Rationals, Fractals and Continued Fractions

Linas Vepstas <linas@linas.org>

12 October 2004

Abstract

Fractals and Continued Fractions seem to be deeply related in many ways. Farey numbers appear naturally in both. The shape of the Devil's Staircase, which occurs in phase-locking dynamic maps, is reminiscent of the Minkowski Question Mark Function. This article attempts to establish that the natural symmetry group of both period-doubling maps and of phase-locking maps is the modular group $SL(2, \mathbb{Z})$, which expains these curious relationships. We classify certain fractals as beloning to different representations of this group; opening the door to a group-theoretical attack on fractals. As a byproduct, I think we expose a fantastic fractal symmetry on the rationals $\mathbb{Q}$ themselves, and thus presumably also on $\mathbb{R}$ and $\mathbb{Q}_p$. This symmetry is never made use of in classical analysis, and probably not in non-Archimedean analysis either. It opens a whole set of new doors. An assortment of other curious observations is presented as well. Its a candy-store of surprises!

1 Intro

THIS IS A DRAFT WORK IN PROGRESS. The intro hasn't been written yet, but if it was, it would work like this: "I've been running a math-art website for over a decade and it's time some of the math behind it got an explanation." For my own benefit, if no one else's. So here it is.

Farey Numbers show up naturally in the analysis of the Mandelbrot set. Farey numbers are used to label the treads of the devil's staircase `http://mathworld.wolfram.com/DevilsStaircase.html`, which is the set of mode-locked regions of the circle map. The devils staircase bears a general resemblance to the Minkowski Question Mark function. Maybe have small pictures of minkowski and devil's staircase here.

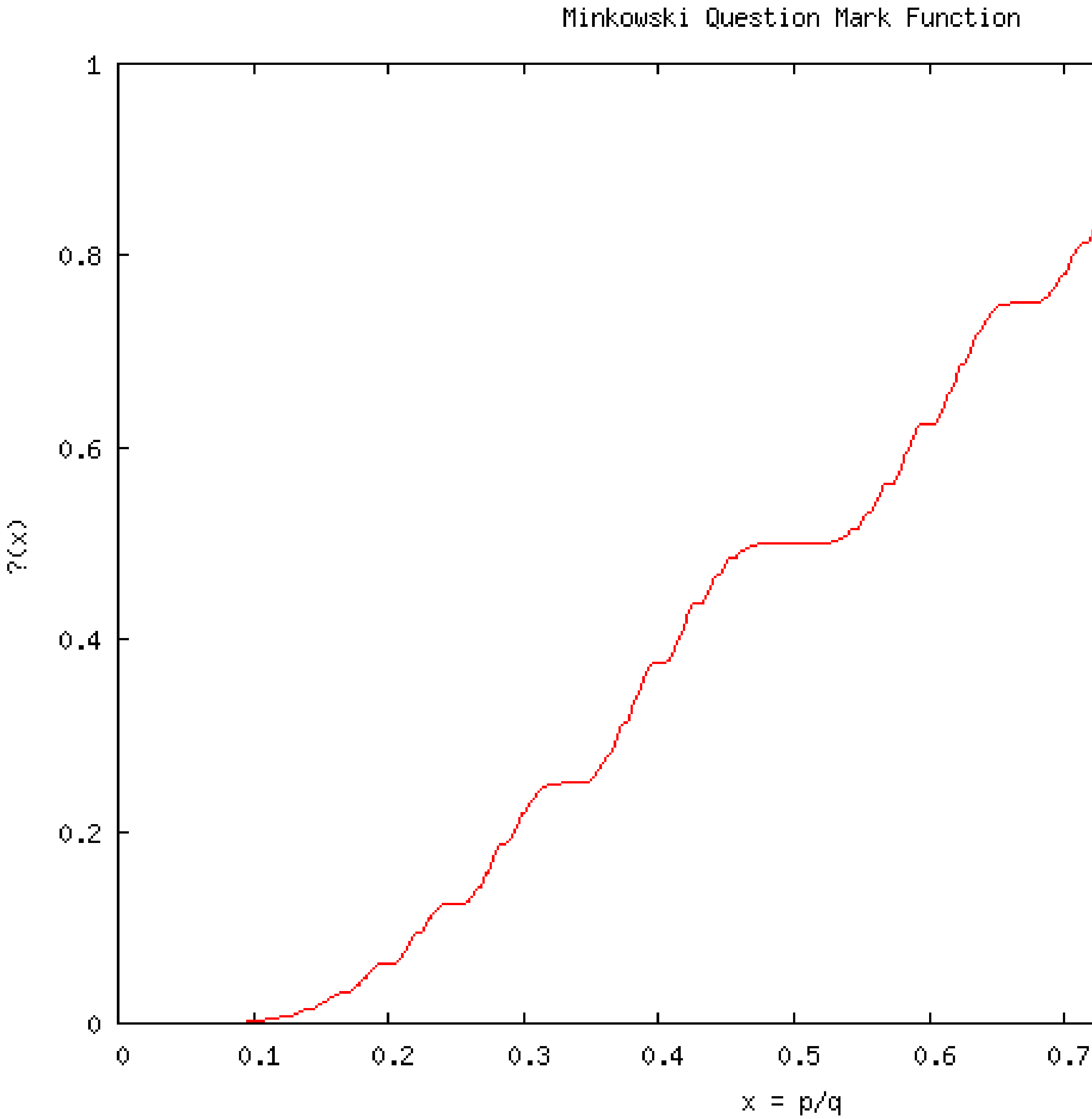
We should mention Sinai's tongue's here, which are the mode-locking regions, and note that they can be approximated with continued fractions, which is something we will show below.

It is the resemblance of these pictures that leads one naturally to the study of continued fractions. several remarkable things pop up: first, that Farey Numbers and continued fractions are deeply related, secondly, that Minkowski Question Mark provides a natural mapping to and from the Farey-number space, and finally, that the symmetry group of the Minkowski Question Mark is the Modular Group, as first pointed out by Georges de Rham in 1957[deR57], who made this observation while studying a class of self-similar Takagi-like curves. It doesn't take long to realize that the modular group is ***the*** symmetry group of a huge class of iterated maps, and can be applied to just about anything that exhibits period doubling or mode locking. In particular, this connection implies that one should explore interesting things in fractals, e.g. Misiurewicz points, and ask what they correspond to in the modular group. and etc. One then is lead to explore a veritable candy-store of possible connections between Modular Forms and Elliptic Curves and Fractals, and ask what each implies in the language of the other.

To get from here to there, we'll try to take a mostly pedagogical approach, keeping things simple and accesible to an amateur mathematician or an attentive undergraduate math-major, using only basic group theory, algebra and analysis.

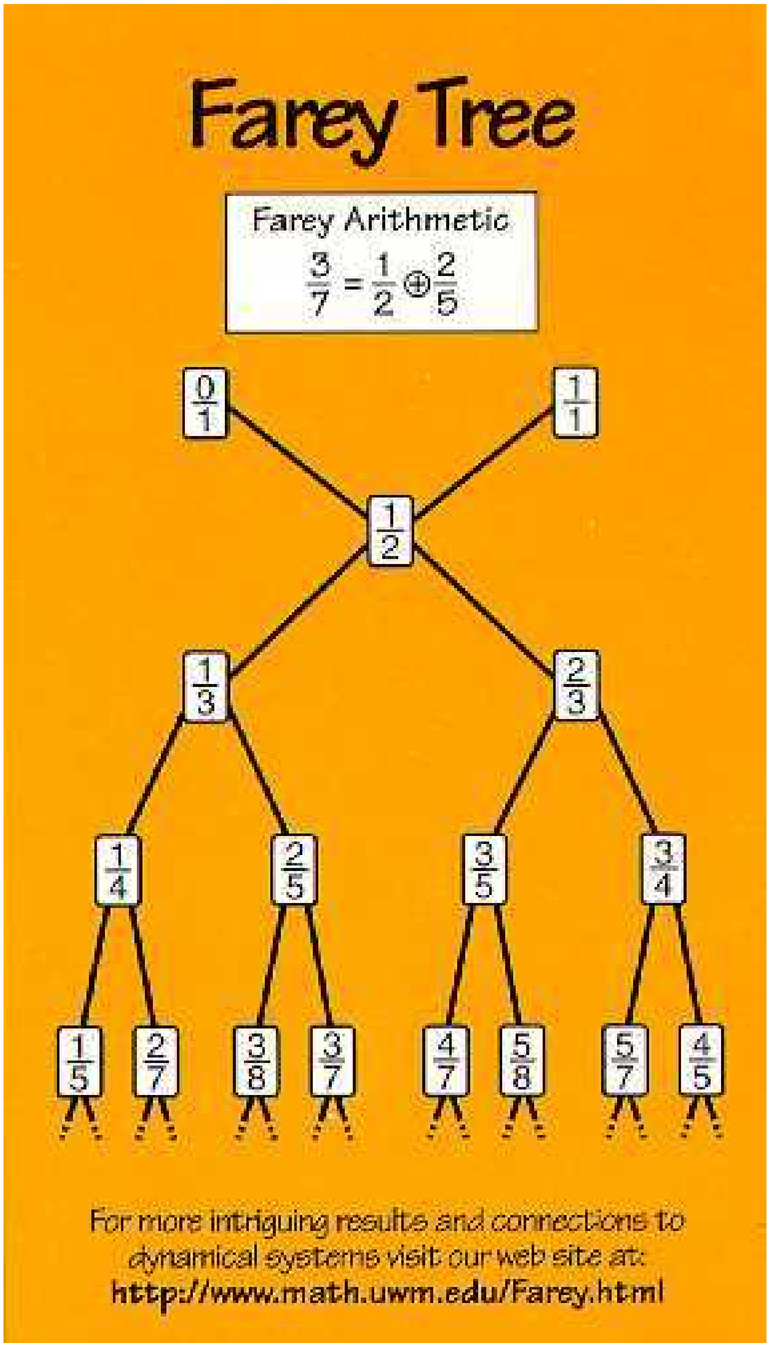
2 The Minkowski Question Mark Function

This section explores a peculiar mapping involving Farey Numbers. Refered to as the Minkowski Question Mark, it is well-defined for all real numbers and maps the unit interval into, and onto, the unit interval. It is continuous and monotonic. The map, pictured below, is highly self-similar. It is also highly singular: all of its derivatives vanish on the rationals. Its self-similarity group can be clearly defined, and essentially that of period-doubling iterated maps, such as the Mandelbrot Set. The fact that it involves the Farey numbers is essentially the "explanation" for why Farey Numbers occur in period-doubling maps.



2.1 Definition of the Question Mark Function

The Minkowski Question Mark map can be constructed in several ways. In all of these constructions, it can be understood to be the mapping between the Stern-Brocot Tree (sometimes called the Farey Tree) and the tree of real numbers expressed as binary, base-2, dyadic fractions. The Stern-Brocot Tree is the tree on which the Farey mediants are naturally organized, and is most easily explained with a picture:



(See also the presenatation on Stern Brocot Trees `http://www.cut-the-knot.org/blue/Stern.shtml` at Cut-the-Knot.) The dyadic tree has the same shape, except that the second row holds the values $1/4$ and $3/4$, and the next row the values 1/8, 3/8, 5/8 and 7/8, and so on. The dyadic tree holds values that are the arithmetic average, rather than the mediant, of the parent values. Both trees are 'binary' in the sense that each branch splits into two; we use the term 'dyadic tree' for the one that holds the 'dyadic' numbers $p/2^k$.

We denote the values of the Minkowski Question Mark function as as $?(x)$. Equating the Stern-Brocot Tree and the Dyadic Tree literally, one gets a recursive definition. Start by defining $?(0) = 0$ and $?(1) = 1$. Then, for a pair of Farey fraction parents p/q and p'/q', it equates the Farey mediant to the arithmetic average:

$$?\left(\frac{p+p'}{q+q'}\right) = \frac{1}{2}\left[?\left(\frac{p}{q}\right) + ?\left(\frac{p'}{q'}\right)\right]$$

Thus, starting with $\frac{0}{1}$ and $\frac{1}{1}$ at the root of the Farey Tree, we have $?\left(\frac{0+1}{1+1}\right) = ?\left(\frac{1}{2}\right) = \frac{1}{2}\left[?\left(\frac{0}{1}\right) + ?\left(\frac{1}{1}\right)\right] = \frac{1}{2}$. Next, we see the Farey number 1/3 is mapped to the binary 1/4, and so on: one quickly sees $?\left(1/n\right) = 2^{-n+1}$. Similarly, 2/3 maps to 3/4, and $?(n/(2n-1)) = 2^{-1} + 2^{-n}$. In general, a rational p/q maps to $?(p/q) = m \cdot 2^{-n}$ for some integers m, n (every rational occurs exactly once on the Stern-Brocot Tree).

