On the Beta Transformation

Linas Vepstas

December 2017

linasvepstas@gmail.com

Abstract

The beta transformation is the iterated map $\beta x \mod 1$. The special case of $\beta = 2$ is known as the Bernoulli map, and is exactly solvable. The Bernoulli map is interesting, as it provides a model for a certain form of pure, unrestrained chaotic (ergodic) behavior: it is the full invariant shift on the Cantor space $\{0,1\}^\omega$. The Cantor space consists of infinite strings of binary digits; it is particularly notable for a variety of features, one of which is that it can be used to represent the real number line.

The beta transformation is a subshift: iterated on the unit interval, it singles out a subspace of the Cantor space, in such a way that it is invariant under the action of the left-shift operator. That is, lopping off one bit at a time gives back the same subspace.

The beta transform is interesting, as it seems to capture something basic about the multiplication of two real numbers: β and x. It offers a window into understanding the nature of multiplication. Iterating on multiplication, one would get $\beta^n x$ – that is, exponentiation; although the mod 1 of the beta transform contorts this in interesting ways.

Analyzing the beta transform is difficult. The work presented here is a pastiche of observations, rather than that of deep insight or abstract analysis. The results are modest. Several interesting insights emerge. One is that the complicated structure of the chaotic regions of iterated maps seems to be due to the chaotic dynamics of the carry bit in multiplication. That is, the multiplication of real numbers, when expressed on Cantor space, requires the use of a carry bit (from long multiplication, as taught in elementary school). The carry bit behaves in a complicated fashion, and seems to be the root cause of chaotic structure. That is, if the carry bit is suppressed, then one obtains only reshufflings on the Cantor space, which have a completely uniform distribution lacking in structure: they are full shifts, not subshifts.

Another interesting, surprising result is that the eigenvalues of the transfer operator of beta transform seem to lie on a circle of radius $1/\beta$ in the complex plane. This is found numerically, using a natural Haar-wavelet-like basis that presents itself naturally for the problem. Given that the transfer operator is purely real, the appearance of such a quasi-unitary spectrum seems surprising. The relationships to orthogonal polynomials, Hessenberg matrixes and Jacobi matrixes is explored.

1 Introduction

The last three or four decades of mathematical research has seen dramatic advances in the theory of subshifts. This text is mostly not about that, except to point out that this theory has very broad and important impact on many branches of physics and mathematics. From the perpective of the amateur enthusiast, the theory of subshifts finally exposes and makes clear some of the mysterious and intriguing behavior of fractals and of chaotic dynamical systems.

This text focuses almost entirely on just one particular map of the unit interval, the β-transform, defined as the iterated map $\beta x \mod 1$. As such, it is an example of an iterated map on the unit interval of the real number line. Such maps have the form

$$f : [0,1] \to [0,1]$$

and the topic is the exploration of the consequence of iterating the map by composing:

$$f^n(x) = (f \circ f \circ \cdots \circ f)(x) = f(f(\cdots f(x) \cdots))$$

Such one-dimensional iterated maps have been heavily studied, and there is a large body of results, interconnecting many different concepts and results from mathematics, and so having a particularly broad range.

This text attempts to report some brand-new results on the β-transform. This is perhaps surprising, as the β-transform can be considered to be fairly well-studied, and fairly well-understood, it being among the very simplest of iterated one-dimensional maps. This text also attempts to report these results in a naive and unsophisticated fashion, in the hope that this makes the text readable for the interested student and casual enthusiast.

Thus, although the author is personally excited by the advances in the field, this text is neither a survey of known results on the β-transform, nor does it much glance at most of the typical avenues that are available for studying one-dimensional maps. It does focus on one particular technique: that of the transfer operator, and thus it contributes to the general "Koopmania". Little prior knowledge is assumed, and the needed concepts are introduced in a very casual and informal way. This will, no doubt, completely discourage and dismay the formally trained mathematician. The best I can offer is to reiterate: "new results", and not along the usual well-worn paths.

This text begins with some pretty pictures, showing the iterated tent and logistic maps, so as to whet the readers appetite for the considerably more dry material that follows. The goal, as always to to simplify and abstract. This is a difficult task.

1.1 Bernoulli shift

The Bernoulli shift (aka the bit-shift map) is an iterated map on the unit interval, given by iteration of the function

$$b(x) = \begin{cases} 2x & \text{for } 0 \le x < \frac{1}{2} \\ 2x-1 & \text{for } \frac{1}{2} \le x \le 1 \end{cases} \tag{1}$$

The symbolic dynamics of this map gives the binary digit expansion of x. That is, write

$$b^n(x) = (b \circ b \circ \cdots \circ b)(x) = b(b(\cdots b(x) \cdots))$$

to denote the n-fold iteration of b and let $b^0(x) = x$. The symbolic dynamics is given by the bit-sequence

$$b_n(x) = \begin{cases} 0 & \text{if } 0 \le b^n(x) < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \le b^n(x) \le 1 \end{cases} \tag{2}$$

Of course, the symbolic dynamics recreates the initial real number:

$$x = \sum_{n=0}^{\infty} b_n(x)\, 2^{-n-1}$$

All of this is just a fancy way of saying that a real number can be written in terms of it's base-2 binary expansion. That is, the binary digits for x are the $b_n = b_n(x)$, so that

$$x = 0.b_0 b_1 b_2 \cdots$$

The Bernoulli shift has many interesting properties, connecting it to the Cantor set and to many self-similar fractals. I have explored these in many other texts, as have other authors, and will not repeat these here. The author is too lazy to provide a bibliography; the reader is directed at search engines.

The current task is to attempt to see how many of these properties still hold in slightly more complex systems, and whether any of the tools used to analyze and solve the Bernoulli shift can be applied to these systems.

1.2 Shift space

The use of the word "shift" here deserves a small bit of formality. A "shift space" can be formally defined to be a set of infinite sequences of a set of N letters (or symbols), together with a shift operator T that takes each sequence, and lops off the left-most symbol. For the Bernoulli shift, there are $N = 2$ letters, taken from the set $\{0,1\}$. For the Bernoulli shift, one is typically interested in the set of all possible infinite sequences: this is the "full shift". One writes $\{0,1\}^{\omega}$ for this shift space, ω denoting countable infinity. For the Bernoulli shift, the map $b(x)$ is the shift operator: it just lops of the left-most symbol.

In general, a shift space does not have to include every possible sequence of symbols; it does, however, by definition, have to be shift-invariant. That is, given some set S of infinite sequences of N symbols, the set S is a shift space if and only if, by loping off the leading symbol of each string, one regains S again. In formulas, a shift space S must obey

$$TS = S$$

For example, $S = \{000\cdots, 111\cdots\}$ contains only two elements: the string of all zeros, and the string of all ones; loping off the leading digit just returns S again. In general, shift spaces may contain a finite, or a countable, or an uncountable number of

elements. In general, one defines the "full shift" as the space N^ω of all possible strings of N symbols. Subsets that are shift spaces are called "subshifts".

