta & -\beta & \alpha - 1 \end{pmatrix}$$

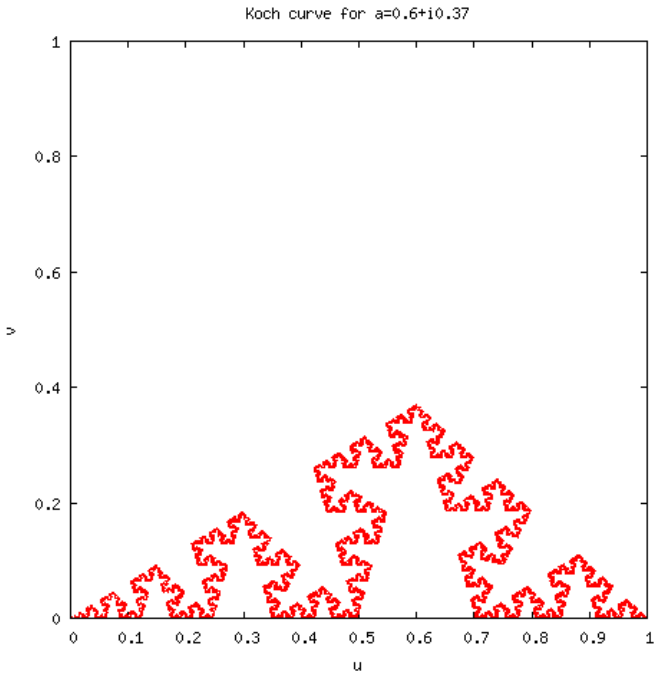
The classic Koch snowflake is regained for $a = \alpha + i\beta = 1/2 + i\sqrt{3}/6$ and the Peano curve for $a = (1+i)/2$. Values intermediate between these two generate intermediate curves, as shown in figure 13 and **??**.

As before, the curves are not left-right symmetric when $\alpha \neq 1/2$, and there is no non-trivial matrix s such that $sL = Rs$. However, when $\alpha = 1/2$, the curves are left-right symmetric, and we can find a matrix r such that $rL = Rr$ and $r^2 = 1$: in fact, this r takes the same form as r_3 before. Thus, the left-right symmetric Koch curve forms yet another 3D representation, given by

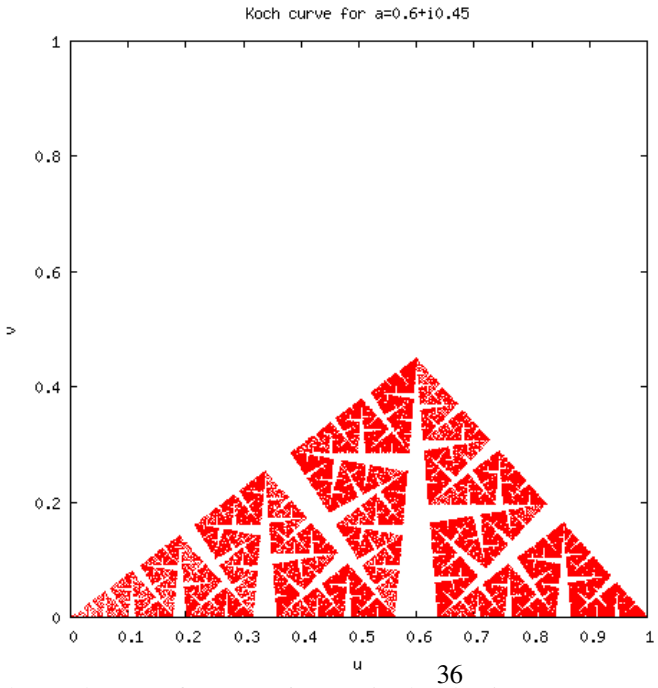
$$g_K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \beta \\ 0 & \beta & -\frac{1}{2} \end{pmatrix}$$

where the subscript *K* denotes "Koch".

Figure 13: Koch Curve



The Koch snowflake curve, constructed for $a = 0.6 + i0.37$. The classic, hexagonal-symmetry curve is regained by setting $a = 0.5 + i\sqrt{3}/6$, which centers the big point at 1/2, and opens the base of the point to run between 1/3 and 2/3'rds.



The Koch curve, for $a = 0.6 + i0.45$. The classic Peano space-filling curve is regained for $a = (1+i)/2$.

Yet another example of a curve generated by means of the de Rham construction is the Minkowski Question Mark function, which is given by the Mobius functions

$$d_0(z) = z/(z+1)$$

and

$$d_1(z) = 1/(z+1)$$

Expressed as the usual 2x2 matrix representations for Mobius transforms, these correspond to the generators L and R given in 1. More precisely, the generated function is actually half the inverse: $d_x = ?^{-1}(2x)$.

From the construction properties, it should now be clear that this generalized de Rham curve construction has the same set of modular-group self-similarities; this essentially follows from the self-similarity properties of the Cantor polynomials. That is, given a contracting group element $\gamma = g^{a_1} r g^{a_2} r g^{a_3} r ... r g^{a_N} \in PSL(2,\mathbb{Z})$, one defines its action in the canonical way, on the parameter space, as an action on dyadic intervals: thus

$$g d_x = d_{x/2}$$

and

$$s d_0 = d_1 \qquad d_0 = s^{-1} d_1$$

Note that the above is not just a statement about some particular value of x, but is rather a statement that holds true for the entire range of parameters $x \in [0,1]$; it is a statement of the self-similarity properties of the curve. In the case of the Koch and Lévy curves, both g and s, and thus any contracting elements γ are expressible as linear affine transformations on the two-dimensional plane. This is essentially an expression of a known result from IFS: these figures are obtainable by iterating on a pair of specific affine transforms.

Homework: write down an explicit expression for a general γ for the Koch and Lévy curves.

The lesson to be learned here bears stating clearly: ***every point on the above-mentioned curves can be uniquely labelled by a real number. The labelling is not abstract, but concrete. The fractal self-similarity of the curves are in unique correspondence to the contracting semi-group of the modular group. To every element of the contracting semi-group, a unique non-degenerate mapping of the plane can be given that exactly maps the curve into a a self-similar subset of itself. The mapping is continuous, and can be expressed in concrete form.***

The iteration of a pair of non-linear mappings to generate a de Rham curve seems to be an unexplored area of mathematics. There is a hint of richness.

10.1 The Total Number of Linear, Planar Dyadic Fractal Curves

The above exposition, in terms of the action of left and right affine transformations, indicates that it is possible to count the total number of uniquely distinct dyadic planar

fractal curves, and to classify them into families. The general linear, dyadic planar fractal curve is given by iterating on

$$d_0 = \begin{pmatrix} 1 & 0 & 0 \\ a & b & c \\ d & e & f \end{pmatrix} \text{ and } d_1 = \begin{pmatrix} 1 & 0 & 0 \\ h & j & k \\ l & m & n \end{pmatrix}$$

where $a, b, \ldots, n$ are taken as real numbers. From this general set, one wants to exclude the cases which are rotated, translated, scaled or squashed versions of one another. The general set appears to have twelve free parameters, from which should be excluded 1 (for rotations) + 2 (for translations) + 2 (for scaling) + 1(for shearing) = 6 non-interesting parameter dimensions. Requiring that the curve be continuous, by using de Rham's continuity condition 15, eliminates two more degrees of freedom. This leaves behind a four-dimensional space of unique fractal curves.