Put another way, given any value x on the Stern-Brocot Tree, $?(x)$ is the value at the corresponding position on the Dyadic Tree. It is convenient in the following to think of $?(x)$ as a map from 'Farey Space' to 'Dyadic Space', and useful to keep in mind which of these two spaces one is currently working in.

An alternate, non-recursive definition for the Question Mark is based on continued fractions. Starting with the expansion $x = [a_1, a_2, ...]$, one writes

$$?(x) = 2\sum_{k=1} (-1)^k \; 2^{-(a_1+a_2+...+a_k)}$$

This sum can be visualized as a count of an alternating sequence of 0's and 1's in the binary expansion of $?(x)$:

$$?(x) = \quad 0.\underbrace{000...0}_{a_1-1}\;\underbrace{11..1}_{a_2}\;\underbrace{00..0}_{a_3}\;\underbrace{11..1}_{a_4}\;\underbrace{00..0}_{a_5}1..$$

In this expnasion, it is useful to remember that $0.111111... = 1.0000...$, in case one has an odd number of terms in the continued fraction expansion.

This sequence of 1's and 0's can be understood to be the set of directions for navigating a tree: on encountering a 0, take the left branch, else take the right branch. Thus, we can write $0.011 = LRR$. Navigating to this spot on the Stern-Brocot Tree, we find 2/5'ths. Navigating to this same position on the Dyadic Tree wee get 3/8'ths. Thus, $?(2/5) = 3/8$.

There is also a well-known representation of the Stern-Brocot Tree on the lattice $\mathbb{Z}\times\mathbb{Z}$, where a fraction p/q is denoted as the ordered pair (p,q). In this representation, the left-right navigation operators L and R have the values $L = \left(\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array}\right)$ and $R = \left(\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right)$ and thus positions are given as elements of $SL\left(2,\mathbb{Z}\right)$. Maybe we'll elaborate on this later xxxx. To do... elaborate.

To summarize, it is this expansion in binary digits that provides the underlying connection between period-doubling maps, such as the Mandelbrot Set, and Farey Numbers. Binary expansions, or code-words, occur naturally in the analysis of Duoady-Hubbard landing rays. We'll demonstrate an explicit mapping in a later section.

2.2 Some Curious Properties of the Question Mark

The Question Mark has the curious property that it maps quadratic irrationals to rational numbers. A quadratic irrational is a number of the form m $p/q + m\sqrt{D}/n$ for integers p, q, m, D, n. Continued fractions whose expansions become eventually periodic are mapped to binary strings that become eventualy periodic. The later are clearly just the binary expansions of rationals whose denominator is not a power of 2. The former are quadratic irrationals, which is known from the theory of Pellian equations: It turns out that every quadratic irrational has a continued fraction expansion that becomes eventually periodic. Note that, as a side effect, this proves that quadratic irrationals are countable in the same way that rationals are.

The Minkowski Question Mark has some very curious analytic properties: the mapping is continuous everywhere. It is also infinitely differentiable on every rational, and all derivatives vanish at every rational; yet the mapping is strictly montonically increasing. Clearly, the function is singular on irrationals. Its hard to define just how. (xxxx Blah blah need to define how).

If there is a 3-adic or p-adic generalization of the Minkowski Question Mark, it is not obvious; one 'obvious' generalization is

$$\sum_{k=1} (-1)^k \; 3^{-(a_1+a_2+\ldots+a_k)}$$

but its highly discontinuous. Other generalizations based on roots of unity in the complex plane also don't seem to work. One might be able to get traction by looking at groups that have $SL(2, \mathbb{Z})$ as a subroup but also have some p-fold symmetry.

2.3 Symmetries of the Question Mark

XXX write the intro again:... the symmetry is the group that maps trees to subtrees so the symmetry group of maping a binary tree onto itself in the modular group. XXX rewrite me. XXX also notice that the tree spans an inf-sup interval which is why this is an interval map and notice that trees don't span any arbitrary interval, but very specific intervals.

It should be immediately appearent from the picture of the Question Mark that it is highly self-similar, and thus should have a group of symmetries. Below, we give an explicit presentation of this group, which turns out to be the Modular Group Gamma. After breifly presenting the modular group a bit, we point out that the 'interesting' subset of the group, that is, 'interesting' as far as fractals are concerned, can be best enumerated as $\mathbb{Q}\otimes\mathbb{N}\otimes\mathbb{Z}_2$. This enumeration will help cement that the modular group is the symmetry group of the Mandelbrot Set. I'm enamored of this development, because the popular literature on fractals always talks about self-similarity without ever actually giving the group structure that's involved. The goal here is to rectify this omission.

We start by noticing that for any real number x having the continued fraction expansion $x = [a_1, a_2, ...]$, that $g(x) = x/(1 + x) = [a_1 + 1, a_2, ...]$. By the definition of $?(x)$ one sees immediately that g is a homomorphism of the Question Mark: $(? \circ g)(x) = ?(g(x)) = ?(x)/2$. Thus, g represents one of the generators of the symmetry group. Denoting repeated iterations of g by g^n we have $g^n(x) = x/(1 + nx)$. Equivalently, for a continued fraction, we have

$$g^n([a_1, a_2, ..]) = [a_1 + n, a_2, ...]$$

Iterating under the questino mark gives $?(x/(1+nx)) = ?(x)/2^n$. The generator g maps intervals to intervals, specifically $g^n : [[0, 1]] \to [[0, \frac{1}{n+1}]]$ where we use the non-standard notation $[[]]$ to denote an interval, to avoid confusion with $[]$ for continued fraction expansions. On the dyadic side, we have $(?g^n) : [[0, 1]] \to [[0, 1/2^n]]$.

One gets the general symmetry group Γ by combining with the reflection operator embodied in the left-right symmetry: $?(1 - x) = 1 - ?(x)$ which we denote with $r(x) = 1 - x$. Note that r commutes with $?$ as we have $(? \circ r)(x) = (r \circ ?)(x)$. Thus, for example,

$$(?rg^n)(x) = (r?g^n)(x) = r\left(\frac{1}{2^n}?\right)(x) = 1 - ?(x)/2^n = ?\left(r\left(\frac{x}{1 + nx}\right)\right) = ?\left(\frac{1 + (n - 1)x}{1 + nx}\right)$$

shows that rg^n is the group element that, under $?$, maps $[[0, 1]] \to [[1, 1 - 1/2^n]]$. We can now write a general group element as $\gamma \equiv g^{a_1}rg^{a_2}r...rg^{a_N}$. Applying the above process, it is relatively straightforward to see that

$$(?\gamma)(x) = (?g^{a_1}rg^{a_2}r...rg^{a_N})(x) = \frac{1}{2^{a_1}} - \frac{1}{2^{a_1+a_2}} + \frac{1}{2^{a_1+a_2+a_3}} - ... + (-)^{N+1}\frac{?(x)}{2^{a_1+a_2+a_3+...+a_N}}$$

To see the other leg of the homomorphism commuative diagram, we start by observing that we can write $g^n(x)$ itself as a continued fraction:

$$g^n(x) = \frac{1}{n + \frac{1}{x}} = [n, x]$$

In the same vein, we have $(rg^n)(x) = [1, n - 1, x]$ and, with a slight abuse:

$$r(x) = \frac{1}{1 + \frac{1}{-1+\frac{1}{x}}} = [1, -1, x]$$

Continuing this excercise, one finds that

$$\gamma(x) = (g^{a_1}rg^{a_2}r...rg^{a_N})(x) = [a_1 + 1, a_2, a_3, ..., a_{N-1}, a_N - 1, x]$$

which is recognizable as an element of the group $SL(2, \mathbb{Z})$. We present this group next.

2.4 The Modular Group Gamma or $SL(2,\mathbb{Z})$

The modular group gamma is normally defined by its canonical representation $SL(2,\mathbb{Z})$. In this standard representation, a general group element is representated by

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with a, b, c, d being integers, and $\det \gamma = ad - bc = 1$. Elements of this group are isomorphic to a certain set of Mobius transforms on the upper-half complex plane; the isomorphism is given by $\gamma(x) = (ax+b)/(cx+d)$. The group elements are generated by a pair of generators. One generator generates the cyclic group of two elements $C_2 \equiv \mathbb{Z}/(2\mathbb{Z})$ and the other generates the subgroup $\mathbb{Z}$. One can choose several different representations for the generators; the choice that is consistent with the previous section is

$$g = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}$$

This matrix representation of the generators can be used to manipulate and evaluate continued fractions; indeeed, the modular group is known as being one way to conveniently work with continued fractions. So, for example, we've already noted above that

$$g([a_1, a_2, ...]) = [a_1 + 1, a_2, ...]$$

Similarly, the action of r is can be written as

$$r([a_1, a_2, ..]) = \begin{cases} [1, a_1 - 1, a_2, ..] & \text{for} \quad a_1 \neq 1 \\ [a_2 + 1, a_3, ...] & \text{for} \quad a_1 = 1 \end{cases}$$

The operator to insert a digit $n \geq 1$ at the front of the continued fraction expansion is thus

$$(g^{n-1}rg)([a_1, a_2, ...]) = [n, a_1, a_2, ...]$$

which we can immediately verify as

$$(g^{n-1}rg)(x) = \left(\begin{pmatrix} 1 & 0 \\ n-1 & 1 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\right) : (x) = \begin{pmatrix} 0 & 1 \\ 1 & n \end{pmatrix} : (x) = \frac{1}{n+x}$$

Here we introduced : to denote the apply operator for the matrices; that is,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : (x) = \frac{ax+b}{cx+d}$$

It should be clear that we can now build up the continued fraction as desired with this insertion operator acting on zero:

$$(g^{j-1}rg)(g^{k-1}rg)(g^{m-1}rg)...(x) = (g^{j-1}rg^k rg^m rg...)(x) = [j-1, k, m, ..., x]$$

One can also find the append operator, which we find by examining

$$[a_1, a_2, ..., a_N, g^k(x)] = [a_1, a_2, ..., a_N + k, x]$$

and

$$[a_1, a_2, ..., a_N, r(x)] = [a_1, a_2, ..., a_N + 1, -1, x]$$

We use these to construct the appending operator $(g^{-1}rg^{k+1})(x) = (kx+1)/x$ which, for $k \geq 1$, acts as

$$[a_1, a_2, ..., a_N, (g^{-1}rg^{k+1})(x)] = [a_1, a_2, ..., a_N, k, x]$$

We can thus build up continued fractions by appending at the back, starting from $(g^{-1}rg)(x) = 1/x = [x]$ as follows:

$$(g^{-1}rg)(g^{-1}rg^{j+1})(g^{-1}rg^{k+1})(g^{-1}rg^{m+1})...(x) = (g^j rg^k rg^m rg...)(x) = [j, k, m, ..., x]$$

arriving at the same answer as before.

2.5 A Theorm about an Isomorphism

Lets restate the above results using modern mathematical notation and language. Let us denote with $\mathbb{R}_D$ the set of real numbers $\mathbb{R}$ that are represented by dyadic expansions; that is, $\forall x \in \mathbb{R}\ \exists x_D \in \mathbb{R}_D$ which represents the real number x as the dyadic expansion x_D with

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

where $b_k \in \{0,1\}\ \forall k \in \mathbb{Z}$. Let us use the notation $\mathbb{R}_C$ to denote the representation of the real numbers as continued fractions, that is, $\forall x \in \mathbb{R}\ \exists x_C \in \mathbb{R}_C$ such that

$$x_C = a_0 + \frac{1}{a_1 + \frac{1}{a_{2+...}}} = [a_0; a_1, a_2, ...]$$

where $a_k \in \mathbb{N}\ \forall k \in \mathbb{N}$. Both of these sets are commonly treated as being isomorphic to the real numbers; they are just different representations of the real numbers. However, we can see that these two are not completely the same, because the group $SL(2,\mathbb{Z})$ has a different action on each. In particular, the group generator g has the representation

$$g_D : \mathbb{R}_D \longrightarrow \mathbb{R}_D$$
$$g_D : x \longmapsto x/2$$

when acting on the set of dyadic numbers, whereas as this same group element has the representation

$$g_C : \mathbb{R}_C \longrightarrow \mathbb{R}_C$$
$$g_C : x \longmapsto x/(x+1)$$

when acting on the set of continued fractions. The other generator r has the same representation on both sets: $r(x) = 1 - x$.