The words "symbolic dynamics" also deserve some mention: given one specific sequence x out of the shift space, one can ponder "where it goes to", as one lops off a symbol at a time. This gives the "symbolic dynamics" or the "point dynamics" of the sequence. The "orbit" is defined as the set $\{T^m x \,|\, \text{integer } m \geq 0\}$– that is, the set of all places that x goes to. There are several possibilities: one is that x is a fixed point, so that $Tx = x$. Another is that x is a repeating sequence of symbols, in which case iteration repeats as well: $T^m x = x$ holds whenever the repeat length is m; this is a periodic orbit. Most importantly, there usually uncountably many non-periodic sequences or orbits. That is, the number of periodic orbits is always countable: one merely arranges them in lexicographic order, and one is done. As Cantor famously demonstrated (and Hilbert so carefully expanded on) this cannot be done for the non-periodic orbits: they are uncountable.

In what follows, the text will in general confine itself to uncountable case. Periodic orbits exist, but will be ignored; in a certain strict sense, they constitute a set of measure zero. A number of glosses like this will be made: for example, the real numbers, and the Cantor space $\{0,1\}^\omega$ are both uncountable; however, they are not in one-to-one correspondence, as some real numbers can have two different representations as bit sequences. Specifically, these are the fractions $(2n+1)/2^m$ for positive integers m,n – they can be validly represented by bit-sequences ending in all-zeroes, or all-ones. There are countably many such fractions, termed the dyadic fractions. For the most part, this difference between the real number line, and the Cantor space will be ignored.

1.3 Downshift

The downshift is similar to the Bernoulli shift, replacing the number 2 by a constant real-number value $1 < \beta \leq 2$. It can be defined as

$$T_\beta(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} \\ \beta\left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \tag{3}$$

This map, together with similar maps, is illustrated in figure 5 below.

Just as the Bernoulli shift generates a sequence of digits, so does the downshift: write

$$k_n = \begin{cases} 0 & \text{if } 0 \leq T_\beta^n(x) < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq T_\beta^n(x) \leq 1 \end{cases} \tag{4}$$

Given the symbolic dynamics, one can reconstruct the original value whenever $1 < \beta$ as

$$x = \frac{k_0}{2} + \frac{1}{\beta}\left(\frac{k_1}{2} + \frac{1}{\beta}\left(\frac{k_2}{2} + \frac{1}{\beta}\left(\frac{k_3}{2} + \frac{1}{\beta}(\cdots)\right)\right)\right)$$

That is, one clearly sees that $T_\beta(x)$ acts as a shift on this sequence:

$$T_\beta(x) = \frac{k_1}{2} + \frac{1}{\beta}\left(\frac{k_2}{2} + \frac{1}{\beta}\left(\frac{k_3}{2} + \frac{1}{\beta}\left(\frac{k_4}{2} + \frac{1}{\beta}(\cdots)\right)\right)\right)$$

In this sense, this shift is "exactly solvable": the above provides a closed-form solution for iterating and un-iterating the sequence.

Multiplying out the above sequence, one obtains the series

$$x = \frac{1}{2} \sum_{n=0}^{\infty} \frac{k_n}{\beta^n} \tag{5}$$

That is, the bit-sequence that was extracted by iteration can be used to reconstruct the original real number. Setting $\beta = 2$ in eqn 2 gives the Bernoulli shift.

Unlike the Bernoulli shift, not every possible bit-sequence occurs in this system. It is a subshift of the full shift: it is a subset of $\{0,1\}^\omega$ that is invariant under the action of T_β. This is explored in greater detail in a later section.

1.4 Density Visualizations

Why is the beta transform interesting to explore? This can be partly illustrated with some graphs. Shown in figure 2 is the "bifurcation diagram" for the beta transform. It visualizes the long-term dynamics of the downshift. Comparing to the usual bifurcation diagram, e.g. for the Feigenbaum logistic map (shown in figure 4) one thing becomes immediately apparent: there are no actual "bifurcations", no "islands of stability", no period-doubling regions. Although there are periodic orbits, these form a set of measure zero: the iteration produces purely chaotic motion for all values of β. Thus, the beta transform provides a clean form of "pure chaos", without the pesky "islands of stability" popping up intermittently.

The visualization of the long-term dynamics is done by generating a histogram, and then taking the limit, as follows. One divides the unit interval into a fixed sequence of equal-sized bins; say a total of N bins, so that each is $1/N$ in width. Pick a starting x, and then iterate: if, at the n'th iteration, one has that $j/N \le b_\beta^n(x) < (j+1)/N$, then increment the count for the j'th bin. After a total of M iterations, let $c(j;M)$ be the count in the j'th bin. This count is the histogram. In the limit of a large number of iterations, as well as small bin sizes, one obtains a distribution:

$$\rho(y;x) = \lim_{N\to\infty} \lim_{M\to\infty} \frac{c(j;M)}{M} \text{ for } \frac{j}{N} \le y < \frac{j+1}{N}$$

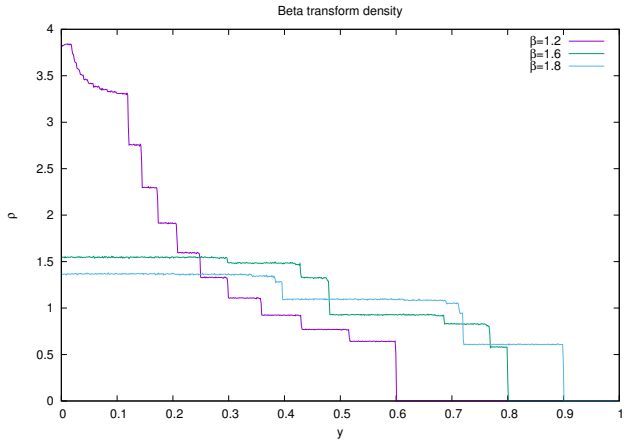
This distribution depends on the initial value x chosen for the point to be iterated; a "nice" distribution results when one averages over all starting points:

$$\rho(y) = \int_0^1 \rho(y;x)\, dx$$

Numerically, this integration can be achieved by randomly sampling a large number of starting points. Observe that $\rho(y)$ is a probability distribution:

$$1 = \int_0^1 \rho(x)\, dx$$

Figure 1: Downshift Density Distribution



The above figure shows three different density distributions, for $\rho_{1.2}(y)$, $\rho_{1.6}(y)$ and $\rho_{1.8}(y)$, calculated numerically. This is obtained by histogramming a large number of point trajectories, as described in the text. The small quantities of jitter are due to a finite number of samples. To generate this figure, a total of $M = 4000$ iterations were performed, using randomly generated arbitrary-precision floats (the Gnu GMP package), partitioned into $N = 800$ bins, and sampled 24000 times (or 30 times per bin) to perform the averaging integral.

This probability distribution is an eigenstate of the transfer operator for the beta transform; the definition of the transfer operator of the beta transform is given later. Probability distributions are the same thing as measures; this particular distribution is invariant under iteration, and thus is often called the invariant measure, or sometimes the Haar measure.

For each fixed β, one obtains a distinct distribution $\rho_\beta(y)$. The figure 1 illustrates some of these distributions. Note that, for $\beta < 1$, the distribution is given by $\rho_\beta(y) = \delta(y)$, a Dirac delta function, located at $y = 0$.