Of this four-dimensional space, one dimension has been explored with the Takagi curves. A second dimension is explored with the Cesàro curves, and a third with the Koch/Peano curves.

The general form may be narrowed as follows. Let d_0 have the fixed point p_0 located at the origin $(u,v) = (0,0)$. This implies that $a = d = 0$. Next let d_1 have the fixed point p_1 at $(u,v) = (1,0)$. This implies that $j = 1-h$ and $m = -l$. Finally, impose the de Rham condition for the continuity of the curve, namely that $d_0(p_1) = d_1(p_0)$. This implies that $h = b$ and $m = e$. Changing symbols, the general form with the endpoints fixed may be written as

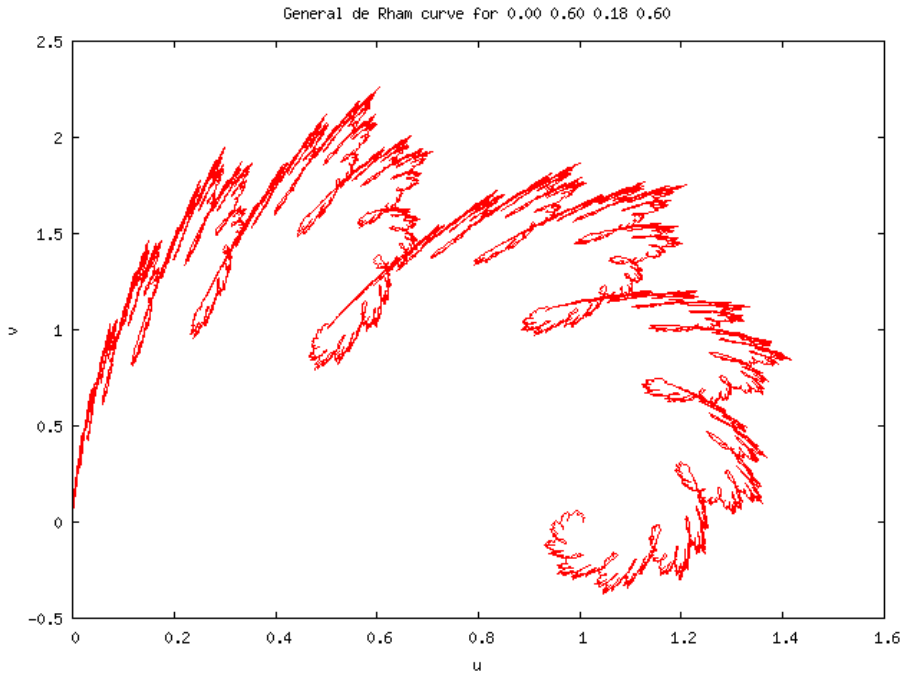
$$d_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & \delta \\ 0 & \beta & \varepsilon \end{pmatrix} \text{ and } d_1 = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1-\alpha & \zeta \\ \beta & -\beta & \eta \end{pmatrix} \tag{17}$$

The half-way point of this curve is located at $1/2 = 0.100\ldots = 0.011\ldots = d_1 d_0 d_0 \ldots = d_0 d_1 d_1 \ldots$ which can be seen to be $(u,v) = (\alpha, \beta)$. Using this last result to fix the location of the half-way point, what remains is a four-parameter family of linear planar fractal curves. Of this space, only about one-fourth the curves are "interesting", as the other half are mirror reflections in the horizontal axis, and another half are reflections about the $u = 1/2$ axis.

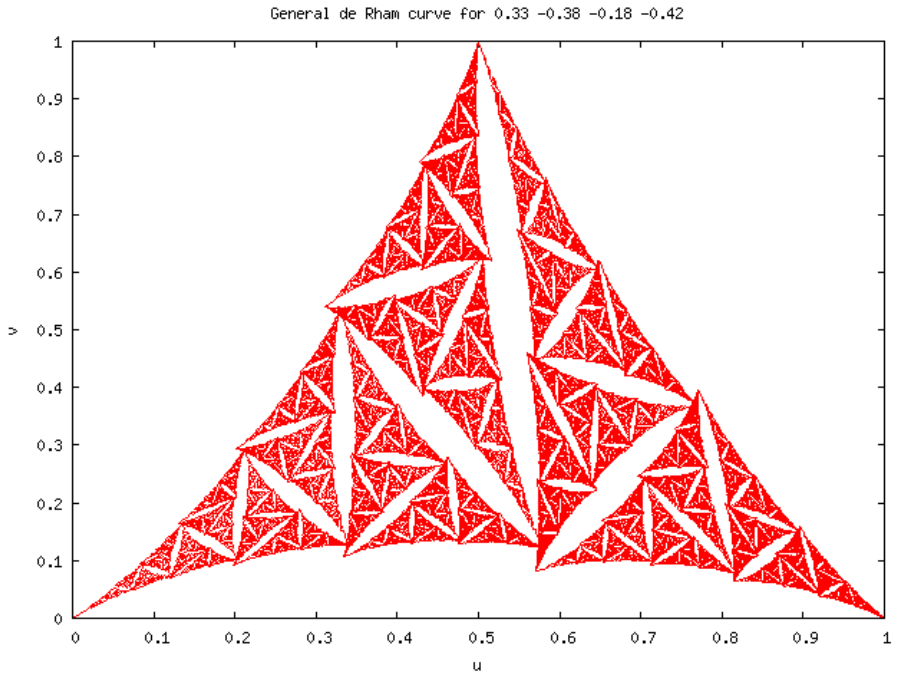
10.2 The Family of Left-Right Symmetric Curves

Of interest is the family of left-right symmetric curves. These are a subset of the above, and may be defined as follows: Given an expansion $d_0^a d_1^b d_0^c \ldots$ for the curve point (u,v), one wants the mirrored expansion $d_1^a d_0^b d_1^c \ldots$ to go to the curve point $(1-u,v)$. Applying this to the half-way point of the curve, one may immediately deduce that $\alpha = 1/2$. Applying this to the $1/4$ and $3/4$ points requires that $d_0 d_1(p_0)$ be mirrored to $d_1 d_0(p_1)$, from whence one deduces that $\eta = \varepsilon$ and $\zeta = -\delta$. Requiring additional symmetry at other points imposes no new constraints. Thus, the most general left-right symmetric plane curves generated by the three-dimensional representation are

Figure 14: Two generalized de Rham curves



A generalized de Rham curve for $\alpha = 0.5$, $\beta = 1$, $\delta = 0$, $\varepsilon = 0.6$, $\zeta = 0.18$, $\eta = 0.6$.