Theorm: The Minkowski Question Mark provides an isomorphism between these two sets. That is, we have the commuting diagram

$$\begin{tikzcd} \mathbb{R}_C \arrow[r, "\gamma_C"] \arrow[d, "?"'] \arrow[dr, phantom, "\bigcirc"] & \mathbb{R}_C \arrow[d, "?"] \\ \mathbb{R}_D \arrow[r, "\gamma_D"'] & \mathbb{R}_D \end{tikzcd}$$

such that $? \circ \gamma_C = \gamma_D \circ ?$ holds $\forall \gamma \in SL(2, \mathbb{Z})$. To be more precise, we really should say that $(? \circ \gamma_C)(x_C) = (\gamma_D \circ ?)(x_D)\ \ \forall \gamma \in SL(2, \mathbb{Z})$ and $\forall x \in \mathbb{R}$ where $x_D \in \mathbb{R}_D$ and $x_C \in \mathbb{R}_C$ are both representations of the same x.

Proof: We've already given the standard representations of γ in these two spaces in previous sections, but we repeat these here just so that this is all crystal-clear. We first decompose an abstract group element $\gamma \in SL(2, \mathbb{Z})$ in terms of the generators g and r, as usual:

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

Then, this γ has the representation γ_D which is

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

and it also has the representation γ_C which is

$$\gamma_C(x) = [a_1 + 1, a_2, a_3, ..., a_{N-1}, a_N - 1, x]$$

and thus we are able to write $?(\gamma_C(x)) = \gamma_D(?(x))$ as the isomorphism. To properly complete this proof, we would need to apply induction, that is, to show that if this equivalence holds for any one given γ and x, then it also holds for $\gamma\gamma'$ for another $\gamma' \in SL(2, \mathbb{Z})$ and that it also holds for $x + y$ where $y \in \mathbb{R}$. We then would need to exhibit one group element for which this holds, and thus, by induction, conclude that it holds for all group elements and reals. One could try to give a very precise proof along these lines, but this is outside the scope of this text. It requires, among other things, talking about the completion of the rationals by the reals, and, if we are not careful, might also invoke the Axiom of Choice. See also the remarks below. QED.

To summarize, we have exhibited an isomorphism of the real numbers onto themselves that demonstrates that there are really two distinct and inequivalent representations of the real numbers. When we say distinct and inequivalent, we really mean it: these two representations transform radically differently under the Modular Group, although they are both isomorphic to that abstract set we call the reals, and are, through the Question Mark, isomorphic to each other.

If we were better mathematicians, we would insist on a more watertight proof of this isomorphism, rather than the assertion that this is so. Unfortunately, we do not have the tools handy to provide this stronger proof. It is widely beleived that both the dyadics and the continued fractions are isomorphic to the reals, but this shouldn't entirely be taken for granted. In fact, we have already commited several grave errors. For one, the set $\mathbb{R}_D$ is larger than the set $\mathbb{R}$: as every student is told, the set $\mathbb{R}_D$ contains the elements 0.11111... and 1.0000... which are clearly distinct in $\mathbb{R}_D$ but represent the same number in $\mathbb{R}$. We have something similar happening in $\mathbb{R}_C$ in that a rational has the multiple representations $[a_0; a_1, a_2, ..., a_n, \infty, a_{n+2}, a_{n+3}, ...]$ where it does not matter what we pick for a_{n+2}, etc. Thus, one of our mistakes was to say $\mathbb{R}_D$ is isomorphic to $\mathbb{R}$. Instead, it is only homomorphic, and there exists a homomorphism $e : \mathbb{R}_D \to \mathbb{R}$ such that $e(0.1111...) = e(1.000...)$, thus implying that $Ker\ e$ is non-trivial. We can then use $Ker\ e$ to induce the desired isomorphism between $\mathbb{R}_D$ and $\mathbb{R}$. The existance of this homomorphism is widely accepted, but I don't know how to get there from first principles. Similar remarks apply to the set $\mathbb{R}_C$. Thus, having to work with quotient spaces introduces complexity into the correct proof. To this volatile mix, we have to add considerations of about the completion of the rationals by the reals. Clearly, the dyadic tree and the Stern-Brocot tree are order-preserving enumerations of the rationals. The correct completion of the dyadic tree appears to be $\mathbb{R}_D$ whereas the correct completion of the Stern-Brocot Tree appears to be $\mathbb{R}_C$, and these two completions are not *prima facae* equivalent. If we can prove that there is only one, unique completion of the rationals by the reals, then we have proved the existance of the Minkowski Question Mark, and we thus arrive at the proof of the homomorphism of $\mathbb{R}_C$ and $\mathbb{R}_D$. But this is far from obvious from this vantage point. One problem is that proofs of completion might surreptiously invoke the Axiom of Choice; what one really wants are proofs of completion that start from a clean base, such as Zermelo-Fraenkel, and arrive at the homomorphism of $\mathbb{R}_C$ and $\mathbb{R}_D$, without invoking the Axiom of Choice. We start laying some of the groundwork for a proof of this type in a later section.

2.6 Fractal Symmetry and the Subset $\mathbb{Q} \otimes \mathbb{N} \otimes C_2$ of the Modular Group

Every group element of $SL(2, \mathbb{Z})$ can be decomposed into a combination of the g and r generators, of the form $g^{a_1} r g^{a_2} r ... r g^{a_N}$ for integers $a_k \in \mathbb{Z}$. We've already discovered the relationship between the modular group and continued fractions, which we gave above as

$$\gamma(x) = (g^{a_1} r g^{a_2} r ... r g^{a_N})(x) = [a_1 + 1, a_2, a_3, ..., a_{N-1}, a_N - 1, x]$$

We saw that this relation gives us the isomorphisms of the Minkowski Question Mark; let us now look at this a little more crefully and precisely, since, among other things, we need to define the domain and range over which the isomorphism is valid.

First, let us consider the group elements with $a_1 \geq 0$, $a_2 \geq 1$, ... $a_{N-1} \geq 1$, $a_N \geq 0$. When the coefficients are zero or positive like this, we find that $\gamma(x)$ maps the unit interval into strictly small intervals; *i.e.* that $0 < \gamma(0), \gamma(1) < 1$. One endpoint of the interval are given by the rational $\gamma(0) = [a_1+1, a_2, a_3, ..., a_{N-1}]$ (note the missing a_N). The other endpoint is $\gamma(1) = [a_1+1, a_2, a_3, ..., a_N]$ when $a_N \geq 1$, although it must be written as $\gamma(1) = [a_1+1, a_2, a_3, ..., a_{N-2}]$ when $a_N = 0$. This last special case is really just a reversal of the two endpoints. For such a 'shrinking' γ, we can now write the isomorphism of the question mark as

$$?(\gamma(x)) = ?(\gamma(0)) + \frac{(-1)^{N+1}}{(2^{a_1+a_2+...+a_N})} ?(x)$$

The best way of understanding this isomorphism is to think of it as an isomorphism of intervals, where we can pick one end of the interval, but are a bit more constrained in picking the other end. That is, we can pick an arbitrary $\gamma(0) = p/q \in \mathbb{Q}$. We get the other end of the interval by picking $a_N \in \mathbb{N}$ a positive integer. Having thus specified the interval, we can, for good measure, choose to exchange the endpoints, that is, pick one of $\{1, r\} \in C_2$.

XXXX explain this: there are two ways to enumerate the tree remaps: tree-root to -tree-root. But the intervals are inf and sup of all the elements in the tree. The alternate map is to specify the inf of the tree (which is actually a real not a rational!!!) in which case one has a choice of tree roots, which are Z. The mirror reflection of the tree is of couse C_2. XXXX reword the above XXXX

To summarize, this set of isomorphisms of the question mark can be enumerated by $\mathbb{Q} \otimes \mathbb{N} \otimes C_2$, which we'd like to call the fundamental set of symmetries. Note that this property of being able to freely pick one endpoint, but being a bit more contstrained with the other is a very important property of this symmetry. ***This property is the fundamental property of self-similar objects and fractals in general: one can freely pick which copy one is to be self-similar with, but then after that, one has little further freedom.*** This is what makes fractals be what they are.

So far, we've only enumerated some, but not all of the modular group. The inverse of our generator g is $g^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. For the group element $\gamma = g^{a_1} r g^{a_2} r ... r g^{a_N}$ we can quickly see that the inverse is $\gamma^{-1} = g^{-a_N} r g^{-a_{N-1}} r ... r g^{-a_1}$. Its easy to see that these define the inverse mappings of smaller intervals onto the unit interval. Note that one needs to be careful when working with $\gamma^{-1}(x)$ because this has poles that are just outside of the interval of the mapping. That is, the domain of the isomorphism is defined only on the interval of the mapping, and not the whole unit interval (its the range, not the domain, that is the whole unit interval). These inverse mappings can also be enumerated by $\mathbb{Q} \otimes \mathbb{N} \otimes C_2$, of course.

All of the remaining group elements that we haven't yet discussed are those with a mixture of positive and negative values for the a_k. These group elements can be obtained by alternately concatenating interval-shrinking and interval-expanding group elements. One has to be considerably more careful in keeping track of the domains and ranges of each mapping. We claim (without proof at this time) that these types of mappings do not introduce any new types of intervals, and so enumeration originally provided, *viz.* $(\mathbb{Q}\otimes\mathbb{N}\otimes C_2)\oplus(\mathbb{Q}\otimes\mathbb{N}\otimes C_2)$ is essentially unchanged.

ToDo: give for example $(rg^{-2}rg^2)(x) = x+2$ and discuss what this means for the question mark, and how we should treat this (to modulo or not to modulo?)

Note also that if we limit ourselves to alternately shrinking and growing interval maps that are stricly contained one inside the other, then we find a proper subroup of the modular group, whose group elements are now enumerated by pairs of intervals or .xxx. be careful here . Again, the domain and range of these sub-symmetries are now limited.

Thus I think we can now conclude that γ is an isomorphism of ?, mapping the unit interval into a subinterval, completing the proof. xxxx finish writing these conclusions.

2.7 Hyperbolic Rotations of Binary Trees

Note that by imposing the modular group symmetry on the real number line, we've essentially introduced a hyperbolic manifold that is homomorphic to the real-number line. The existance of this hyperbolic manifold and its negative curvature esentially 'explains' why trajectories of iterated functions have positve Lyapunov exponents. Of course they do, since thier 'true' trajectories should be consideed to live on the hyperbolic manifold rather than on the real-number line. We try to make this clear below.

One can get a much better sense of the hyperbolic nature induced by this symmetry group by looking at the discrete 'rotations' of the binary tree. Rotations are frequently used in computer science algorithms to rebalance finite binary trees while at the same time preserving the order of the elements in the tree. We can think of a rotation as kind of like draping the flexible, droopy tree over a peg, letting gravity do its job, and declaring the node on the peg as the new root of the tree. In fact, we have to do a bit of minor surgery to get this right; we ave to cut a branch and re-attach it at the free spot where the old root used to be.

Thus for example, lets rotate the dyadic tree so that the node at $1/4$ becomes the new root. We do this by chopping off the tree rooted at $3/8$'ths and re-attaching at as the left subtree of the old root at $1/2$. (xxx we desperately need a diagram here) Denoting this rotation with the symbol θ, we have an isomorphism of trees: that is,

$$\theta\left(\frac{1}{2}\right) = \frac{1}{4}$$

$$\theta\left(\frac{1}{4}\right) = \frac{1}{8} \qquad \theta\left(\frac{3}{4}\right) = \frac{1}{2}$$

$$\theta\left(\frac{1}{8}\right) = \frac{1}{16} \quad \theta\left(\frac{3}{8}\right) = \frac{3}{16} \quad \theta\left(\frac{5}{8}\right) = \frac{3}{8} \quad \theta\left(\frac{7}{8}\right) = \frac{3}{4}$$

and the rest of the tree hanging as normal under these nodes. Its not hard to see that this rotation is order-preserving, that is, $\theta(x) < \theta(y)$ whenever $x < y$ and ths monotonic, one-to-one and onto. This is the strecth-and-shrink map

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

which is stretching in one interval and shrinking in another. Since we've seen that the Modular Group maps intervals to intervals, some stretching and some shrinking, we immediately recognize that three group elements were used to construct this map. Since we've already enumerated the modular group in terms of maps of intervals, we see that the three different group elements making up this map are the one that re-parented the 1/8 tree at 1/4, the 3/8 tree at 5/8 and the 3/4 tree at 7/8. In terms of tree surgery, it is enough to specify these three remapings, as the rest of the binary trees hanging below are structurally unaltered. Note that the 3/8'ths to 5/8'ths mapping does not change the denominator: this map neither stretches nor shrinks, its a lateral translation.