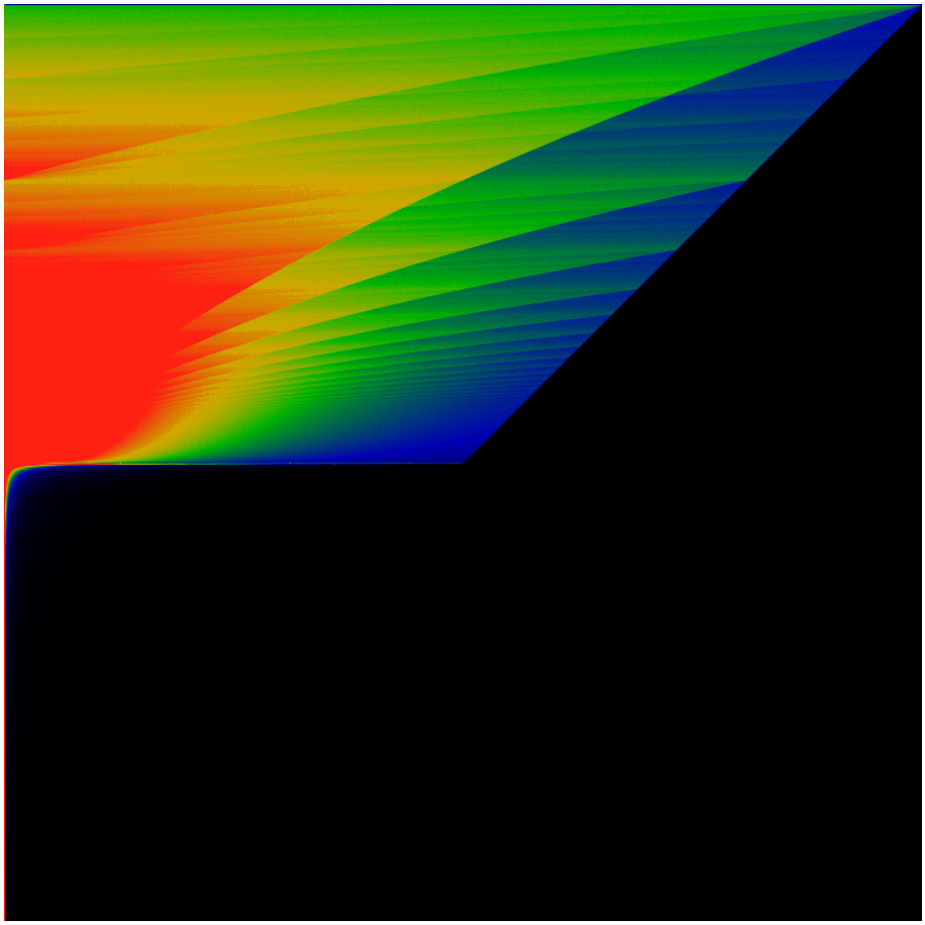
The general trend of the distributions, as a function of β, can be visualized with a Feigenbaum-style "bifurcation diagram", shown in figure 2. This color-codes each distribution $\rho_\beta(y)$ and arranges them in a stack; a horizontal slice through the diagram corresponds to $\rho_\beta(y)$ for a fixed value of β. The term "bifurcation diagram" comes from its use to visualize the logistic map iterator.

1.5 Tent Map

The tent map is a closely related iterated map, given by iteration of the function

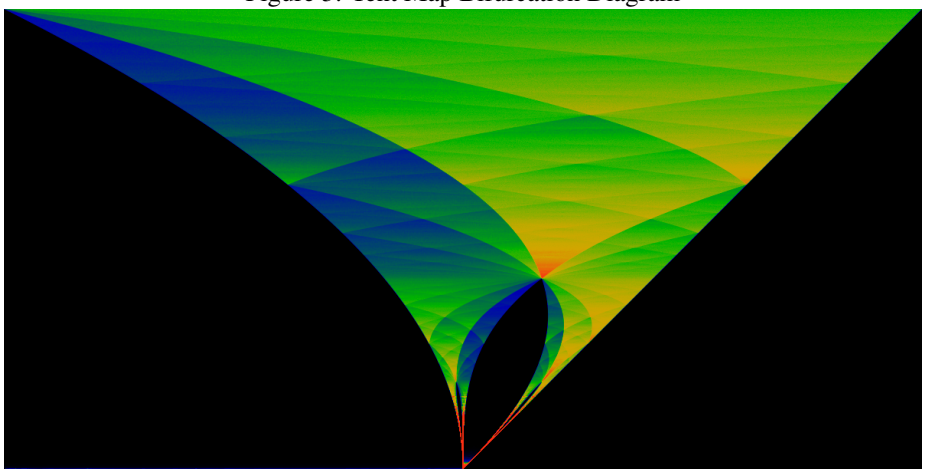
$$t_\beta(x) = \begin{cases} \beta x & \text{for } 0 \le x < \frac{1}{2} \\ \beta\,(1-x) & \text{for } \frac{1}{2} \le x \le 1 \end{cases}$$

Figure 2: Downshift Bifurcation Diagram



This figure shows the density $\rho_\beta(y)$, rendered in color. The constant β is varied from 0 at the bottom to 2 at the top; whereas y runs from 0 on the left to 1 on the right. Thus, a fixed value of β corresponds to a horizontal slice through the diagram. The color green represents values of $\rho_\beta(y) \approx 0.5$, while red represents $\rho_\beta(y) \gtrsim 1$ and blue-to-black represents $\rho_\beta(y) \lesssim 0.25$. The diagram is "interesting" only for $1 < \beta$; for smaller β's, one has that $\rho_\beta(y) = \delta(y)$, indicated by the column of red pixels on the left side of the image. The lines forming the fan shape are not actually straight, they only seem to be; in fact, they have a slight curve. This means that one cannot apply simple-minded guess-work to discover the overall diagram shown here.

Figure 3: Tent Map Bifurcation Diagram



The bifurcation diagram for the tent map. The value of β runs from 1 at the bottom of the image, to 2 at the top. The color scheme is adjusted so that green represents the average value of the distribution, red represents areas of more than double the average value, while blue shows those values that are about half the average value. Note that this is a different color scheme than that used in figure 2; that scheme would obliterate the lower half of this figure in red.
The black areas represent parts of the iterated range that are visited at most a finite number of times. To the left, a straight line indicates that after one iteration, points in the domain $\beta/2 \leq x \leq 1$ are never visited. To the right, points in the domain $0 \leq x \leq \beta(1-\beta/2)$ are never visited more than a finite number of times.

Its similar to the downshift, except that the second arm is reflected backwards, forming a tent. The bifurcation diagram is shown in figure 3. Its is worth contemplating the similarities between this, and the corresponding downshift diagram. Clearly, there are a number of shared features.

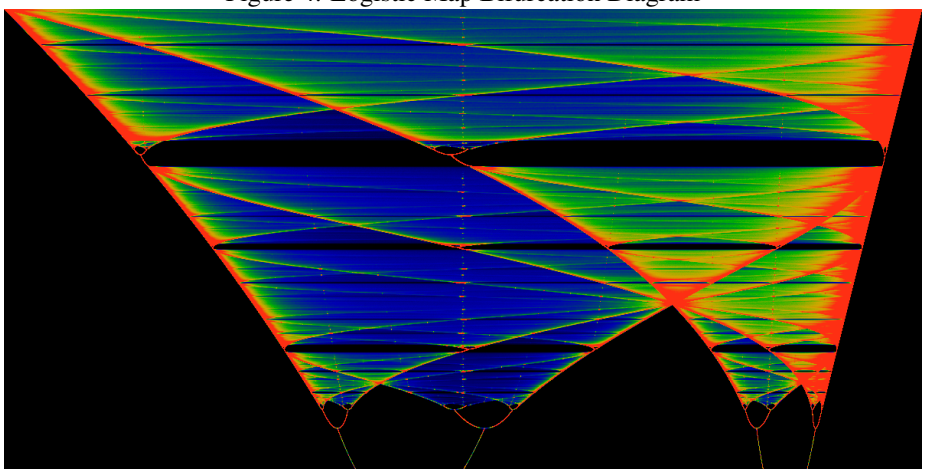
1.6 Logistic Map

The logistic map is related to the tent map, and is given by iteration of the function

$$f_\beta(x) = 2\beta x(1-x)$$

It essentially replaces the triangle forming the tent map with a parabola of the same height. That is, the function is defined here so that the the same value of β corresponds to the same height for all three maps. Although the heights of the iterators have been aligned so that they match, each exhibits rather dramatically different dynamics. The β-transform has a single fixed point for $\beta < 1$, and then explodes into a fully chaotic regime above that. By contrast, the logistic map maintains a single fixed point up to $\beta = 3/2$, where it famously starts a series of period-doubling bifurcations. The onset