A generalized de Rham curve for $\alpha = 0.5$, $\beta = 1$, $\delta = 0.33$, $\varepsilon = -0.38$, $\zeta = -0.18$, $\eta = -0.42$.

generated by

$$d_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \delta \\ 0 & \beta & \varepsilon \end{pmatrix} \text{ and } d_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -\delta \\ \beta & -\beta & \varepsilon \end{pmatrix}$$

The Cesaro, Takagi and Koch curves may all be seen to be special cases of this form. It can furthermore be readily verified that $d_0 = rd_1r$, and so this family of left-right symmetric curves has a representation in terms of g and r.

10.3 Convergence

It should be clear that particularly large values of the parameters may lead to divergence upon iteration. To get absolute convergence, one wants the final form of the affine matrix $d_x = d_{b_1} \circ d_{b_2} \circ \cdots$ to be

$$d_x = \begin{pmatrix} 1 & 0 & 0 \\ u & 0 & 0 \\ v & 0 & 0 \end{pmatrix}$$

so that the fixed point is (u,v) independent of the starting point on the plane. In order to guarantee absolute convergence, it is clear that each of the two transforms should be contracting; this is guaranteed if the absolute value of each of the two eigenvalues of the two matrices

$$\begin{pmatrix} \alpha & \delta \\ \beta & \varepsilon \end{pmatrix} \text{ and } \begin{pmatrix} 1-\alpha & \zeta \\ -\beta & \eta \end{pmatrix}$$

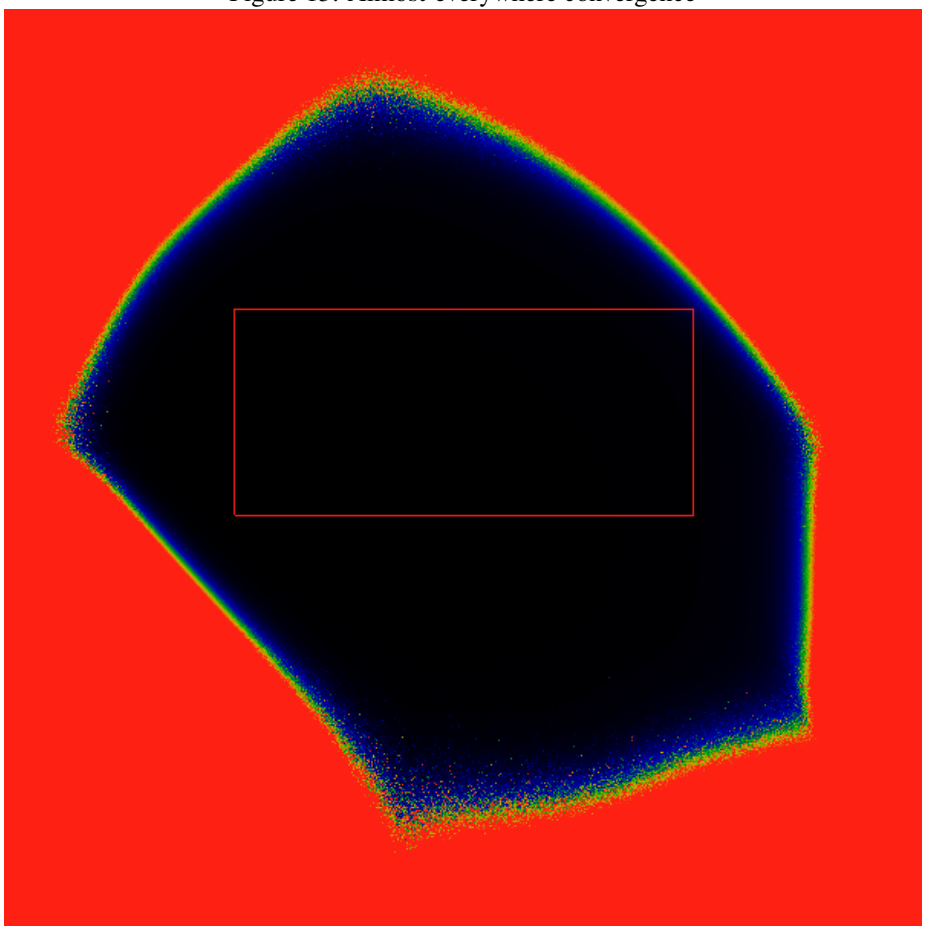
are less than one. The eigenvalues are readily obtained, and the parameter space is bounded by eight quadratic surfaces. One may wonder if there is a region of "almost-everywhere" convergence, and numerical exploration indicates that there is, as illustrated in the figure 15.

11 Markov chains

Markov chains are commonly studied in the theory of probability. Of all possible Markov chains, the chains for a stationary process are most frequently discussed, because they have a simple, tractable theory. In this section, we note that the de Rham construction can be used to create a set of non-stationary Markov chains that are none-the-less tractable. These are essentially a minor variant on the above construction, but are noted because of the importance of Markov chains in general.

A Markov chain may be defined in several ways. For the purposes here, a Markov chain may be defined as a sequence of $n \times n$ matrices, the probability transition matrices, having the property that they preserve the norm of probability vectors. A probability vector is very simply a vector $(p_1, p_2, \ldots, p_n)$ where all of the entries are real and positive, and sum to one. One may then deduce that the transition matrix must have entries that are real, and lie between zero and one, with the entries in each column summing to one as well. The sequence of matrices are multiplied together to form the

Figure 15: Almost-everywhere convergence



This figure shows a region of almost-everywhere convergence, with the region of absolute convergence inscribed as a perfect rectangle. The parameter space explored here has $\alpha = 0.5$, $\beta = 1$,$\delta = 0$ and $\eta = 0.6$ held fixed, ε varied along the horizontal, from -2 to +2, and ζ from -2 to 2 along the vertical. The black area indicates the region of (ε, ζ) for which equation 16 converged for all attempted 4621 trial values of x (that is, for all values $x = k/4621$with integer $0 < k < 4621$). The fuzziness around the edges and particularly the corners indicates that some care must be taken with the notion of everywhere: in this case, the measure of the set of diverging values can be argued to be less than 1/4621. This slice through the four-dimensional space seems "generic", in that other slices are not dissimilar (and in fact, all slices seem five-sided, indicating possibly some regular polytope).

chain. Properties of the chain that may be studied are expected value of the probability vector (the fixed point), the means-square variance.

It is readily seen that the general form of a two-dimensional dyadic chain will be

$$d_0 = \begin{pmatrix} a & b \\ 1-a & 1-b \end{pmatrix} \text{ and } d_1 = \begin{pmatrix} c & d \\ 1-c & 1-d \end{pmatrix} \tag{18}$$

The fixed point $p_0$is given by

$$p_0 = \begin{pmatrix} \frac{b}{1-a+b} \\ \frac{1-a}{1-a+b} \end{pmatrix}$$

and likewise for p_1. Although there is not overt requirement for the resulting de Rham curve to be continuous, one might ask what constraints continuity imposes. The answer is that one must have

$$d = c - (1 - a + b)$$

The requirement that $0 \leq d \leq 1$ translates into the condition

$$c + a \leq 1 + b$$

A pair of typical figures are shown in figure 16.