Note that the map is clearly invertible, with its inverse being

$$\theta^{-1}(x) = xxx$$

Notice that the shrinking portion of the map is just the inverse of the stretching portion of the map. Iterating these maps just makes them even more hyperbolic. The figure1 shows some of these, with crosses denoting the endpoints of the straight-line segments.

The reflection map $r(x) = 1 - x$ conjugates θ to its inverse, that is, $r\theta^n r = \theta^{-n}$.

We demonstate a more complex mapping by rebalancing the binary tree so that the 3/8 node becomes root. This requires two tree movements: switching the 7/16 node so that it lives under 1/2, and switching the 5/16 node so that it lives under the now empty slot at 1/4. This is showin in figure (XXX need that figure.) This rotation is given by the mapping

$$\eta(x) = \begin{cases} x & \text{for} \quad 0 \le x \le 1/4 \\ x/2 + 1/8 & \text{for} \quad 1/4 \le x \le 3/4 \\ 2x - 1 & \text{for} \quad 3/4 \le x \le 1 \end{cases}$$

xxx stuff

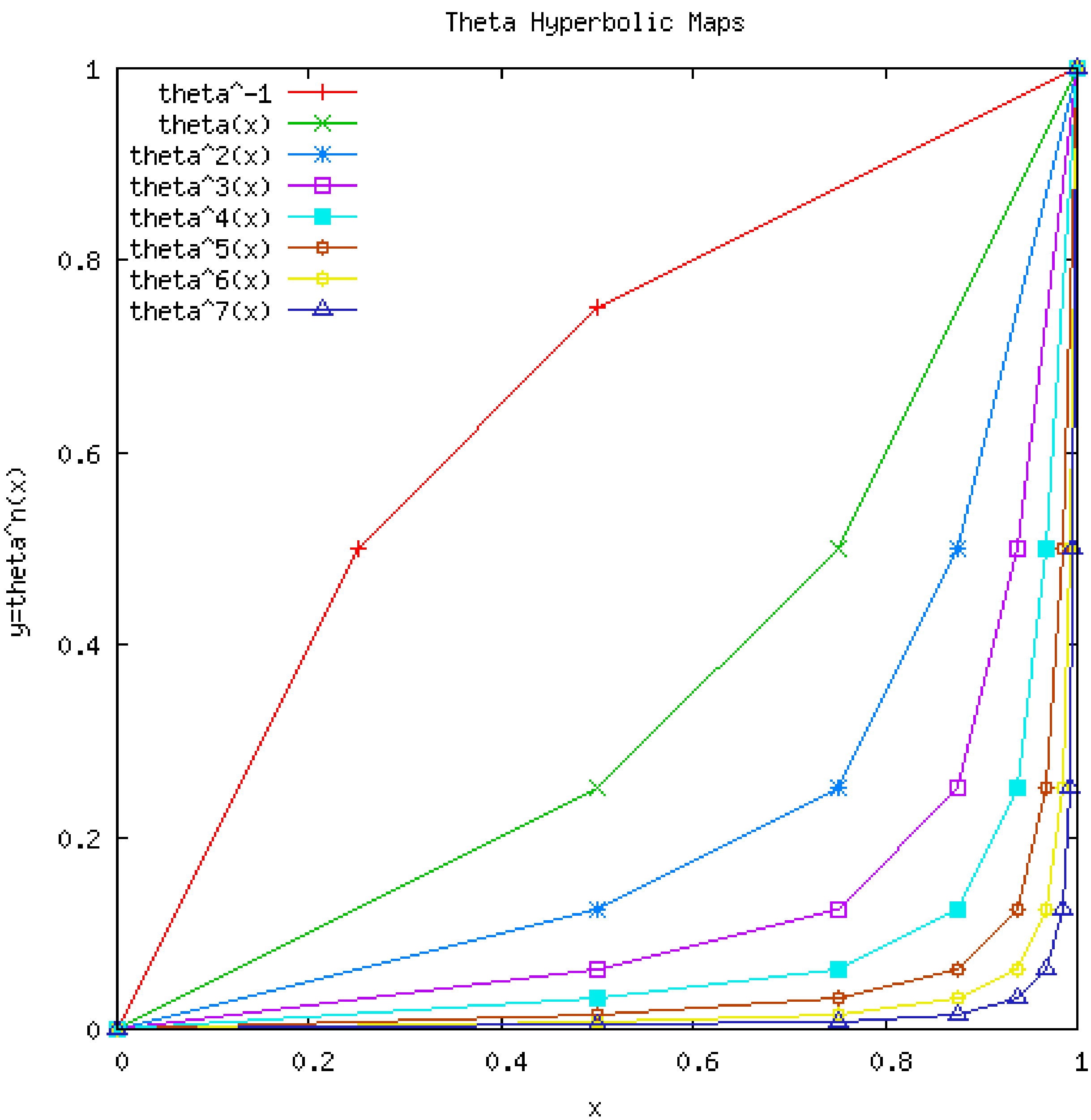
xxx

Looking at these, it becomes clear that one can insert hyperbolas wherever one wishes, as long as the resulting map is monotonically increasing.

And so we can see the negative curvature. we also have an explicit metric on this space. Note that this in turn induces a metric on the Stern-Brocot Tree. We can consider the path of 'geodesics' under this map, say of iterating θover and over. clearly these 'geodesics' separate.

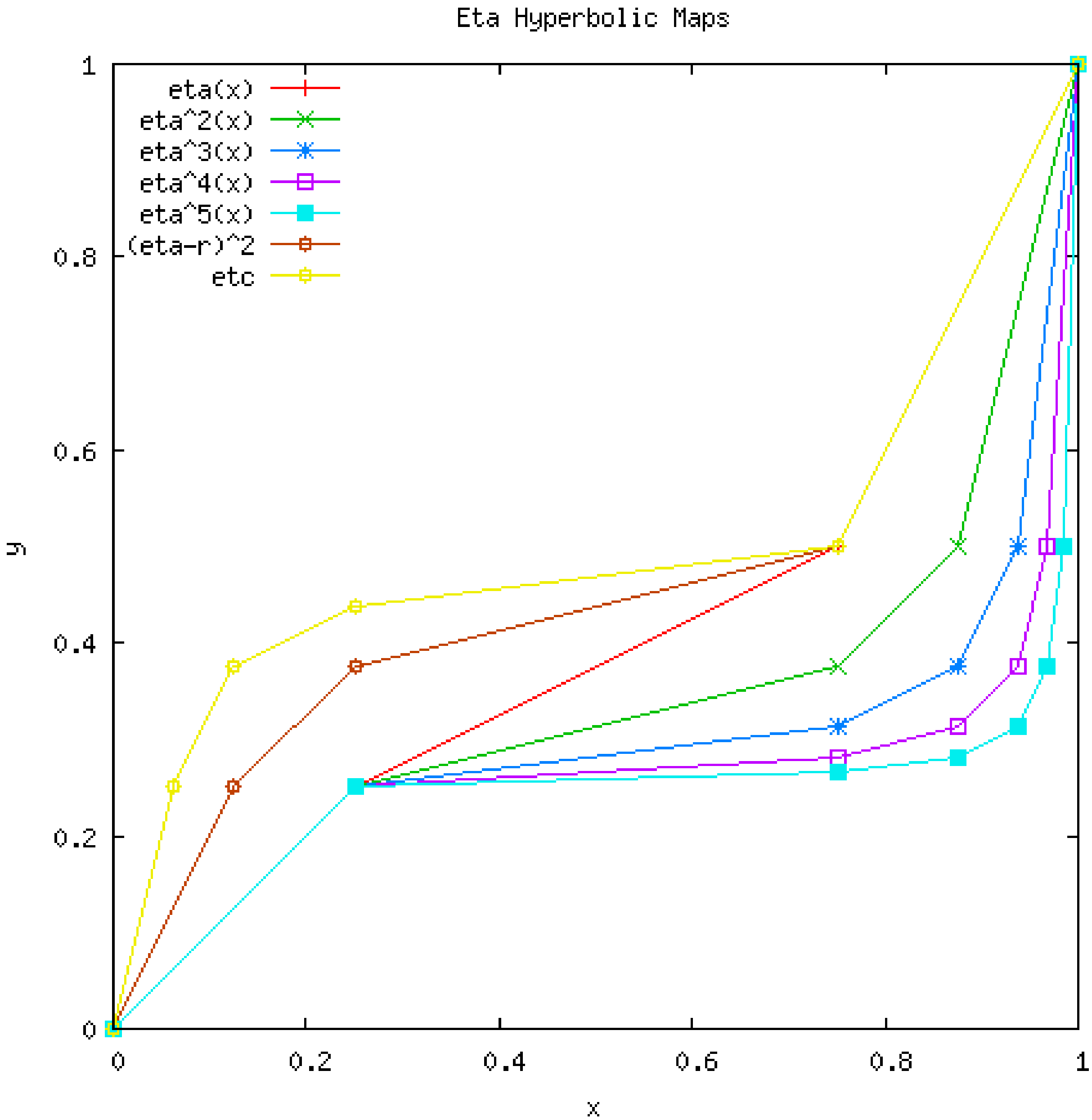
Note that this is essentially a model of two-dimensional space-time, where the Lorentz transformations are given by elements of the modular group.

Figure 1:



This figure shows $\theta^n(x)$ for $n = -1$ and positive n up to 7. Note that these maps are symmetric around the $y = 1 - x$ line, and that the endpoints of each segment lie on the hyperbola $xy = 2^{-n-2}$, with the exception of the enpoints at 0 and 1. Since each map contains more segments, each map becomes a better approximation to the hyperbola.

Figure 2: eta map



This figure shows two different hyperbolic sequences, one for the iterated eta, and also $r\eta r\eta$ and that sequence/

We of course are now begging to ask about 3+1 "spacetime" generated by the complex numbers, (xxx this is actually called the Picard group see Fricke and Klien, circa 1897.) which is generated by $SL(2, ZxZ)$ which is a subgroup of $GL(2, C)$. Studying how rotations work on this manifold would be interesting, as well as defining precisely its relation to minkowski spacetime.xxx move this to the todo-list. Higher dimensional manfolds seem to be generated by quaternions and the octonians. We also can't help but take a general pot-shot and exclaim 'of course quantum mechanics is chaotic: Hamiltonian evolution takes place on a fundamentally hyperbolic manifold, viz. the Minkowski spacetime'. The wave functions will of course be fractal. In some freaky way, this is also why we get quantization: we don't truly have the general symmetries, we have instead interval-mapping symmetries that are limited to a discrete (but infinite) set of intervals.

2.8 Symmetries Induced on $\mathbb{Q}$ and $\mathbb{R}$ (and $\mathbb{Q}_p$?)

Note that the correspondances presented here induce fractal isomorphisms of $\mathbb{Q}$ onto itself. This was already hinted at in the previous section, where we were trying to just plain-old count the rationals, and not getting what every middle and high-school teacher teaches students to expect. This symmetry would also imply a symmetry on the closure in reals $\mathbb{R}$. Exactly what it implies for the non-Archimedean closures $\mathbb{Q}_p$ is not yet clear to me. This symmetry is never made use of in classical analysis.