Figure 4: Logistic Map Bifurcation Diagram



The logistic map bifurcation diagram. The value of β runs from 1.75 at the bottom of the image, to 2 at the top. The color scheme is adjusted so that green represents the average value of the distribution, red represents areas of more than double the average value, while blue shows those values that are about half the average value. Clearly, the orbits of the iterated points spend much of their time near the edges of the diagram.

---

of chaos is where the bifurcations come to a limit, at $\beta = 3.56995/2 = 1.784975$. Within this chaotic region are "islands of stability", which do not appear in either the β-transform, or in the tent map. The tent map does show a period-doubling regime, but in this region, there are no fixed points: rather, the motion is chaotic, but confined to multiple arms. At any rate, the period doubling occurs at different values of β than for the logistic map.

The bifurcation diagram is shown in figure 4. Again, it is worth closely examining the similarities between this, and the corresponding tent-map diagram, as well as the β-transform diagram. Naively, it would seem that the general structure of the chaotic regions are shared by all three maps. Thus, in order to understand chaos in the logistic map, it is perhaps easier to study it in theβ-transform.

The general visual similarity between the figures 2, 3 and 4 should be apparent, and one can pick out and find visually similar regions among these three illustrations. Formalizing this similarity is a bit harder, but it can be done: there is a way to make all three of these maps be "topologically conjugate" to one-another. This is perhaps surprising to some readers, but is based on the observation that the "islands of stability" in the logistic map are contuable, and are in one-to-one correspondance with certain "trouble points" in the iterated beta transformation. These are in turn in one-to-one correspondance with rational numbers. With a slight distortion of the beta transformation, the "trouble points" can be mapped to the islands of stability, in essentially the same way that "phase locking regions" or "Arnold tongues" appear in the circle map. But this is all for a later section, again, mentioned here only to whet the appetite.

1.7 Beta Transformation

After exactly one iteration of the downshift , all initial points $\beta/2 \le x \le 1$ are swept up into the domain $0 \le x < \beta/2$, and never leave. Likewise, the range of the iterated downshift is $0 \le x < \beta/2$. Thus, an alternative representation of the downshift, filling the entire unit square, can be obtained by dividing both x and y by $\beta/2$ to obtain the function

$$t_\beta(u) = \begin{cases} \beta u & \text{for } 0 \le u < \frac{1}{\beta} \\ \beta u - 1 & \text{for } \frac{1}{\beta} \le u \le 1 \end{cases} \tag{6}$$

which can be written more compactly as $t_\beta(x) = \beta x \mod 1$. In this form, the function is known as the beta transform, and is often called the β-transformation, as if to present a typesetting challenge to publishers. The orbit of a point x in the downshift is identical to the orbit of a point $u = 2x/\beta$ in the beta transformation.

Although the downshift and the β-transformation are essentially "the same function", this text works almost exclusively with the downshift. There is no particular technical reason for this; it is rather due to happenstance.

After a single iteration of the tent map, a similar situation applies. After one iteration, all initial points $\beta/2 \le x \le 1$ are swept up into the domain $0 \le x < \beta/2$. After a finite number of iterations, all points $0 < x \le \beta(1-\beta/2)$ are swept up, so that the remaining iteration takes place on the domain $\beta(1-\beta/2) < x < \beta/2$. It is worth defining a "sidetent" function, which corresponds to the that part of the tent map in which iteration is confined. It is nothing more than a rescaling of the tent map, ignoring those parts outside of the above domain that "wander away". The sidetent is given by

$$s_\beta(u) = \begin{cases} \beta(u-1)+2 & \text{for } 0 \le u < \frac{\beta-1}{\beta} \\ \beta(1-u) & \text{for } \frac{\beta-1}{\beta} \le u \le 1 \end{cases}$$

Performing a left-right flip on the side-tent brings it closer in form to the beta-transformation. The flipped version, replacing $u \to 1-u$ is

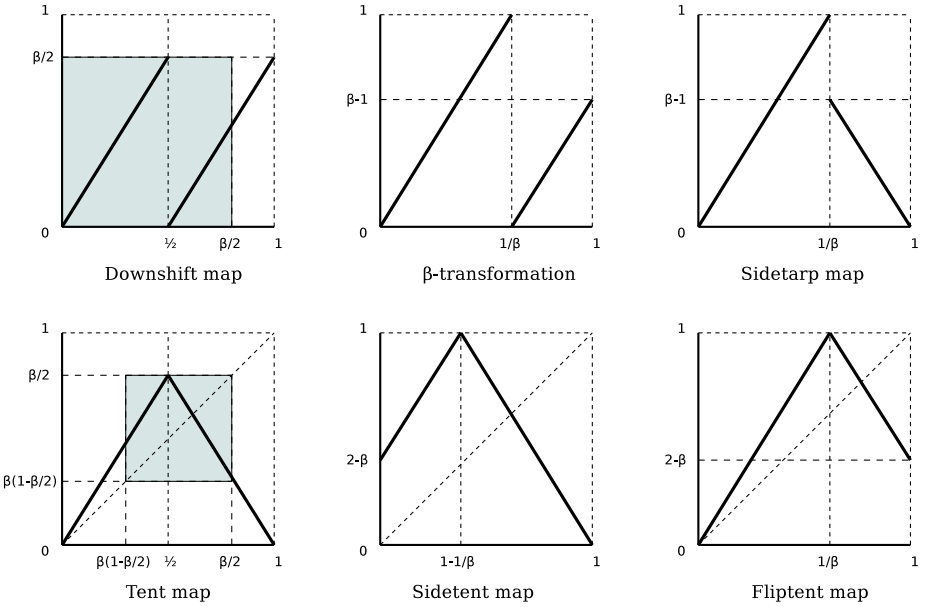
$$f_\beta(u) = \begin{cases} \beta u & \text{for } 0 \le u < \frac{1}{\beta} \\ 2 - \beta u & \text{for } \frac{1}{\beta} \le u \le 1 \end{cases}$$

The tent map (and the flipped tent) exhibits fixed points (periodic orbits; modelocking) for the smaller values of β. These can be eliminated by shifting part of the tent downwards, so that the diagonal is never intersected. This suggests the "sidetarp":

$$a_\beta(u) = \begin{cases} \beta u & \text{for } 0 \le u < \frac{1}{\beta} \\ \beta(1-u) & \text{for } \frac{1}{\beta} \le u \le 1 \end{cases}$$

The six different maps under consideration here are depicted in figure 5. It is interesting to compare three of the bifurcation diagrams, side-by-side. These are shown in figure 6.