Similar considerations may be given for $n \times n$ matrices instead of 2×2 matrices.

12 Symbolic dynamics

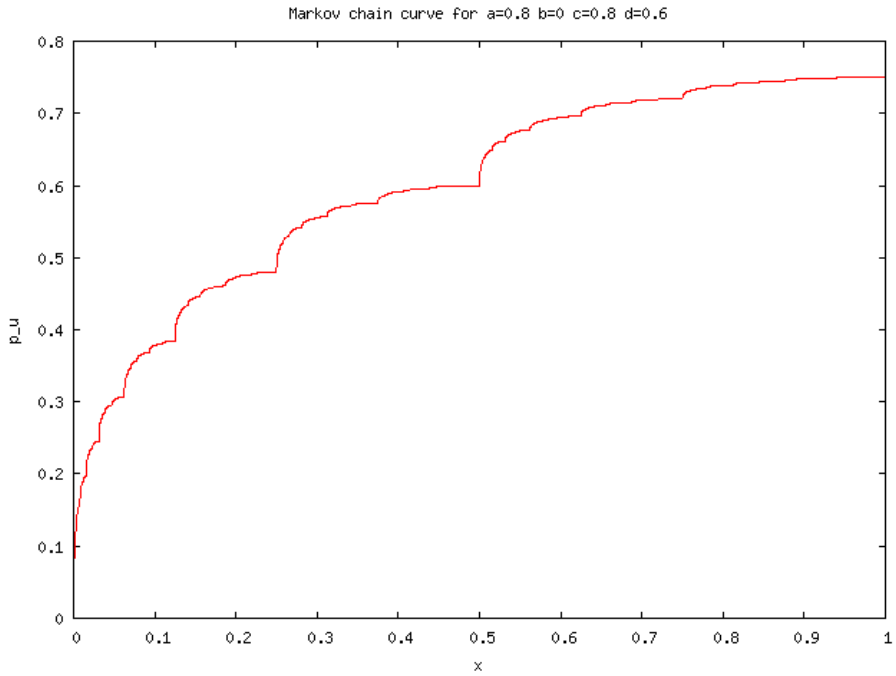
The construction and symmetry properties are not limited to continuous curves. Any system with a symbolic dynamics in two symbols is naturally parameterized through real numbers. Consider, for example, the symbolic dynamics of the tent map of unit height. By picking an initial value, and then iterating on the tent map, one will get a sequence of values. For each iteration, write down the letter *A* if the value is less than 1/2, and write the letter *B* if the value is more than 1/2. The result of the iteration will be an infinitely long string composed of the letters *A* and *B*, sometimes called the *orbit* of the initial value. This string is the so-called symbolic dynamics of the iterated unit-height tent map. It can be shown that the symbolic dynamics of the iteration precisely and completely defines the iteration.

What is notable here is that any string in two letters can be exactly mapped to a real number, by interpreting it as the dyadic expansion of that number. In this case, let c_k be zero if the k'th letter is *A*, and let it be one if the k'th letter is *B*. We then define the value of a symbolic string to be

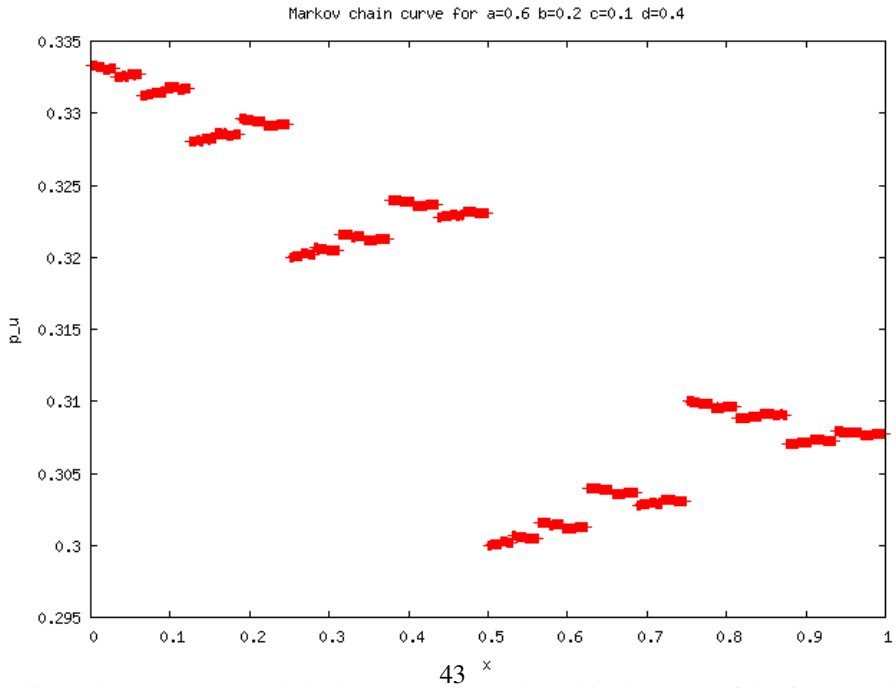
$$s(x) = \sum_{n=1}^{\infty} c_k 2^{-k} \tag{19}$$

where x was the initial value that was iterated. The figure 17 shows the graph of the above, after iterating the tent map of unit height.

Figure 16: Markov chain curves

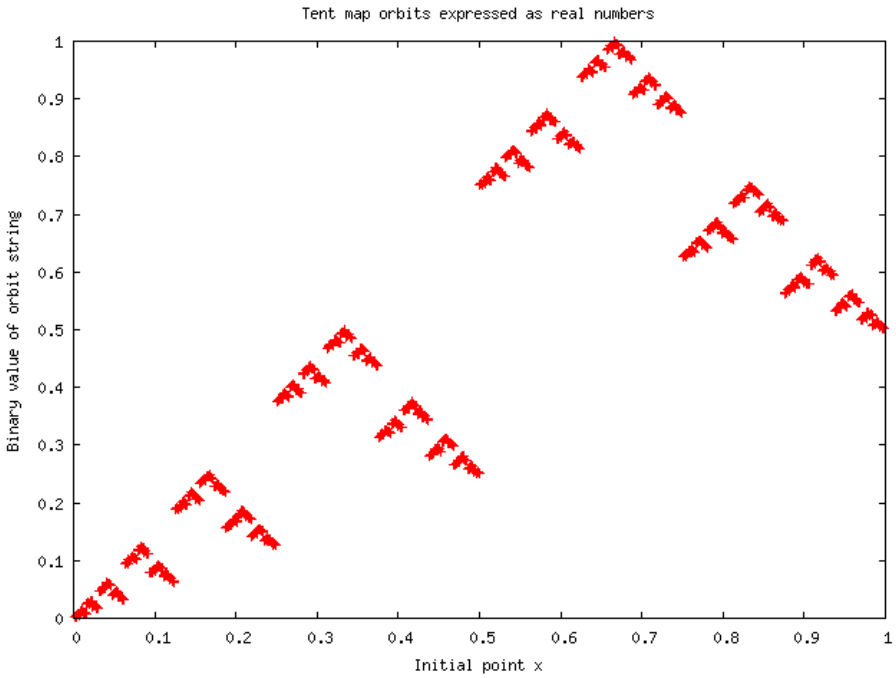


A continuous Markov chain de Rham curve. Plotted is the value of the first coordinate of the fixed point versus the curve coordinate x. This curve corresponds to values for equation 18 of $a = 0.8$, $b = 0$, $c = 0.8$, $d = 0.6$.