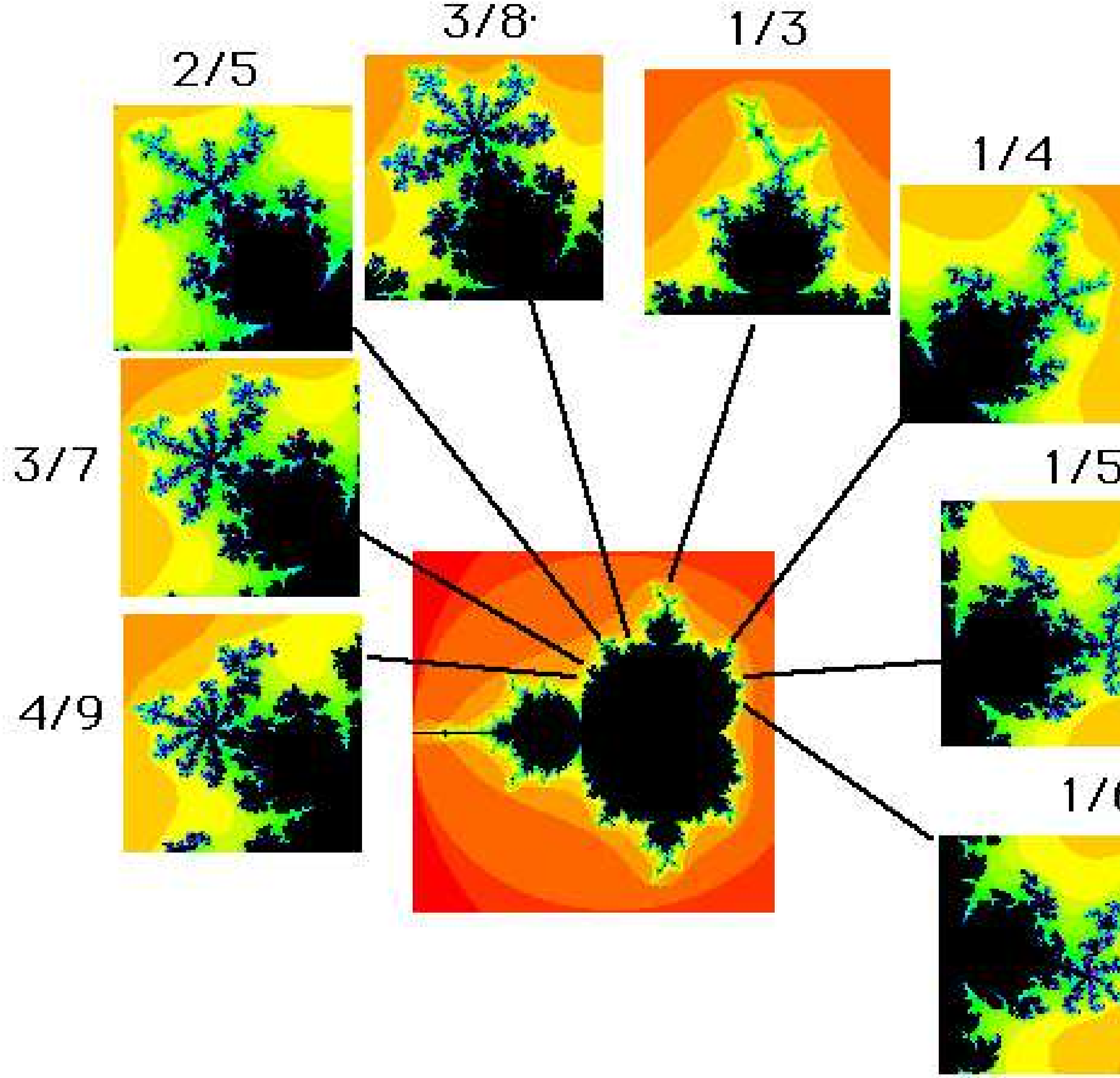
The field structure of the rationals in turn induces a field structure on a certain subroup of the modular group, specfically, the interval-mapping subgroup. Albeit this would be a highly fractal structure ... I have no clue of what this field structure implies for modular forms or goodies like the Wierstrass elliptic integrals, etc. This would be a great TBD with a whole-lotta partyin goin on.

2.9 Fundamental Domain

Blah Blah. The hyperbolic plane, the tesselation of the hyperbolic plane by hyperbolic triangles.

3 Symmetry of the Mandelbrot Set and the Circle Map

To be written. First we need to explain how to count with Farey numbers on the M-Set. Here is a picture stolen from Robert L. Devaney's website The Fractal Geometry of the Mandelbrot Set :



This mapping and method of counting on the buds is well-known in the literature and is explored in many ways.

Next we note that enumerating the group elements of the modular group as $\mathbb{Q} \times \mathbb{N}$ plays directly into the celebrated self-similarity of the Mandelbrot set.

What does this mean for the symmetries of the Mandelbrot set? it means that the modular group maps buds to buds. Note, however, that the $\mathbb{Q} \times \mathbb{N}$ parameterization of the modular group means that one can map one bud onto any another, but that the mapping has to be onto the whole bud, and not just a portion of it. Put another way, it is a mapping of intervals to intervals. However, we can't just map any-old-interval anywhere we want. We can map an interval to any starting point that we wish (this is the $\mathbb{Q}$ part of the symmetry, but after specifying the starting point, we have only a discrete set of intrvals we can map to: this is the $\mathbb{N}$ part of the enumeration. The buds appear discretely on the boundaries of the main cardiod.

3.1 Douady-Hubbard Rays

Douady Hubbard rays are a mapping of the unit circle to the boundary of the Mandelbrot Set. One of the properties is that a pair of rays pinches off each bud of the M-Set. Note that both of the rays occur at external angles that are also rationals. This now induces some additional symmetries on the rational numbers, since each bud can be labelled with a rational, and thus, for each rational, we associate the upper and lower rays. These two map are highly discontinuous. But that is OK, because the discontinuities pinch off the bud in the middle. We use the code-word substituion techniques of Doaudy-Hubbard to paste into the discontinuities. Since we are free to paste in any code-word desired, we have, in a certain sense, total and complete randomness, which is then propagated everywhere through the self-similarity.

Here is a bizarro philosophical thought: we have complete freedom to choose the codewords to be pasted into the period-doubling symmetry, in the sense of philosophical free-will: we can choose as we wish. The result of the application of our free will is multiplied through the modular group so that choice now appears everywhere among the rationals. But we know that the closure of rationals is the reals, and we know that the reals are "unknowable" in the Turing-computable sense. Thus, we conclde that the application of free-choice results in something unknowable. How strange.

Here's how code-word substitution works: start with the dyadic (base-2) expansion for $0 \leq x \leq 1$

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} = 0.a_1 a_2 a_3 ...$$

where the a_k can only take on the values 0 and 1. Then pick any two integers $h, l \in \mathbb{N}$ and a third integer m such that $h, l < 2^m$ is satisfied (alternately, we start by picking m first, then picking h, l to satisfy this). We then substitute the value of 0 in the dyadic expansion of x with l and the binary value of 1 in the expansion with h, padding each appropriately:

$$B_{h,l,m}(x) = \sum_{n=1}^{\infty} \frac{h a_n + l(1 - a_n)}{2^n 2^{nm}}$$

which we now see as a highly discontinuous function of x. To get a better idea of why its called "codeword substitution", lets work an example. Let $x = 0.10101010101..._2$ and pick $m = 4$ and $h = 7$ and $l = 4$. Note that, in binary, $h = 0111_2$ and $l = 0100_2$ and so we have

$$B_{7,3;4}(0.1010101..._2) = 0.0111\ 0100\ 0111\ 0100\ 0111\ ..._2$$

Lets now work backwards and get the question mark: thus

$$?([2, 1, 1, 1, ...]) = 0.1010101..._2$$

and

$$?([2, 3, 1, 1, 3, 3, 1, 1, 3, 3, 1, 1, ...]) = B_{7,3;4}(0.1010101..._2)$$

blah blah why are we doing this again?

3.2 Islands, Sinai's Tongues

blah blah write this section.

See the pictures of phase-locking in the Circle map at the URL Circle Map `http://www.linas.org/art-gallery/basic/basic.html` and notice the similarities to the gaps pictured at Farey Room Pictures `http://www.linas.org/art-gallery/farey/fthumb.html` Yada Yada. The mode-locked regions are called Sinai's toungues. This implies that the Modular Group is the symmetry group of the phase locked-loop (the circle map) as well as the Mandelbrot Set.

Notice 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 Form symmetry?

What about Misiurewicz points? Misiurewicz points on the M-set are the pre-periodic points. As such we expect to be able to label them in some way with xxx ???

3.3 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,... I'm not sure I think this is a manifestation of the fact that the modular group can be written as the free product of the two and three cycles C_2 and C_3.

3.4 Iterating on the Hyperbolic Plane

Of course, now that we know about the modular group and the hyperbolic plane, we state that we really want to study the iteration of one-dimensional maps like the logistic equation or the tent map by studying how it bounces around on the fundamental domain. That is, each time we iterate a 1-D map, we are bounced to somewhere else on the unit interval. This point on the unit interval is included in an enumerable set of line segments, which can be maped back to the whole unit interval via the modular group. Thus, a single point is a representative element of the equivalence class of all the points that the modular group maps it to. We want to study how the logistic equation acts on these equivalence classes, instead of asking how it acts on single points.

Blah Blah Blah. Define what equivalence class means when a group acts on a set. viz define the quotient space.

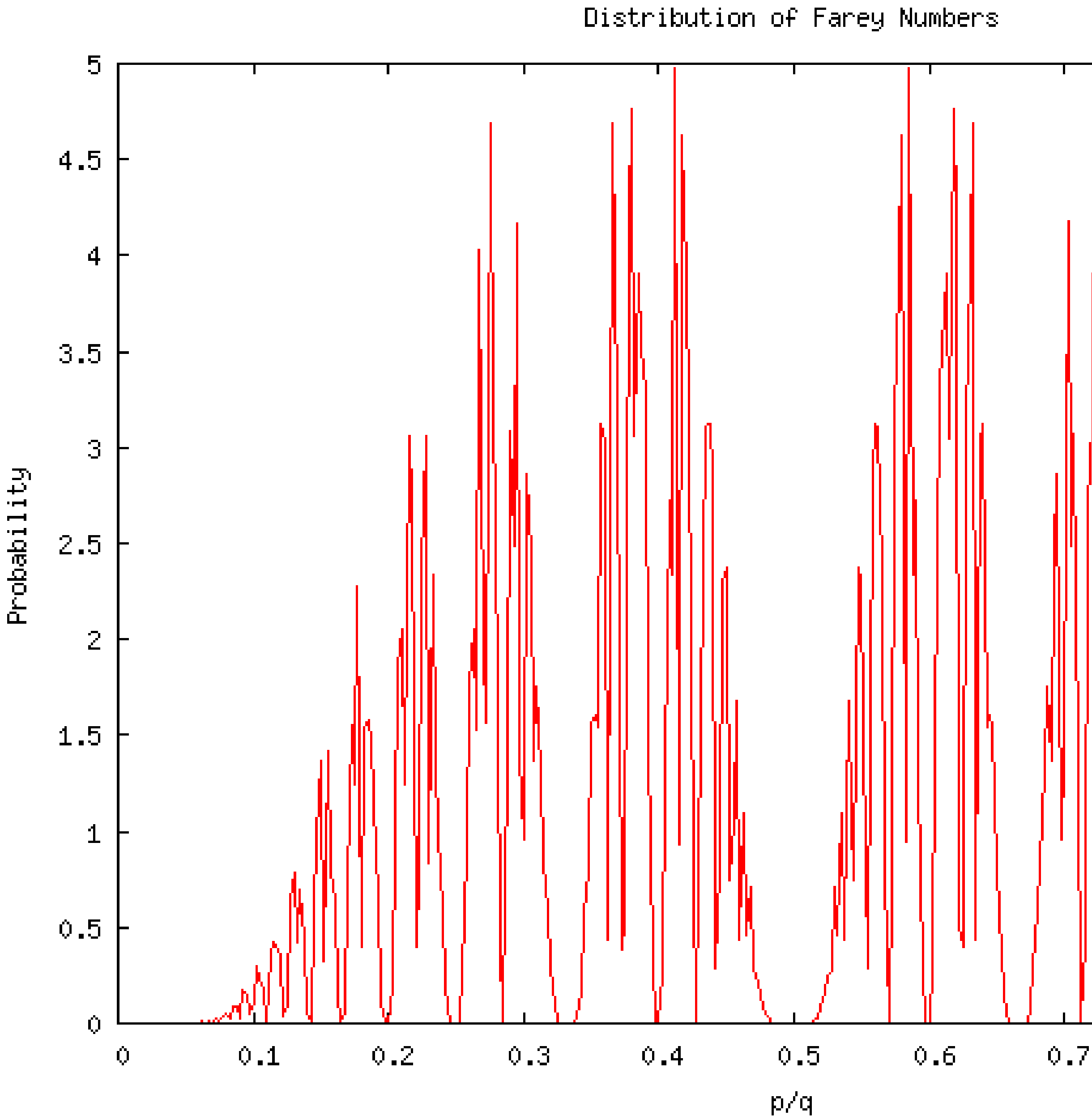
This is essentially how we get related to Julia sets. After we iterate a map like the logistic map once, we then ask what other points are in the same equivalence class, under the action of the modular group. That is, we now studay iteration in the quotient space, where the modular group provides the quotient. This quotient map is just the map of all hyperbolic triangles back to the fundamental domain. And so we really study iteration in this quotient space. Blah blah. Provide pretty pictures. We now ask what do the points on the boundary of the fundamental domain correspond to, blah blah, these are the Misiurewicz points, which is why they relate to Julia sets; they essentially paste together the pieces.

4 Distributions

Blah blah blah. Its all interconnected, isn't it?