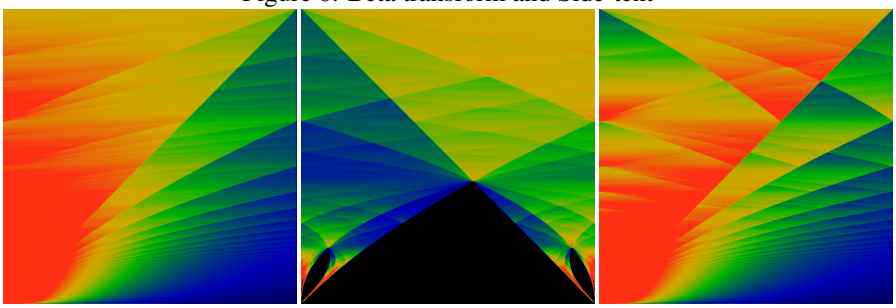
Figure 5: Iterated piece-wise linear maps



The downshift map, shown in the upper left, generates orbits that spend all of thier time in the shaded area: a box of size $\frac{\beta}{2} \times \frac{\beta}{2}$. Enlarging this box to the unit square gives the β-transformation. The tent map resembles the downshift, except that one arm is flipped to make a tent-shape. After a finite number of iterations, orbits move enitrely in the shaded region; enlarging this region to be the unit square gives the sidetent map. Flipping it left-right gives the fliptent map. Although it is not trivially obvious, the fliptent map and the sidetent map have the same orbits, and thus the same bifurcation diagram.

The bottom three maps all have fixed points and periodic orbits, esssentially because the diagonal intersects the map. The top three maps have no periodic orbits, and are purely chaotic, essentially because the diagonal does not intersect them. Note tht the slopes and the geometric proportions of all six maps are identical; they are merely rearrangents of the same basic elements.

Figure 6: Beta transform and Side-tent



The left figure shows the bifurcation diagram for the β-transform, as it is normally defined as the $\beta x \mod 1$ map. It is the same map as the downshift, just rescaled to occupy the entire unit square. In all other respects, it is identical to 2.

The middle figure is a similarly-rescaled tent map, given the name "side tent" in the main text. It is essentially identical to 3, with the middle parts expanded and the sides removed. In both figures, β runs from 1 at the bottom to 2 at the top. The right-hand-side figure is the "sidetarp", clearly its an oddly-folded variant of the beta transform.

1.8 Beta Transformation Literature Review and References

The β-transformation, in the form of $t_\beta(x) = \beta x \mod 1$ has been well-studied over the decades. The bit expansion 4 was introduced by A. Renyi[1], who demonstrates the existance of the invariant measure. The ergodic properties of the transform were proven by W. Parry[2], who also shows that the system is weakly mixing.

An explicit expression for the invariant measure was given by W. Parry[2], as

$$\rho_\beta(y) = \frac{1}{F}\sum_{n=0}^{\infty}\frac{d_n(y)}{\beta^n}$$

where d_n is the digit sequence

$$d_n(y) = \begin{cases} 0 & \text{if } t_\beta^n(1) \le y \\ 1 & \text{otherwise} \end{cases}$$

and F is a normalization constant.

The β-transformation has been shown to have the same ergodicity properties as the Bernoulli shift.[3] The symbolic dynamics of the beta-transformation was analyzed by F. Blanchard[4]. A characterization of the periodic points are given by Bruno Maia[5]. The fact tht the beta shift, and its subshifts are all ergodic is established by Climenhaga and Thompson[6]. The zeta function, and a lap-counting function, are given by Lagarias[7]. The Hausdorf dimension, the topological entropy and genreal notions of topological pressure arising from conditional variational principles is given by Daniel Thompson[8]. A proper background on this topics is given by Barreira and Saussol[9]. None of these topics are actually touched on below.

2 Symbolic Dynamics

The Bernoulli shift corresponds to the sequence of binary digits of a real number. Such sequences can be imagined to belong to the space of all possible sequences of binary digits, the Cartesian product of infinitely many copies of the set containing two elements:

$$\{0,1\} \times \{0,1\} \times \{0,1\} \times \cdots = \{0,1\}^{\omega} = 2^{\omega}$$

This space has a natural topology, the product topology, which differs sharply from the natural topology on the real-number line. Essentially all of the strange phenomena of fractals and of iterated functions follows from the product topology on this sequence.

One notable effect that can be explained in terms of the product topology is the fractal self-similarity of many kinds of fractals: this arises from the self-similarity of the product space under the action of a shift: specifically, the left-shift, which discards the left-most digit, and shifts the rest over by one. The shift operator itself is that operator that performs this shift; self-similar fractals can be seen to be eigenstates of the shift operator.

Another notable effect is the close proximity of the Cantor set to the proceedings. In a certain sense, the Cantor set can be understood to be the most basic manifestation of the product space. When attuned to its presence, it can be seen everywhere throughout the proceedings.

A third byproduct is the manifestation of the infinitely-deep binary tree. This arises when the set $\{0,1\}$ of the product space is re-interpreted as the set $\{L,R\}$ of left-right moves. At each point in a binary sequence, one can make a choice of one of two things: to move left or right. This naturally suggests a binary decision tree.

A fourth byproduct is the presence of some implicit, ambient hyperbolic space. The infinite binary tree, when drawn on flat two-dimensional space, simply "runs out of room", as each subsequent branching pushes closer together. The infinite binary tree can be embedded in the simplest hyperbolic space, the Poincaré disk or upper-half-plane, in such a way that the distance, the spacing between two neighboring nodes is always the same. Visually, this takes the form of some prototypical M.C. Escher drawing, of a repeated fractal form moving out to the edge of a disk. This makes the self-similar shape of the infinite binary tree manifest: as one moves from one location to another, one always sees "the same thing" in all directions: the space is homogeneous.

The rational numbers play a very special role in the infinite binary tree. Dyadic rationals, of the form $(2p+1)/2^n$ for integers p and n correspond to bit sequences (eqn 2) that terminate in all-zeros after a finite number of moves. That is, after an initial "chaotic" sequence, they settle down to a fixed point of period one. General rational numbers p/q behave similarly, in that after an initial "chaotic" sequence, they settle down to periodic orbits of some fixed period. The bit-sequence becomes cyclic. This cyclic behavior implies that most of classical number theory can be dragged into the proceedings. Any particular statement that classical number theory makes with regard to rational numbers, or even modular forms, can be promptly ported over to a statement about the bit-sequences and the orbits of the Bernoulli shift, usually taking on a strange and unrecognizable form.

All of these things go together, like hand in glove: whenever one is manifest and

visible, the others are lurking right nearby, in the unseen directions. All of these things can be given a formal and precise definition, and their explicit inter-relationships articulated. This has been done by a wide variety of authors over the last four decades; a proper bibliography would be overwhelming. I have written on all of thee topics, trying to present them in the simplest, most jargon-free way that I can, in a dozen different texts available wherever you found this one. The ideas will not be repeated here; they are not immediately useful to the current proceedings. None-the-less, the general interplay between all of these concepts is extremely important to understand, and burbles constantly under the surface of the current proceedings. In essence, shifts and subshifts are interesting precisely because they touch on so any different topics; and, conversely, so many different areas of mathematics can inform the subshift.