A discontinuous Markov chain de Rham curve. Plotted is the value of the first coordinate of the fixed point versus the curve coordinate x. This curve corresponds to values for equation 18 of $a = 0.6$, $b = 0.2$, $c = 0.1$, $d = 0.4$.

Figure 17: Tent Map



The graph of the values corresponding the orbit if an initial point, under iteration of the tent map. In fact, the orbits in this system are exactly solvable. If we expand the initial value as

$$x = \sum_{n=1}^{\infty} b_n 2^{-n}$$

then the c_k of equation 19 are given by

$$c_{k+1} = b_k \text{ XOR } b_{k+1}$$

for $k \neq 0$ and with $c_1 = b_1$. This result is easily obtained by noting that the unit-height tent map is just $\tau(x)$ of the previous sections, and that $\tau^{k+1}(x) = \tau(2^k x)$, as previously noted.

---

13 Miscellany

A compendium of miscellaneous related topics and to-do items that come to mind.

Homework: Curiously, the representations of g for the Takagi curve resemble the form that a translation would take in a linear projective space. That is, one has non-zero elements only along the diagonal, with the exception of the last row, whose non-zero elements can be interpreted as a translation in the given direction by a given amount. In other words, the matrices g look like affine transformations that encode a translation with a funny scale. By combining appropriately with r, show how to encode a rotation, while minimizing the shear.

Homework: The group elements generated by g then look like an iterated affine transformation, reminiscent of the Barnsley IFS constructions[1]. One can then take a set of arbitrary (contractive) group elements $\gamma_1, \gamma_2, ...\gamma_k \in SL(2,\mathbb{Z})$ and iterate on these, noting that they generate a subset of $SL(2,\mathbb{Z})$ that consists of contractive affine transformations. The homework problem is then to repeat the Barnsley analysis on affine transformations, and to restate his "Collage Theorem" for these affine maps, this time, not for 2D images, but rather for an arbitrary N-dimensional representation.

Homework: Describe the measure of the generated Barnsley IFS fractal set, as a function of the generators $\gamma_1, \gamma_2, ...\gamma_k$.

Homework: Affine transformations imply a projective space. That is, without the coupling to the Takagi Curve, we would be reasoning about matrix representations of polynomials, which is a very traditional subject. Can something be made out of this projective nature?

13.1 IFS Construction

The deRham construction is broadly applicable to pairs of iterated dyadic affine transformations, if one drops the requirement that the generated object be a continuous line. That is, given any pair of transforms d_0 and d_1, one can label a string of iterations of these with a real number expressed in dyadic form. Such transformations, known as *dimers*, include the Lévy dragon, the Harter-Heighway dragon, and a number of others, all of which are known to tile the plane. In essence, one associates to each map a measure; the sum of the measures reproduces the plane.

The curiosity of this construction is that each measure can be broken down, in turn, but a labelling of the real number corresponding to the string of iterations. What remains to be shown is that these measures are not only self-similar, but that the semi-group of self-similarities is again the contracting semi-group of the modular group. But, by its very construction, this demonstration is straightforward.

Homework: Specify the Hausdorff dimension of the generated measures, as a function of the affine parameters of the generating matrices, and the single linear dyadic parameter. This is a non-trivial exercise.

13.2 Weierstrass-Hardy Function

Perform the same analysis on the Weierstrass-Hardy function, which has an obvious fractal self-similarity; can we analyze it in the same way, using the same mechanisms? This analysis appears to generalize to most holomorphic functions. In a certain sense, most holomorphic functions have an innate symmetry of this type.

13.3 Batrachions and Ladders of Symmetry

A Batrachion is an integer sequence that hops from one integer to another using a simple recurrence relation. The Hofstadter-Conway 10,000-Dollar Sequence in particular is known to generate successive approximations to the Takagi Curve. That is, one has a "ladder" of approximations to

$$t_w(x) = \sum_{k=0}^{\infty} w^k \tau\left(2^k x\right)$$

by terminating the sum to a finite number of terms.

Homework: Describe the set-theoretic implications of having a ladder of progressive symmetries on Batrachion sequences. That is, at each level, one can have more and more elements from the modular group come into play, expressing the symmetries available at that level. Develop a general theory of ladders of subsets of the modular group, if possible.

XXX This finite-to-infinite-limit inter-relation is essentially the same phenomenon as the relationship between binomial coefficients (Pascal's Triangle) and the Sierpinski Gasket, shown by Pietgen etal. The 10,000-Dollar Sequence is essentially the a sum over the modulo function on rows of binomial coefficients. In the limit, the projection of the Sierpinski Gasket onto the unit interval yields a multi-fractal measure whose integral is essentially the Takagi Curve.

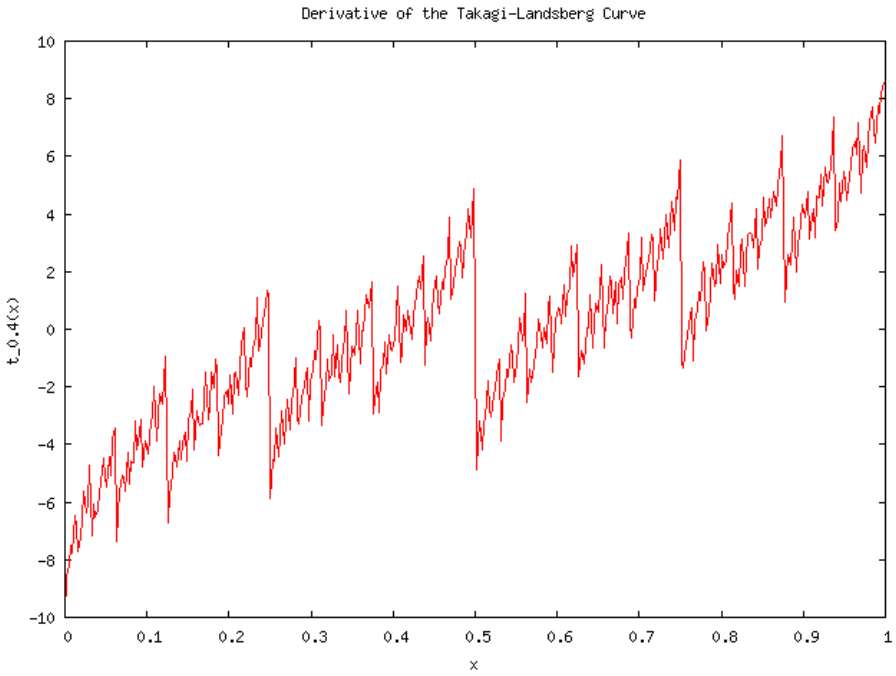
All of this discussion needs to be moved over to the chapter that covers the Sierpinski Gasket vs. Modular Group discussion.