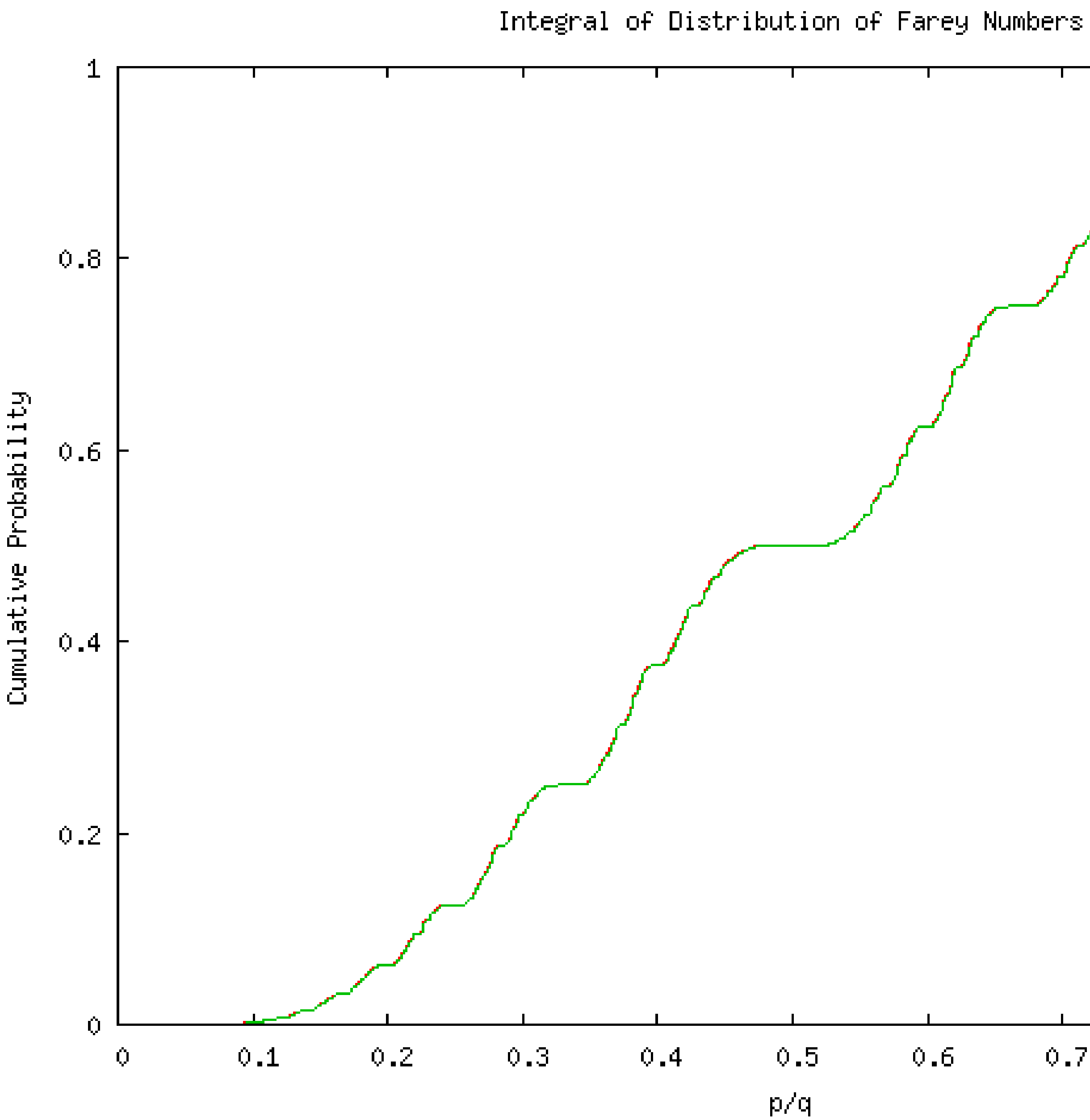
4.1 Distribution of Rationals in the Farey Tree

One way to enumerate all of the rationals is by placing them on the Farey Tree (or the Stern-Brocot Tree). One can prove that the Farey Tree enumerates all of the rationals in the unit interval exactly once (Proof at Cut-the-Knot `http://www.cut-the-knot.org/blue/Stern.shtml`). However, the distribution of the rationals on the Farey Tree is highly non-uniform, if we walk that tree breadth-first. The graph below shows that distribution.



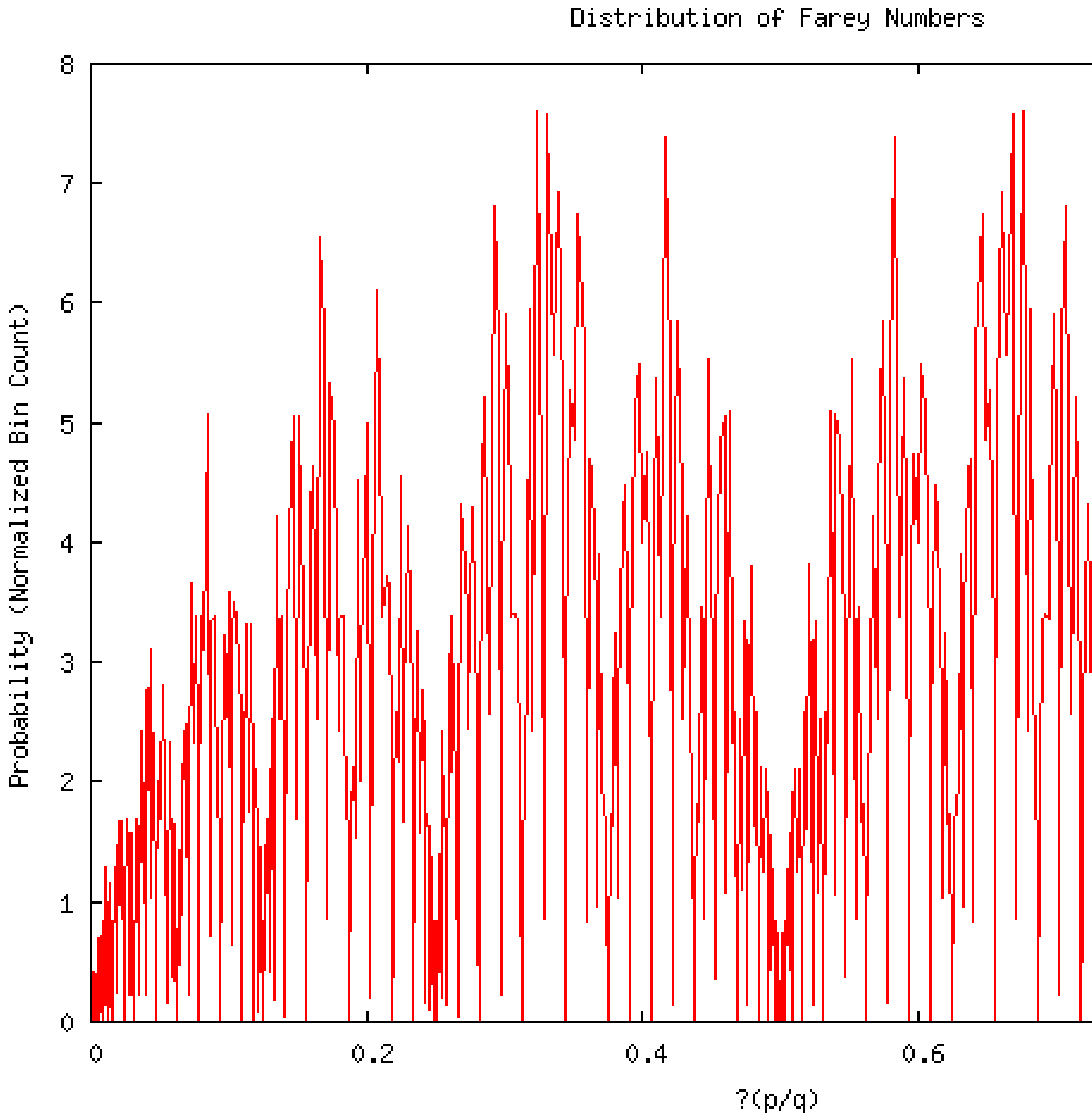
Superficially, the shape of this curve seems independent of the bining; although, upon very close inspection, one can see that is not quite true. In particular, one can deduce that in the limit of an 'infinite' number of bins, that this 'curve' is equal to exactly zero at every rational p/q, and infinite at every irrational. But no matter; for a finite number of bins, we can proceed.

This shape is remarkably suggestive. It appears to be the derivative $?'(x) = d?(x)/dx$ of the continued fraction mapping, although using classical analytic techniques, this derivative is not definable. Indeed, performing a simple numeric integral of the above, and comparing it to the exact expression shows a perfect match, as witnessed in the graph below.



The continued fraction mapping is infinitely differentiable on every rational, and all derivatives vanish at every rational; yet the mapping is strictly montonically increasing. Just examing the graph visually suggests that the derivatives must be non-zero "somewhere", although its hard to define how to value these derivatives. Yet, visually comparing the Farey mapping to the distribution of the Farey Fractions, it seems that the distribution of Farey's might be a good estimate of that first derivative. Clearly, classical analytic techniques are sorely lacking and inapplicable; we need to find a set of tools where we can talk about and manipulate the so-called 'first derivative' in a meaningful way.

Below is another curious picture, where we've relabelled the x axis by passing it through $?(x)$; one could say that it is a picture of the Jacobian $(?' \circ ?^{-1})(x)$



This function appears to vanish only at the dyadics. Note the superficial resemblance of the profile to the Takagi curve. By visually comparing this function and the Takagi curve (also called the *blancmange curve*), one can intuitively sense that there must be some tangent-space-to-Exp-space type relation between the two. The trick is to write down this relation.

This is really a very curious circumstance. The above graphs show only the distribution of a finite number of points (they were, after all, computer generated). A finite set of points on the real number line is zero; yet we are using it to induce a measure that can be used for integration. How is this? The statistical properties of a finite sampling provide us with information about the limit. Its as if we were working with an anti-Cantor-set construction: we start with a dust, and average over intervals to discover analytic properties.

4.2 The Takagi-Landsberg Curve

The Takagi-Landsberg curve was first explored by Teiji Tagaki, circa 1903 as an example of a continuous-everywhere, differentiable-nowhere curve [Man88]. 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 we should diagram the standard contruction here). Equivalently, one assigns geometrically smaller displacements to each level of the dyadic tree; that is, one picks some $w < 1$, and assigns to each dyadic number $x = p/2^n$ the value

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

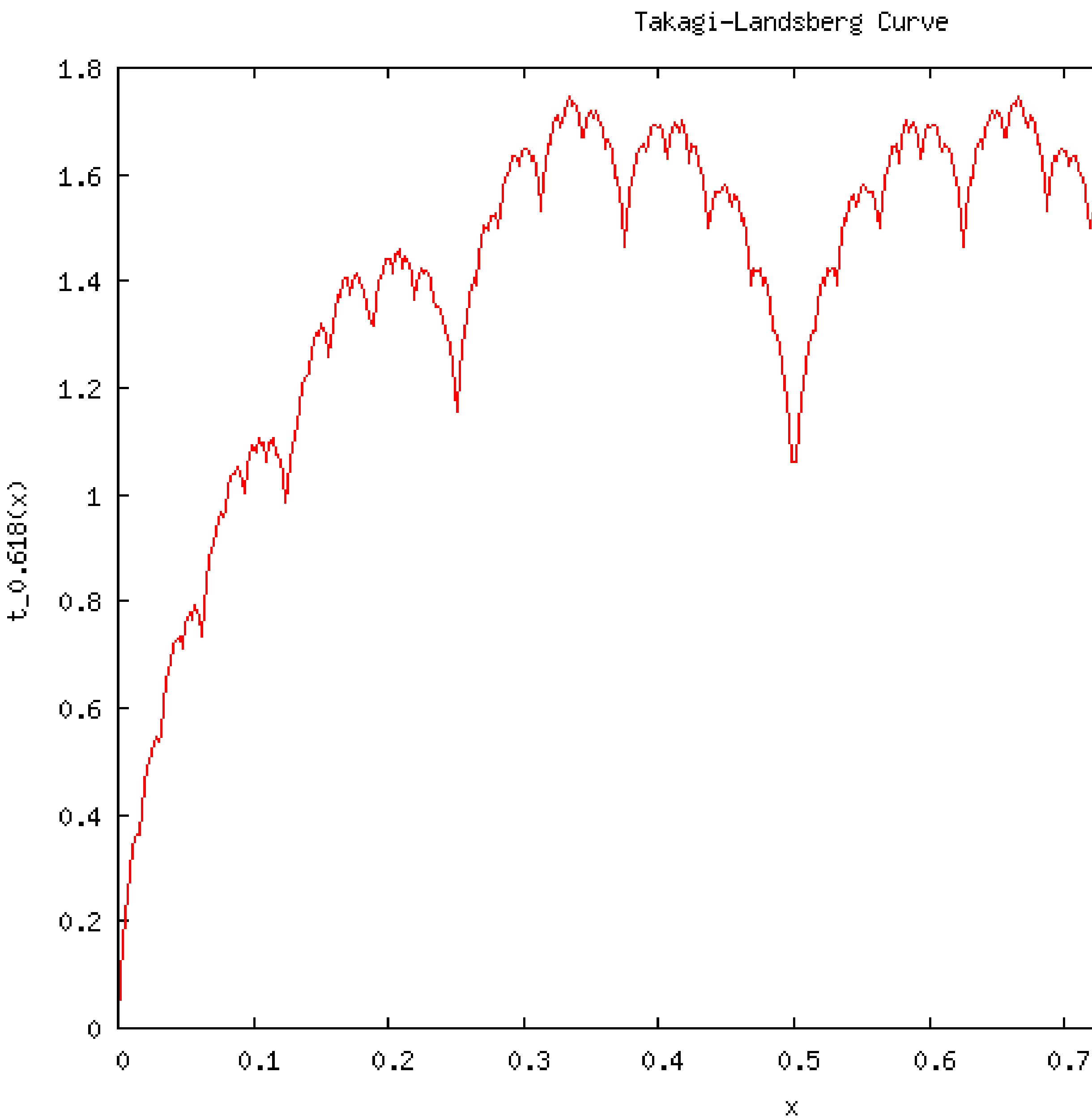
One extends this curve to all reals by 'linear interpolation' between the dyadic values, as