2.1 Symbolic Dynamics

Given that iteration can generate strings of binary digits, and that these can be reassembled back into rel numbers, it is interesting to ask what those mappings look like. To firm up the notation, let$(b_n) = (b_0, b_{1,}, \cdots)$ denote a sequence of bits (or symbols) and write

$$x_\beta\left((b_n)\right) = \frac{1}{2}\sum_{n=0}^{\infty}\frac{b_n}{\beta^n}$$

as the real number generated from that sequence. Conversely, given a real number x, let $\left(k_{n;\beta}(x)\right)$ denote the sequence of bits obtained by iterating the downshift on x with constant β; that is, the sequence generated by eqn. 4. The bit sequence for $\left(k_{n;2}(x)\right)$ is just the bit sequence $\left(b_n(x)\right)$ generated by eqn 2. The transformations between symbol sequences and real numbers make sense only when $1 < \beta \leq 2$.

Two interesting functions can be considered. One is the compressor

$$\mathrm{cpr}_\beta(y) = x_2\left(\left(k_{n;\beta}(y)\right)\right)$$

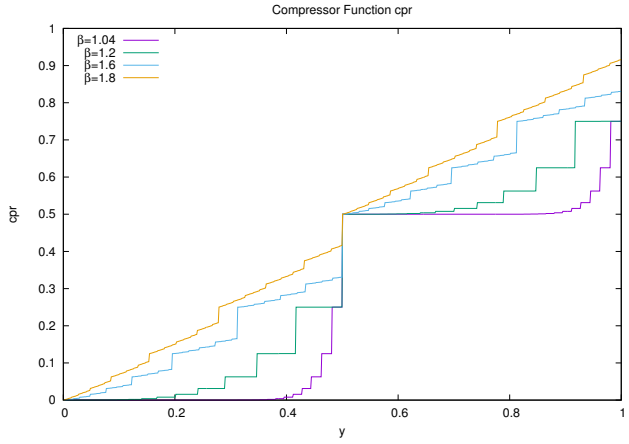
and the other is the expander

$$\mathrm{pdr}_\beta(y) = x_\beta\left(\left(k_{n;2}(y)\right)\right) \tag{7}$$

The terms "compressor" and "expander" are being invented here to indicate negative and positive Lyapunov exponents associated with the two functions. For almost all y, the compressor function is pushing nearby points closer together; the total measure of the range of the compressor function is less than one. Likewise, for almost all y, the expander function is pushing nearby points apart. These two functions are illustrated in figures 7 and 8.

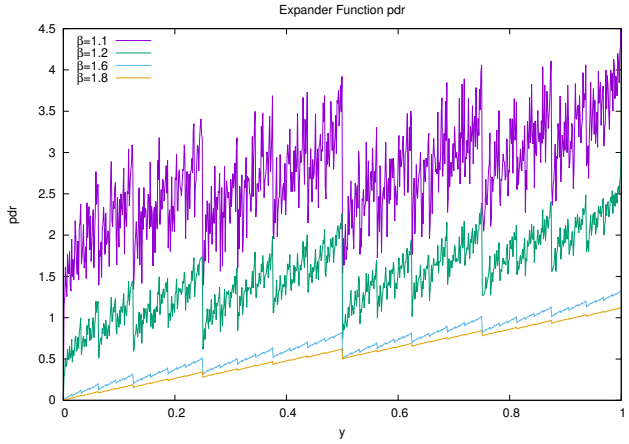
The two functions are adjoint; specifically, one has that $\mathrm{pdr}_\beta\left(\mathrm{cpr}_\beta(y)\right) = y$ but that $\mathrm{cpr}_\beta\left(\mathrm{pdr}_\beta(y)\right) \neq y$. The former relation is equivalent to eqn. 5. Not all possible sequences of bit strings appear in the downshift sequence $\left(k_{n;\beta}(x)\right)$; that is, this function is not a surjection onto $\{0,1\}^\omega$. This manifests itself as the gaps in the range of the compressor function, clearly visible in figure 7. If a sequence of bits is viewed

Figure 7: Compressor Function



This illustrates the compressor function for various values of β. As should be clear, almost all input values are mapped to a set of discrete output values.

Figure 8: Expander Function



This illustrates the expander function for various values of β. As should be clear, almost all neighboring input values are mapped to wildly different output values.

---

as a sequence of left-right moves walking down a binary tree, this implies that some branches of the tree are never taken, and can be pruned. Only branches on the right are ever pruned: That is, there can be arbitrarily long sequences of zeros in the expansion, but the longest possible sequence of 1's is always bounded. The longest run of 1's possible is the largest value of n that satisfies

$$2 \geq \frac{1+\beta+\beta^2+\cdots\beta^{n-1}}{\beta^{n-1}}$$

Solving, the bound is

$$n = 1 + \left\lfloor \frac{-\log(2-\beta)}{\log\beta} \right\rfloor \tag{8}$$

That is, every n'th right branch is pruned from the binary tree. For example, a run of three 1's in a row is possible only for $\beta \geq \left(1+\sqrt{5}\right)/2 \approx 1.618034$ the Golden Ratio. The range of $\mathrm{cpr}_\beta(y)$ is most of, but not all of the Cantor set. The figure 9 visualizes the range of the compressor as a function of β.

2.2 Shifts with holes

Viewed as a shift space, as opposed to a cut-down binary tree, the trimming can be thought of as a punching of holes into the full shift. This requires a bit of mental gymnastics. Let (a,c) be an (open) interval on the real number line: $(a,c) = \{x \,|\, a < x < c\}$. Given the Bernoulli shift $b(x) = T_2(x)$ from eqns 1 or 3, consider the set

$$\mathscr{I}(a,c) = \{x \,|\, b^n(x) \notin (a,c) \text{ for any } n \geq 0\}$$

That is, as one iterates on some fixed x, one requests that no iterate $b^n(x)$ ever lands in the interval (a,c). In essence, one has punched a hole in the unit interval; this corresponds to a "hole" in the full Bernoulli shift. The set $\mathscr{I}(a,c)$ is what remains after punching such a hole.

How can this be visualized? Considering the case $n = 0$, its clear that $\mathscr{I}(a,c)$ cannot contain (a,c). That is, $\mathscr{I}(a,c) \cap (a,c) = \emptyset$. For $n = 1$, the interval (a,c) can come from one of two places: either from $\left(\frac{a}{2}, \frac{c}{2}\right)$ or from $\left(\frac{a+1}{2}, \frac{c+1}{2}\right)$, and so neither of these can be in $\mathscr{I}(a,c)$. Continuing, for $n = 2$, the intervals $\left(\frac{a}{4}, \frac{c}{4}\right)$, $\left(\frac{a+1}{4}, \frac{c+1}{4}\right)$, $\left(\frac{a+2}{4}, \frac{c+2}{4}\right)$ and $\left(\frac{a+3}{4}, \frac{c+3}{4}\right)$ must also be gone. Continuing in this fashion, one proceeds with an infinite hole-punch: to obtain $\mathscr{I}(a,c)$, one just cuts out (a,c) and everything that iterates to (a,c). For the holes, write