13.4 Derivatives of the Takagi Curve

One can also explore the derivative of the Takagi Curve, which is reasonably well defined for $w < 1/2$; one takes the derivative under the summation, and notes the derivative of a triangle wave is a square wave. This is shown in figure 18. Derivatives for higher w require analytic continuation past $w = 1/2$, which is always possible if x is held fixed.

Homework: Work out the transformation of the derivative curve under the modular group. This can be bit bit confusing, because there are two potential candidates to distinguish. Given a general element $\gamma \in SL(2,\mathbb{Z})$ one can define $\gamma_1 d\overrightarrow{v}/dx \equiv d\gamma\overrightarrow{v}/dx$ for the vector $\overrightarrow{v} = a + bx + ct_w(x)$ as the rule for how the 3-rep elements transform, and one can also define $\gamma' t'_w(x) \equiv t'_w(\gamma x)$ as the rule for how the derivative transforms. Note that $\gamma' \neq \gamma_1$. Keep in mind that the derivative

Figure 18: Derivative of the Takagi Curve



This graph shows $dt_w(x)/dx$ for a value of $w = 0.4$.

is ill-defined only on the dyadics, which is a set of measure zero. The derivative is well-defined on a set of measure one.

13.5 3-period Mode Locking

Note that some chaotic maps are seen to have period-doubling paths to chaos that include a factor of three: viz. 3,6,12,24,... Perhaps this is a manifestation of the fact that the modular group can be written as the free product of the two and three cycles $\mathbb{Z}_2$ and $\mathbb{Z}_3$?

14 Analytic Continuation

So far, we have been looking at the Takagi curve as a function of x, holding w fixed. One may also explore the dependency on w for fixed x. For any fixed x, the Takagi curve is analytic in w, by construction, with radius of convergence of 1. One task is to describe the the set of x for which the function may be analytically continued to values of $|w| > 1$. There are many easy cases. For example, $t_w(0) = 0$ and $t_w(1/2) = 1$ are trivial, and $t_w(1/4) = w + 1/2$ is clearly entire. Continuing, we have

$$t_w\left(\frac{1}{2^n}\right) = \frac{1}{2^{n-1}}\,\frac{1-2^n w^n}{1-2w}$$

which is a polynomial of finite degree, without any poles (there is no pole at $w = 1/2$), and thus entire. Its not hard to see, for any integer p, that $t_w(p/2^n)$ will be a finite polynomial in w. By contrast,

$$t_w\left(\frac{1}{3}\right) = \frac{2}{3(1-w)}$$

has a simple pole at $w = 1$. Other values include

$$t_w\left(\frac{1}{5}\right) = \frac{2}{5}\cdot\frac{1+2w}{1-w^2}$$

and

$$t_w\left(\frac{1}{7}\right) = \frac{2}{7}\cdot\frac{1+2w+3w^2}{1-w^3}$$

and

$$t_w\left(\frac{1}{9}\right) = \frac{2}{9}\cdot\frac{1+2w+4w^2}{1-w^3}$$

and

$$t_w\left(\frac{1}{11}\right) = \frac{2}{11}\cdot\frac{1+2w+4w^2+6w^3+5w^4}{1-w^5}$$

all of which appear to have simple poles at various roots of unity. Note that for all of these cases, the limiting asymptotic behavior for large w is finite.

It could be interesting to provide a description in terms of the reduced characters. XXX do this.

Given any one such value, one may deduce the related function using formula 9 given previously:

$$t_w\left(\frac{x}{2^n}\right) = \frac{x}{2^{n-1}}\left(\frac{1-2^n w^n}{1-2w}\right) + w^n t_w(x)$$

Thus, at least for many rational values of x, it appears that the Takagi curve may be analytically extended to regions $|w| > 1$, where it appears to be a rational function of w.

However, the rational numbers are a set of measure zero in the reals; can anything be done for the irrationals? Some numerical explorations can provide insight.

Homework: Describe the analytic properties of $dt_w(x)/dx$ considered as a function of w, with x held fixed.

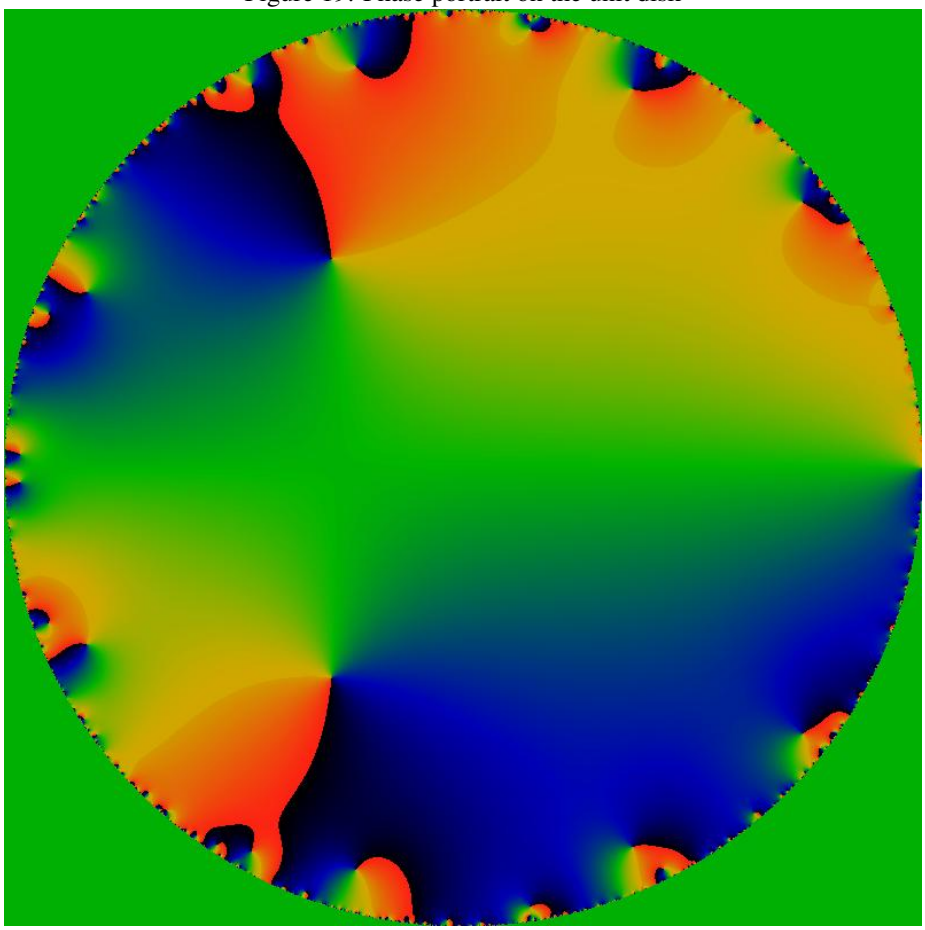
15 Conclusions

The seeming general applicability of this mode of analysis seems to beg the question, are there any self-similar curves that do NOT have the modular group symmetry?