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

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

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

Note that insofar as the triangle curve is differentiable, we could say that the Takagi-Landsberg curve is differentiable (along x) as well, in that it has both leftward-going and rightward-going derivatives, although these derivatives differ at each dyadic point. We can choose to define the derivative dt_w/dx by passing the d/dx past the usmmation above, to act on the triangle-wave, to give a square wave. Note that $t_{1/4}(x) = 4x(1-x)$ is a parabola (the mid-point subdivision construction of a parabola is attributed to Archimedes[Man88]). A prototypical Takagi-Landsberg curve is pictured below:



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 $f(0) = f(1)$. Discontinuous functions f yeild 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 below.

Of the many possible different choices for $f(x)$, the traingle 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 in k times gives $\tau^k(x) = \tau(2^k x)$. Thus, we may write

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

In passing, we should also note another curious sum

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

Also in passing, we note that the iterated tent map is a shift-state of the Bernoulli Operator.

4.3 Symmetries of the Takagi-Landsberg Curve

In this section, we will show that the Takagi-Landsberg curve transforms as a three-dimensional representation of the modular group. This provides an example where the action of the group can be precisely and completely specified. Consider first the generator that is the interval contraction $g : [0,1] \to [0,1/2]$. It acts on the Takagi map as

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

Iterating on this generator gives

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

Next, we notice that the action of g on the expression $a + bx + ct_w(x)$ is linear and is also closed, thus forming a vector space. That is, we have discovered that the Takagi Curve transforms under the three-dimensional represention of the modular group. To be explicit, we write a general element as the dot-product of a column of real numbers times a row of basis vectors

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

Then, we find that g and r have the representations

$$g = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & w \end{pmatrix} \quad \text{and} \quad r = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

by which we mean, for example,

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

and also that

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

which we recognize as the dyadic representation g_D from previous sections. The Takagi Curve is then seen to commute with the generator $(gt_w)(x) = (t_w g)(x)$ in this representation. By contrast, the curve has mirror symmetry, and thus kills the reflection operator: $(t_w r)(x) = t_w(x) \neq (rt_w)(x) = 1 - t_w(x)$.

Thus, in the above, we have demonstrated the explict symmetry relationship of the Takagi Curve, and furthermore, have indicated that it, together with 1 and x, are the basis vectors of a three-dimensional representation of the modular group. The matrix representations of g and r make it very easy to compute how the curve transforms under any given element of the modular group.

Theorm: Every real value of w generates a unique, inequivalent representation of the modular group.

Proof: If two representations were equivalent, then we could find a matrix U such that $g_v = U^{-1} g_w U$ for $v \neq w$ and $r = U^{-1} r U$. We 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 $d = k = 0$. Next, equating $Ug_v = g_w U$ one finds that $l = m = 0$. This means that $detU = 0$ and thus U is not invertible: there is no equivalency transformation between two different representations. *QED.*

xxxxx more junk below, needs fixing.

We can now take an explicit hyperbolic symmetry, and apply it directly to the Takagi curve. Start with the $\theta(x)$ previously given in the hyperbolic section up above. Evaluate $t_w(\theta(x))$ to get ...

We really need a proof that *any* hyperbolic curve generated by $SL(2, Z)$ does the right thing here. we can do this recursively by noting that each hyperbolic is a list of straight line segments, so that at each point, we are performing a linear rescaling.

4.4 Analytic Continuation of the Takagi Curve

For any fixed x, the Takagi Curve is analytic in w, having a simple pole at $w = \exp(i\theta)$. We want to do two things here: first, give θ as a function of x, second, analytically continue to regions $|w| > 1$.

4.5 The Frobenius-Perron Operator

This section provides a basic review of the Frobenius-Perron operator and its use in the description of fractals and chaotic iterated maps. No results are presented here; rather the goal is to provide the background for later sections.

The Frobenius-Perron operator of a function provides a tool for studying the dynamics of the iteration of that function. If one only studies how a point value jumps around during iteration, one gets a very good sense of the point dynamics but no sense of how iteration acts on non-point sets. If the iterated function is applied on a continuous, possibly even smooth density, then one wants to know how that smooth density evolves over repeated iteration.

If we consider a smooth density $\rho(x)$ as a set of values on a collection of points, we can take each point and iterate it to find its new location, and then assign the old value to the new location. Of course, after iteration, several points may end up at the same location, at which point we need to add thier values together. Lets write the new density as $\rho_1(x)$, with the subscript 1 denoting we've iterated once. We can express this idea of iterating the underlying points, and then assigning thier old values to new locations as

$$\rho_1(x) = \int dy\, \delta\left(x - g(y)\right)\ \rho(y)$$

where $g(x)$ is the iterated function. To get $\rho_n(x)$, one simply repeats the proceedure n times. In more abstract notation, one writes

$$\left[U_g \rho\right](x) = \rho_1(x)$$

to denote this time evolution. The notation here emphasizes that $U_g :\ f \mapsto U_g f$ is an operator that maps functions to functions: written formally, we have $U_g : \mathcal{F} \to \mathcal{F}$ where $\mathcal{F} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ is the set of all functions. In analyzing U_g, we will often be interested in how it acts on the subset of square-integrable functions, or possibly just C^∞ functions or polynomials or the like. Repeated iteration just gives the time-evolution of the the density; that is,

$$U_g^n \rho \equiv \underbrace{U_g \circ U_g \circ \ldots \circ U_g}_{n \text{ times}} \circ \rho = \rho_n$$

where iteration is just ordinary operator multiplication.

To understand U_g, one typically tries to understand its spectrum, that is, its eigenvalues and eigenfunctions. In most cases, one finds that U_g is contractive in that it has one eigenvalue equal to one and all the other eigenvalues are real and smaller than one. However, one must be terribly careful here, as there are landmines strewn about: the actual spectrum, and the nature of the eigenvalues, depends very much on the function space choosen. Typically, when acting on polynomials, one gets discrete, real eigenvalues for U_g. When acting on square-integrable functions, one seems to usually get a continuous set of complex-valued eigenvalues. This is because one can often find shift-states ψ_n such that

$U_g \psi_n = \psi_{n-1}$, in which case one can construct eigenfunctions $\phi(z) = \sum_n z^n \psi_n$ whose complex eigenvalues z form the unit disk. It is often considered to be a mistake to try to analyse U_g acting on a finite grid of discrete points, such as one might try on a computer: it is all to easy to turn this into an excercise of analysing the permutation group on a set of k elements, of which any student knows that the eigenvalues are the k'th roots of unity.

Since U_g is a linear operator, it induces a homomorphism in its mapping, and so one should study its kernel $Ker\ U_g = \{f \mid U_g f = 0\}$ to gain insights into its symmetry as well as to express more correctly the quotient space. Insofar as the iterated map might represent a dynamical system, one knows that symmetries lead to conserved currents, via Noether's theorm, and sometimes to topologically-conserved (quantum) numbers, winding numbers or other invarients.

Finally, we note that since U_g looks like a time-evolution operator, we are tempted to write

$$U_g^t = \exp\ tH_g$$

for some other operator H_g. Since U is in general not unitary, H is not (anti-)Hermitian. As before, describing the eigenvalues and eigenfunctions of H is a useful excercise. Also, any group of symmetries on U should express themselves as an algebra on H and these might provide an alternate path for exploring and describing the fractal in question.

In practice, when one is given an iterated map $g(x)$, one computes the Frobenius-Perron operator as

$$\left[U_g \rho\right](x) = \sum_{x' : x = g(x')} \frac{\rho(x')}{|dg(x')/dx'|}$$

which provides an expression for U_g acting on a general function ρ. If one is interested in U acting on polynomial or C^∞ functions, then one immediately writes the Taylor (or McLauren) series

$$\rho(x) = \sum_{n=0}^{\infty} \frac{\rho^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} a_n x^n$$

and substitutes this in to get the matrix form of U:

$$\left[U \rho\right](x) = \sum_{m=0}^{\infty} b_m x^m = \sum_{m=0}^{\infty} x^m \sum_{n=0}^{\infty} U_{mn} a_n$$

Equating each power of x^m we get

$$\left. \frac{1}{m!} \frac{d^m \left[U \rho\right](x)}{dx^m} \right|_{x=0} = \sum_{n=0}^{\infty} U_{mn} \left. \frac{1}{n!} \frac{d^n \rho(x)}{dx^n} \right|_{x=0}$$

as the matrix equation for the transformation of polynomials, expressed in classical notation. Completely equivalently, we can use the Dirac bra-ket notation, writing down the operator in space coordinates:

$$\delta\left(g(x)-y\right) = U_g(x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \langle x|m\rangle \langle m|U_g|n\rangle \langle n|y\rangle$$

where $U_{mn} = \langle m|U|n\rangle$ and $\langle x|m\rangle = x^m/m!$ and $\langle n|y\rangle = \delta^{(n)}(y)/n!$, the latter being the n'th derivative of the Dirac delta function. Then, integrating over y, we regain the previous expressions for the operator in Hilbert space:

$$\begin{aligned}
\left[U_g\rho\right](x) &= \int dy \quad U_g(x,y)\rho(y) \\
&= \int dy\, \delta\left(g(x)-y\right)\rho(y) \\
&= \sum_{m,n=0}^{\infty} \frac{x^m}{m!}\, U_{mn} \int dy\, \frac{\delta^{(n)}(y)}{n!}\rho(y) \\
&= \sum_{m,n=0}^{\infty} \frac{x^m}{m!}\, U_{mn}\, \frac{1}{n!}\, \left.\frac{d^n\rho(y)}{dy^n}\right|_{y=0} \\
&= \sum_{m,n=0}^{\infty} \frac{x^m}{m!}\, U_{mn}\, \frac{\rho^{(n)}(0)}{n!}
\end{aligned}$$

The above sets of expressions, although in a certain sense are 'equivalent', as we did use the equals sign, do in fact do a great deal of violence because they hide or incorrectly equate the function spaces on which the operator U_g acts. That is, whenever $\rho(x)$ is not differentiable or is otherwise singular, the expansion in derivatives is not justified; by contrast, U_g acting in the space of square-integrable functions may have non-differentiable eigenfunctions (which turn out to be, of course, fractals, which is the whole point of this excercise!)

If one is very lucky, one finds that U_{mn} is upper-triangular, in which case it can be solved immediately for its eigenfunctions, and its eigenvalues already lie on the diagonal. We will find that we get lucky in this way for the Bernoulli operator, and for the "singular sawtooth" operator, but not for the Gauss-Kuzmin-Wirsing operator. Of course, it is known that a complete solution of the GKW should lead directly to a proof of the Riemann Hypothesis, so getting lucky would be truly lucky indeed.

4.6 The Frobenius-Perron Operator of the Bernoulli Map

The Bernoulli map is an exactly solvable example of deterministic chaos. Its theory has been widely explored[Dri99], and we recap here only insofar as it gives a simple example of some of the more subtle upcoming issues.