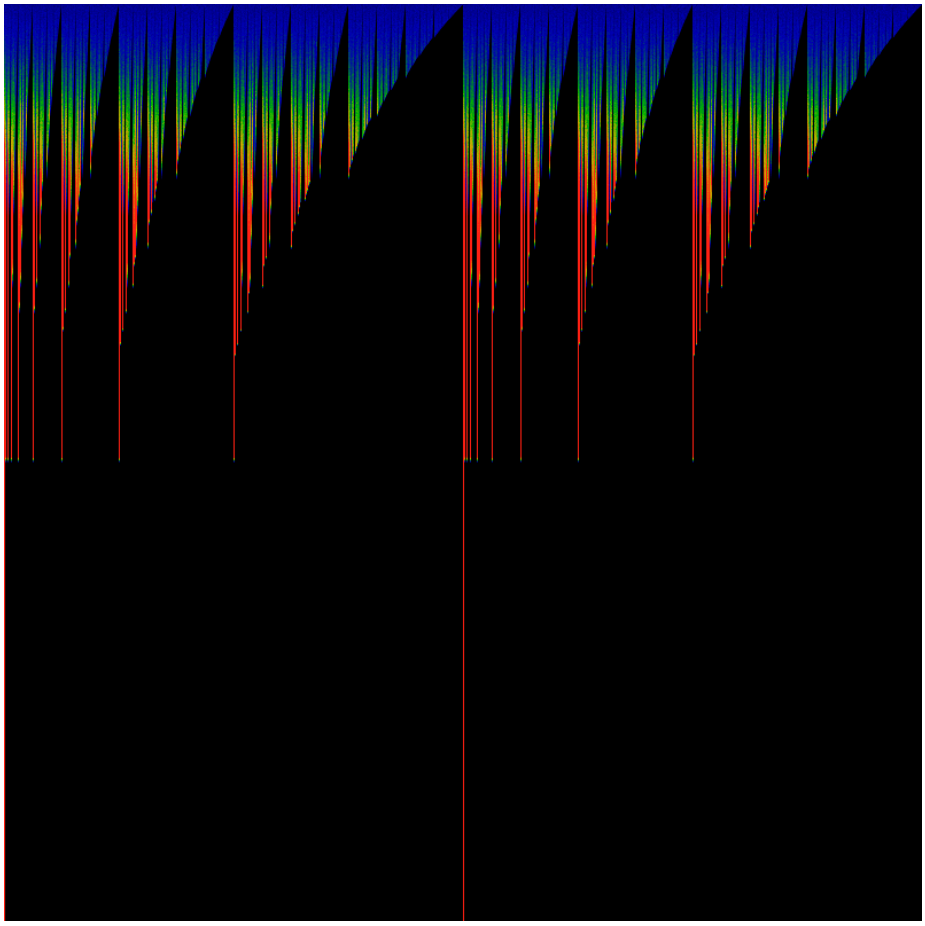
$$\mathscr{H}(a,c) = \bigcup_{n=0}^{\infty} \bigcup_{k=0}^{2^n-1} \left(\frac{a+k}{2^n}, \frac{c+k}{2^n} \right)$$

and for the interval with the holes punched out:

$$\mathscr{I}(a,c) = [0,1] \setminus \mathscr{H}(a,c)$$

where $\bigcup$ denotes set-union and $\setminus$ denotes set subtraction. It is not hard to see that, in the end, this forms a contorted Cantor set, using the standard midpoint-subtraction algorithm. The canonical Cantor set is built by taking $(a,c) = \left(\frac{1}{3}, \frac{2}{3}\right)$.

Figure 9: Range of the compressor



This figure illustrates a color coded visualization of the range of the compressor function. As before β varies from 0 at the bottom to 2 at the top, and y varies from 0 on the left to 1 on the right. In general, the compressor function maps intervals of the real number line to single points; the color corresponds to the size (the measure) of the intervals that were mapped to that particular point. Blue corresponds to a compression of the measure by about 1, green to a compression of about 2-3, and yellow-red to a compression greater than that.

Note that both $\mathscr{H}(a,c)$ and $\mathscr{I}(a,c)$ are subshifts: applying the left-shift to them just returns the same set again. Bot are invariant under the action of the shift operator. In formulas,

$$b\mathscr{H}(a,c) = \mathscr{H}(a,c)$$

and

$$b\mathscr{I}(a,c) = \mathscr{I}(a,c)$$

where, for notational simplicity, the parenthesis are not written, so that for the set S, write $bS = b(S)$. As shifts, its more appropriate to view both as sets of bit-sequences, so that the proper relationship between one and the other should have been written as

$$\mathscr{I}(a,c) = \{0,1\}^{\omega} \setminus \mathscr{H}(a,c)$$

How should these subshifts be visualized as strings? Let $(b_n(x))$ be the bit sequence generated by x, for some $a < x < c$. The cut operation states that such strings can never occur anywhere in $\mathscr{I}(a,c)$. Explicitly, $\mathscr{I}(a,c)$ never contains sequences of the form $d_0 d_1 d_2 \cdots d_k b_0(x) b_1(x) b_2(x) \cdots$ for any arbitrary leading bits $d_0 d_1 d_2 \cdots d_k$.

How should these subshifts be visualized as binary trees? The simplest case to visualize is to take $a = m/2^n$ and $c = (m+1)/2^n$ being dyadic rationals, for some integers m, n. In this case, one takes the bit-expansion for both have the same n leading bits: one starts at the root of the tree, and walks down the binary tree, making left-right moves in accordance with this sequence, and after n moves, arrives at a node above a subtree. Just cut out this subtree, in it's entirety. That's the first cut. Now repeat the process, for the left and right subtrees, from off the root, *ad infinitum*. For a and c not dyadic rationals, the process is more complicated. If a and c are ordinary rationals, thus having a repeating bit-sequence, one performs in the same way, but cyclically walking down the side branches of subtrees. For a and c irrational, the algorithm is considerably more complicated, and is left as an exercise for the reader :-).

A general classification of shifts with holes, for the beta transform, was performed by Lyndsey Clark[10].

2.3 Generalized compressors and expanders

The range of the compressor function is a shift with a hole. Specifically, for a given β, the range of cpr_β is $\mathscr{I}\left(\frac{\beta}{2}, \frac{1}{2}\right)$. The construction for shifts with holes can then be applied to construct generalized compressor and expander functions. One way, which is really rather cheesy, but it works, is to define the function

$$\mathrm{dcpr}_{\beta,\gamma}(a;x) = \sum_{n=0}^{\infty} \left[\frac{1}{\gamma^{n+1}} \sum_{k=0}^{2^n-1} \delta\left(x - \frac{a+k}{\beta^n}\right) \right]$$

and then define the generalized compressor as

$$\mathrm{cpr}(a;x) = \int_0^x \mathrm{dcpr}(a;y)\, dy$$

That is, as one walks along the unit interval, from left to right, one picks up points with weights on them, obtaining a generalized Devil's staircase (Cantor-Vitali) function. This generalization does not seem to be terribly useful here, and is left to rot.

2.4 Self-similarity

Subshifts are, by definition, self-similar. If S is a subshift, and T is the shift operator, then $TS = S$ is a part of the definition of the subshift. It is fun to see how this actually manifests itself on the unit interval.

So, the two functions cpr and pdr are self-similar. The pdr function demonstrates classic period doubling self-similarity: namely, under $g(x) = x/2$, it behaves as

$$\left(\mathrm{pdr}_\beta \circ g\right)(x) = \mathrm{pdr}_\beta\left(\frac{x}{2}\right) = \frac{1}{\beta}\mathrm{pdr}_\beta(x)$$

while under reflection $r(x) = 1 - x$, it behaves as

$$\left(\mathrm{pdr}_\beta \circ r\right)(x) = \mathrm{pdr}_\beta(1-x) = \frac{\beta}{2(\beta-1)} - \mathrm{pdr}_\beta(x)$$

Note that

$$\lim_{x\to 1} \mathrm{pdr}_\beta(x) = \frac{\beta}{2(\beta-1)}$$

The full dyadic monoid is generated by the generators g and r; see other posts from me for lengthly expositions on the structure of the dyadic monoid and its relationship to the Cantor set and a large variety of fractals.