At first, one is tempted to state that curves generated by *random* subdivision clearly cannot have this symmetry, but in fact, it is far from clear that this is true. As we've considered in an earlier chapter, the "gaps" of the continued fraction are *randomly* distributed on the unit square, and thus one has a Cauchy sequence that converges arbitrarily close to any given random value. The limit point is representable by a rational. This implies that *any* stochastic process on the unit interval can be represented by a strictly monotonically increasing sequence of rationals. This sequence of monotonically increasing rationals encodes the 'information' contained in the stochastic process, and through a standard δ, ε type proof, can be shown to be made arbitrarily accurate. But because this sequence of rationals was constructed explicitly through the use of continued fractions, it has inherently built into it the symmetry of the Modular Group. That is, *even random walks* have the modular group symmetry. This symmetry is clearly hidden (perfectly hidden, one might say) when looking at the stochastic process itself, but should manifest itself whenever the stochastic process is used for some purpose. For example, in a differential equation, the symmetry will show up as a scaling property or a power-law noise spectrum. For a fractal constructed randomly, for example, thorough DLA (Diffusion Limited Aggregation), a similar set of scaling and self-similarity relationships will become manifest. In the case of DLA, the ability to analyze in terms of the equipotential, and to draw the Duoady-Hubbard rays that map the dendrites to the unit circle are already the strongest hint that the underlying symmetry group is the modular group. The "gotcha" is that the symmetry difficult to manifest in practice, since it requires identifying the rational associated with a particular dendritic shape, which is a computationally difficult proposition.

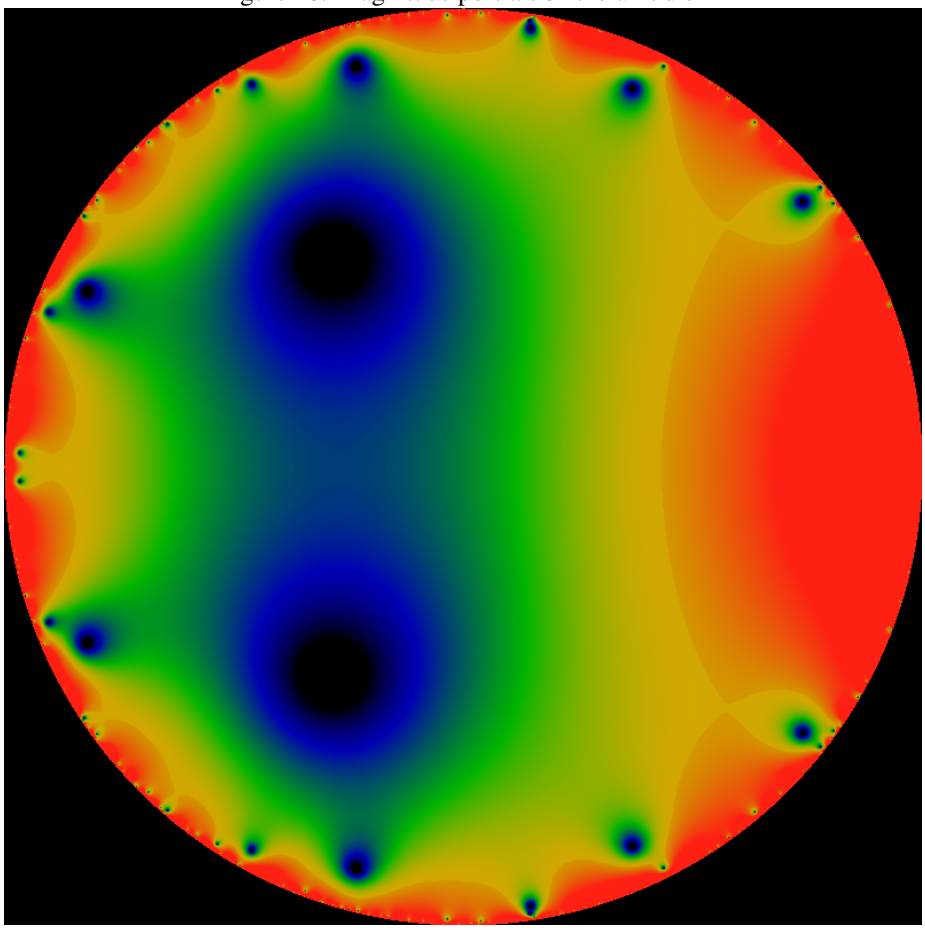
A different example of this manifestation of symmetry in randomness is through the Barnsley IFS ideas. There, a suitably chosen set of "random" affine matrices can

Figure 19: Phase portrait on the unit disk



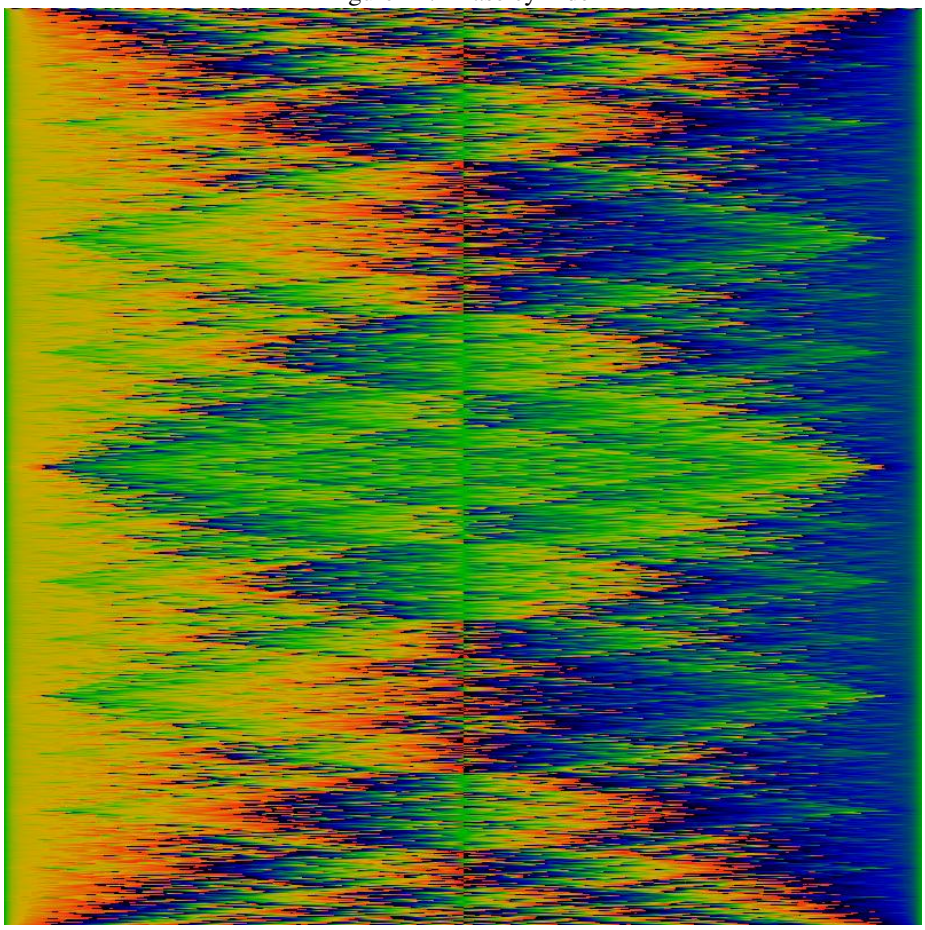
This figure shows a phase portrait of $t_w\,(0.8959848989828067)$on the unit disk $|w| < 1$. The color coding is such that black corresponds to $\arg t_w \approx 0$, green to $\arg t_w \approx \pi$ and red to $\arg t_w \approx 2\pi$. Simple zeros are clearly visible inside the disk, as endpoints of lines where the phase wraps from black to red. An arrangement of a large number of poles along the boundary $|w| = 1$ is shown in the next figure. The sums are not convergent for $|w| > 1$ and are not shown. This particular value of x is representative of what the disk likes for general, irrational x, although each and every value of x produces disks that have the zeros located in completely different places.