The Bernoulli map is given by

$$b(x) = 2x - \lfloor 2x \rfloor$$

and can be thought of as poping the leading digit off of the binary expansion of x. This map has a positive Lyapunov exponent and is highly chaotic, as, in a certain sense, one can say that the digits of the binary expansion of some 'arbitrary' number are 'unpredicatable', and that the orbits of two close-by numbers will eventually become 'uncorellated' (after suitably defining what we mean by 'arbitrary' and 'unpredictable').

The Frobenius-Perron operator of the Bernoulli map is given by

$$\left[U_B f\right](x) = \frac{1}{2}\left[f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right)\right]$$

Note that the the iterated tent map behaves sort-of like a shift state, in that

$$\left[U_B \tau^k\right](x) = \tau^{k-1}(x)$$

although it does not terminate properly for a shift state:

$$\left[U_B \tau\right](x) = \frac{1}{2}$$

(a true shift state would vanish on the final iteration). We see that the Takagi curve transforms as

$$\left[U_B t_w\right](x) = 1 - \frac{w}{2} + w t_w(x)$$

under the Bernoullli operator.

Blah blah blah. How much more here do we want to say? Well, Provide the matrix elements. Provide the discrete eigenfunctions, which are amazingly also named after Bernoulli: the Bernoulli polynomials. What is the kernel of U_B? It is the set of functions that have odd symmetry. This implies that 'half' of all square-integrable functions are in the kernel. This is a huge space. The quotient space of the implied isomorphism thus has the Bernoulli polynomials as the representative elements. This is I think the correct way to relate coordinate space to the Hilbert space, is by means of the quotient space generated by the kernel of the time-evolution operator. This provides the isomorphisms, and any other square-integrable function is thus expandable as a linear combination of the polynomial eigenfunctions plus the kernel. Talk about H and its spectrum. Symmetry symmetry talk about the symmetry group of the kernel.

The continuous-spectrum eigenfunctions of the Bernoulli map are of course a generalizastion of the Takagi-Landsberg curve to complex values, and are of course the famous and beautiful Levy Dragons (reference the Paul Levy 1938 paper here) (actually Levy was trying to generalize the Koch curves). These eigenfunctions are

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

which have eigenvalue z: that is $[U_B \phi_{z,l}](x) = z\phi_{z,l}(x)$. These are quite beautiful fractals provide a picture here. Symmetry is the name of the game here.

Note that when we go to look at the tesselatation of the hyperbolic plane by the hyperbolic triangle under the symmetry of the modular group, I think we get the tesselation of the plane with Levy's Dragons, which I think is the point that Paul Levy was trying to make.

This also means that Bournoulli map also makes for a good toy example of quantum chaos. In quantum chaos, the eigenfunctions must form a complete set in the sense that for any given point in space, there must exist at least one eigenfunction that is non-vanishing on that point. In other words, the eigenfunctions of a quantum-chaotic system must tesselate the space in which they live. The toy example of this kind of tesselation is the Levy Dragon.

4.7 The Singular Sawtooth

The singular sawtooth is given by the dyadic-space version of the frac operator, and is thus a model of the Gauss-Kuzmin-Wirsing operator. It can be solved exactly; unfortunately, while one might think that there is a similarity transform to take it back to GKW, it turns out this similarity transform is dastardly singular, being just the Jacobian of the Minkowski question mark $(?' \circ ?^{-1})(x)$ Of course, we knew there was trouble because the eigenvalues

$$\lambda_k = \sum_{n=1}^{\infty} \frac{1}{n^k (n+1)^k}$$

are not the eigenvalues of GKW. We can see (trivially) that the ratio of eigenvalues settles down to $\lambda_k / \lambda_{k+1} = 2$ whereas for the GKW the ratio is 2.65 (which is what, I'm not sure, is close to Khinchin's constant 2.685 xxxx) confusing.... Feigenbaum's constant as the ratio of period doubling is $\delta = 4.6692...$ which is 2+2.6692... so which is it, exactly?

We'd like, among other things, to present the shift operator, so that can write down the analogue of the Levy-Dragon solutions for this operator, and possibly take a guess at those of the GKW.

blah blah blah provide the details of the solution here.

4.8 Other Trees of Rationals

Recall that we were able to construct other trees that are isomorphic to the Stern-Brocot tree by passing the values in the Stern-Brocot Tree through a ratio of polynomials, thus mapping rationals to rationals. Monotonically-increasing functions will always perserve monotonicity of the question mark. Other mappings, such as the cubic $x^3 - x$, are in general not globally monotonic, but they are always locally monotonic. In other words, $?((2x-1)^3 - x + 1)$ is montonic between the same regions that $(2x-1)^3 - x + 1$ is, and thus it does not yeild a fundamentally new function. Thus, these ploynomial-based mappings of rationals into rationals do not generate another self-similar map, they are distorted versions of the Minkowski question mark.

Here's a mapping that is continuous and is also fractally non-monotonic (and is thus self-similar): its $x-?(x)$ which looks sine-like and we want to exhibit the cosine-like so that we can build up fractally-periodic basis functions.

4.9 Cantor Dusts

Speaking of Cantor sets, this begs a question. Can we find functions that are continuous and monotinically increasing, yet all of whose derivatives are identically zero on a Cantor dust? I think we can. This section is also to be written. Sigh.

5 Permutations

Stuff about permutation groups.

5.1 Hilbert Lattice and Permuation Groups

Note that continued fractions have a representation as a lattice in the infinitie-dimensional Hilbert space $\mathbb{N}^\infty = \mathbb{N} \times \mathbb{N} \times ...$ where $\mathbb{N} = \mathbb{Z}^+$ the set of positive integers. Note that this implies that each point on this lattice can be given a unique value using the continued fraction mapping, and that each value appears only once on the lattice. In other words, the continued fraction brackets are an operator that maps this Hilbert space to the real numbers, i.e. $[] : \mathbb{N}^\infty \rightarrow \mathbb{R}$. What makes the [] operator so interesting is that it is 1-1 and onto, and is thus invertible, at least on the irrationals, if we brush aside a little problem with the rationals. Note that in this mapping, the number 0 is a kind-of "essential singularity" at ∞, since the continued fraction expansion for 0 is $0 = [\infty, a_2, a_3, ...]$ for arbitrary integers a_k. Indeed, all rationals suffer in this way: a finite contined fraction is just one where ∞ appears somewhere in the expansion, i.e. $[a_1, a_2, ..., a_N] = [a_1, a_2, ..., a_N, \infty, a_{N+2}, ...]$. We can work around this difficulty by taking the kernel of [] and defining a quotient space $\mathbb{N}^\infty / \ker[]$, and insisting that the operator [] is 1-1 and onto in this quotient space. Homework: prove that this quotient space is validly defined and has the needed vector-space properties. In the subsequent discussion, we'll intentionally confuse $\mathbb{N}^\infty$ with this quotient space, wherever its convenient. In the following, we also make a convenient little confusion between the continued fraction map defined on the unit interval, and the map defined for non-negative reals: viz. $[a_0; a_1, a_2, ...] \equiv a_0 + [a_1, a_2, ...]$ where we use the semicolon to distinguish this zero'th element.

Note that this lattice is a subspace of the Hilbert space of "generalized functions", in that any map $\mu : \mathbb{N}^\infty \to \mathbb{R}$ induces a map $\tilde{\mu} : \mathbb{R} \to \mathbb{R}$, through the isomorphic operator []. The normal definition of a generalized function is that its a map $\tilde{f} : \mathcal{F} \to \mathbb{R}$ where $\mathcal{F}$ is the space of ordinary functions $\{f \mid f : \mathbb{R} \to \mathbb{R}\}$. The canonical example is the Dirac delta oprator $\delta(x)$ which is zero everywhere, except at $x = 0$ where it is infinite.

It seems intuitive that all other mappings from $\mathbb{N}$ to $\mathbb{Z}$ that satisfy this invertability property are obtainable through the action of the permutation group, although one should always be careful with such statements.. The permutation group elements are the shift operator $h(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$ which acts as $h : [a_1, a_2, a_3, ...] \mapsto [a_2, a_3, ...]$ and the permuation operator $S_{p,q} : [a_1, ..., a_p, ..., a_q, ...] \mapsto [a_1, ..., a_q, ...a_p, ...]$. Homework: give the analytic form for S. Note that $S_{p,q}(x)$ is discontinuous, but I think only on a countable set of points, and thus can be used to form well-defined integrals using classic techniques.

Note that more generally, we can now define $h : SL(2, \mathbb{Z}) \to SL(2, \mathbb{Z})$ as the operator that lops off the first term of teh expansion. This has a kernel that contains the group element $g^n rg$ and so we ask about the group structure that results from this.

In particular, let us look at the family of analytic functions blah blah blah that is produced by invoking the permutation group on blah blah. What are the symmetries.

6 Conclusions

Apologies for the format of this paper; its like the Picard Theorm suggests: this paper will touch on all possible topics an infinite number of times as it approaches the essential singularity of being completed.

Blah Blah Blah, The modular group is the symmetry group of the hyperbolic plane, and trajectories on hyperbolic surfaces are known the have positive Lyapunov exponents, so once again, we have another curious tie into chaotic behvaiour.

Note that this also means that all of the results from modular forms now have some sort of interpretation in the fractal world. The program is now how to map those results into statements about classical chaotic dynamics and more importantly, for quantum chaos. Do statements about Hamiltonian dynamics have corresponding statements in terms of modular forms? If so, what are they? What about the converse?

7 Open Questions

Some unanswered questions of varing depth and complexity.

7.0.1 Multiplicity (Degeneracy) of Rationals

If we enumerate the rationals p/q for all values of $p, q \in [1..N]$, there will be many duplicates after reducing by eliminating common factors. In the limit of $N \to \infty$, what is the multiplicity (degeneracy) of any given rational? What is the group of symmetries of the resulting fractal?

7.0.2 Sagher Enumeration

Sagher enumerates rationals m/n with the integer $m^2n^2/q_1..q_k$ where q_i are the prime factors of n. Make a graph of this distribution.

7.0.3 Simplicity

The simplicity $w(p/q)$of a farey number is given by $w(p/q) = 1/pq$. The weighted simplicity is $W(p/q) = w(p/q)2^{-N(p/q)-1}$ where $N(p/q)$ is the row number on the Farey Tree. Make a graph of this function. Relate it to the rest. Note that the sum of $w(p/q)$ on one row of the farey sums to 1, and that thus, the sum of $W(p/q)$over all rationals equals 1. See Cut the Know for the correct attribution for this concept.

7.0.4 The Gauss-Kuz'min Wirsing Distribution

The Gauss-Kuzmin-Wirsing (GKW) distribution gives likelyhood of the occurance of any given partial sum in a continued fraction expansion. If we were instead to flatten that distribution, and work backwards, this would imply the existance of a measure/distribution on the rationals. What is that distribution? If we were to replace the GKW distribution by some arbitrary (monotonically-creasing, continuous) mapping, what would eb corresponding measure be that is induced? Can we analyse the these measures in terms of the eigenfunctions of the GKW?

7.0.5 Relation to Cryptography

Cryptographic hashes are also known to distribute points randomly on the unit square; but they do so only on a fixed (albeit very large) lattice. What are the relationships between the randomness of the gap and crypto-type systems? For example, we could study the behaviour of the gap on all rationals which have the same large prime as the denominator. Is there a relationship to elliptic curves (which are known to be related to continued fractions in other ways, e.g through modular forms)? Can we say that the uniform distribution of gaps is somehow the 'fundamental theorm of cryptography', implying that there are always unbreakable codes because the gaps are dense on the unit square?

7.0.6 Algebraic Pellian-like Equations

Given arbitrary real x, y, one can find a sequence of rationals that converge on x and whose gaps converge on y. Can this sequence be expressed as approximants to some algebraic equation, somewhat in analogy to the Pellian equations? What would that series of equations be?

References

[asdf] Here is a very similarly titled paper with a very different subject matter: Continued Fractions and Chaos `http://www.cecm.sfu.ca/organics/papers/corless/confrac/html/confrac.html` by Robert M. Corless

[Dri99] Dean Driebe Fully Chaotic Maps and Broken Time Symmetry Kluwer Academic

[Man88] Benoit Mandelbrot, in *The Science of Fractal Images, ed. Heinz-Otto Peitgen, Dietmar Saupe,* (Springer-Verlag, 1988) p. 246

[deR57] Georges de Rham, On Some Curves Defined by Functional Equations (1957), reprinted in Classics on Fractals, ed. Gerald A. Edgar, (Addison-Wesley, 1993) pp. 285-298