Figure 20: Magnitude portrait on the unit disk



This figure shows a magnitude portrait of $t_w(0.8959848989828067)$on the unit disk $|w| < 1$. The color coding is logarithmic, with black denoting $|t_w| \approx 0$ and red denoting poles. Zeros are clearly visible inside the disk, with an accumulation of poles on the boundary $|w| = 1$. The sums are not convergent for $|w| > 1$ and are not shown. This particular value of x is representative of what the disk likes for general, irrational x, although each and every value of x produces disks that have the zeros located in completely different places.

Figure 21: Phase cylinder



This figure shows a combined w and x portrait. Along the horizontal, $w = |w| \exp i\theta$ with $|w| = 0.94$ fixed, and θ runs from 0 to 2π. Along the vertical, x running from 0 to 1. The color coding shows the arg, with black corresponds to $\arg t_w \approx 0$, green to $\arg t_w \approx \pi$ and red to $\arg t_w \approx 2\pi$. The dependency on x is not only discontinuous, but depends very much on how x is sampled: choosing a vertical height of a prime number of pixels yields a different image than one where the eight is a composite number of small primes. In this case, a height of $n = 801$ pixels was chosen. This value was chosen because the resulting samples are more representative of having x lie on true irrationals.

be used to generate a remarkably varied set of self-similar images, such as fern-leaf shapes, maple-leaf shapes, and cloud-shapes. A more general theorem due to Barnsley states that any two-dimensional image (photographic snapshot) can be encoded (represented) with a set of affine transformations. For a while, there was great excitement with the idea of fractal IFS image compression as file-format for encoding images for the Internet or for general use as a compression algorithm. There apparently do exist some reasonably efficient algorithms that can take a photograph, and given a desired degree of accuracy $\varepsilon > 0$, extract the affine transformations that represent that photograph. By construction, the iterated affine transformations will be self-similar. But we saw above that any given set of "general, random" affine matrices correspond one-to-one to a set of of "general, random" (contractive) elements of $SL(2,\mathbb{Z})$. Through this exercise, we conclude that the self-symmetry of the Barnsley IFS fractals are in fact a manifestation of modular group symmetry, expressed through a choice of "random" generators.

From this exercise, one can now imagine that a suitably modified form of the Barnsley IFS image compression algorithm can be used to find the modular-group generators of the DLA fractals. To be precise, my intent is not merely to run snapshots of DLA through IFS, its rather to ask, what is the set of generators that will reproduce a given DLA tree, to a given accuracy? These generators then, by construction, correspond to elements of the modular group. One can then (potentially) create new theorems about DLA by picking "randomly" from the modular group, and exploring the consequences. This wouldn't be easy, but it would be interesting.

Pursuing the idea of symmetry in randomness in a different direction, its already known that stochastic differential equations have various scaling properties. We can now suspect that these scaling properties are the result of a "hidden" self-similarity. Let us now chose some arbitrary but small set of generators from the modular group, and use these to generate a self-similar, but superficially stochastic, seemingly random process. How does the stochastic differential equation behave upon input of any such given stochastic-but-self-similar process? Can we decompose the equation into a summation of eigen-equations, each reacting in a specific way to a specific generator of the stochastic noise? Realistically, this would be a rather difficult research program. However, I find that the potential for doing this sort of analysis surprising and intriguing.

Returning a bit closer to earth, we note the phase-locked regions of the Circle Map (see, for example, the pictures at the URL Circle Map `http://www.linas.org/art-gallery/basic/basic.html`) can be numbered with the Farey Numbers and are also qualitatively similar and suggestive of the gaps in continued fractions (see for example, the pictures at Farey Room Pictures `http://www.linas.org/art-gallery/farey/fthumb.html`). Reaching, but not too far, this implies that the Modular Group is the symmetry group of the Circle Map (the phase locked-loop of electronics) as well as the Mandelbrot Set. We then note that the mode-locked regions of the circle map are exactly those regions that are *not* chaotic: in other words, these are the so-called "islands of order" that can be found in the "sea of chaos". One of course sees such islands in other iterated equations, such as the iterated Feigenbaum map, and one is naturally lead to ask, is the modular group the symmetry group that maps islands into islands, for any period doubling map? What about the converse? Are there any maps, with islands, that do not have the Modular Group as its underlying

symmetry?

References

[1] Michael F. Barnsley. Fractal modelling of real world images. In Dietmar Saupe Heinz-Otto Peitgen, editor, *The Science of Fractal Images*, pages 219–242. Springer-Verlag, 1988.

[2] Georg Cantor. On the power of perfect sets of points (1884). In Gerald A. Edgar, editor, *Classics on Fractals*, pages 11–23. Addison-Wesley, 1993.

[3] Georges de Rham. On some curves defined by functional equations (1957). In Gerald A. Edgar, editor, *Classics on Fractals*, pages 285–298. Addison-Wesley, 1993.

[4] Dean J. Driebe. *Fully Chaotic Maps and Broken Time Symmetry*. Kluwer Academic Publishers, 1999.

[5] Stefano Isola. On the spectrum of farey and gauss maps. preprint, between 2000 and 2004.

[6] Benoit Mandelbrot. In Dietmar Saupe Heinz-Otto Peitgen, editor, *The Science of Fractal Images*, page 246. Springer-Verlag, 1988.

[7] Teiji Takagi. A simple example of a continuous function without derivative. *Proc. Phys. Math. Japan*, 1:176–177, 1903.

[8] Karl Weierstrass. On continuous functions of a real argument that do not have a well-defined differential quotient (1872). In Gerald A. Edgar, editor, *Classics on Fractals*, pages 2–9. Addison-Wesley, 1993.