Here, g is the generator that corresponds to the shift operator T. The notation g is used only to stay consistent with other things that I've written. The generator r indicates that the subshift is also invariant under reflection; in this case, under the exchange of the symbols $0 \leftrightarrow 1$ in the corresponding shift.

The function cpr also exhibits self-similarity, although it alters (expands) what happens on the x axis. Several self-similarities are apparent. First, for $0 \le x \le 1$, one has

$$\mathrm{cpr}_\beta\left(\frac{x}{2}\right) = \frac{1}{2}\mathrm{cpr}_\beta\left(\frac{\beta x}{2}\right)$$

Equivalently, for $0 \le y \le \beta/2$ one can trivially restate the above as

$$\mathrm{cpr}_\beta\left(\frac{y}{\beta}\right) = \frac{1}{2}\mathrm{cpr}_\beta(y) \tag{9}$$

Although it follows trivially, this restatement helps avoid later confusion.

The left and right halves are identical to one-another, but offset:

$$\mathrm{cpr}_\beta\left(\frac{1}{2} + \frac{x}{2}\right) = \frac{1}{2} + \mathrm{cpr}_\beta\left(\frac{x}{2}\right)$$

It follows that

$$\mathrm{cpr}_\beta\left(\frac{1}{2} + \frac{y}{\beta}\right) = \frac{1}{2} + \frac{1}{2}\mathrm{cpr}_\beta(y)$$

Combining the above results into one, one has that

$$\text{cpr}_\beta\left(\frac{y}{\beta}\right) + \text{cpr}_\beta\left(\frac{1}{2}+\frac{y}{\beta}\right) = \frac{1}{2} + \text{cpr}_\beta(y)$$

This last form is interesting, as it makes an appearance in relation to the transfer operator, defined below.

2.5 Other things with similar symmetry

The cpr curve is just one that belongs to a class of such curves. As an example, one may construct a Takagi (blancmange) curve by iterating triangles whose peak is located at $1/\beta$. The Takagi curve is an example of a curve transforming under a 3-dimensional representation of the dyadic monoid; the cpr curves transforms under a two-dimensional representation. See my paper on the Takagi curve for details. Figure 10 shows such a curve. Denote by $\text{tak}_{\beta;w}(x)$a curve constructed in this fashion. The transformation properties of this curve include self-similarity on the left, as

$$\text{tak}_{\beta;w}\left(\frac{x}{\beta}\right) = x + w\,\text{tak}_{\beta;w}(x)$$

for $0 \le x \le 1$ and self-similarity on the right, as

$$\text{tak}_{\beta;w}\left(\frac{1}{\beta} + x\left(1-\frac{1}{\beta}\right)\right) = 1 - x + w\,\text{tak}_{\beta;w}(x)$$

Both of these properties follow directly from the construction of the curve; they can be taken as the defining equations for the curve. That is, the curve can be taken as that function which satisfies these two recursion relations.

The derivative of the skew Takagi curve is shown in figure 11, and, for lack of a better name, could be called the skew Haar fractal wavelet. It can be defined as the formal derivative

$$\text{har}_{\beta;w}(x) = \frac{d}{dx}\text{tak}_{\beta;w}(x)$$

This formal derivative is well-defined, as the skew Takagi is smooth and piecewise-linear almost everywhere; the places where it has corners is a dense set of measure zero. That is, the derivative is defined everywhere, except on a set of measure zero, which happens to be dense in the unit interval.

Note that the Haar fractal wavelet is piece-wise constant everywhere. It is constructed from a "mother wavelet" given by

$$h_\beta(x) = \begin{cases} \beta & \text{for } 0 \le x < \frac{1}{\beta} \\ \frac{-\beta}{\beta-1} & \text{for } \frac{1}{\beta} \le x \le 1 \end{cases} \tag{10}$$

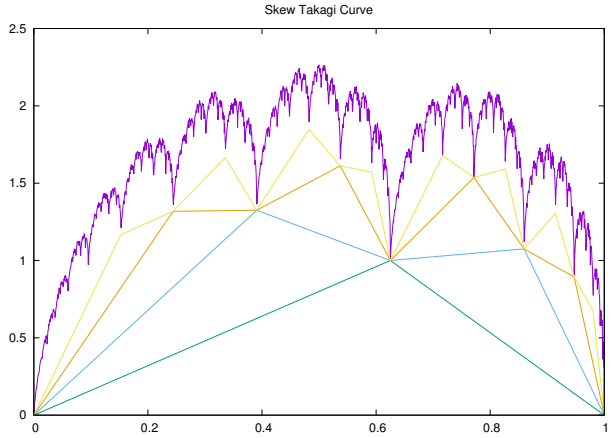
which is then iterated on to form the fractal curve $\text{har}_{\beta;w}(x)$. The self symmetries are

$$\text{har}_{\beta;w}\left(\frac{x}{\beta}\right) = \beta + w\,\text{har}_{\beta;w}(x)$$

and

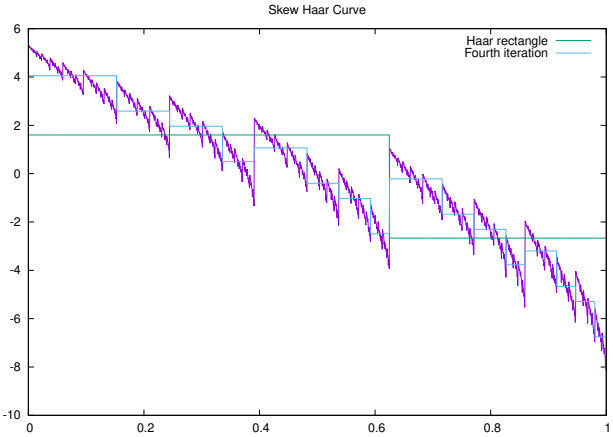
$$\text{har}_{\beta;w}\left(\frac{1}{\beta} + x\left(1-\frac{1}{\beta}\right)\right) = -\frac{\beta}{\beta-1} + w\,\text{har}_{\beta;w}(x)$$

Figure 10: Skew Takagi Curve



This figure shows a skew Takagi curve, and the first four steps of its construction. The initial triangle is of height 1; the apex is located at $1/\beta$, for $\beta = 1.6$ in this figure. Subsequent triangles obtain a height of w=0.7 above the apex point, and are similarly skew.

Figure 11: Skew Haar Wavelet



This figure shows the derivative of the skew Takagi curve. Note that it is piece-wise constant everywhere. The mother wavelet is shown, as well as the fourth iteration. The specific values graphed are $\beta = 1.6$ and $w = 0.7$.

2.6 Periodic Orbits

The Bernoulli shift, given by eqn 2, generates every possible bit-sequence. As was observed in a previous section, not every possible bit-sequence occurs in the downshift. The longest sequence of all-ones possible was given by eqn 8. Arbitrary finite lengths of zeros do appear; but are there fixed points, i.e. sequences that terminate in all-zeros? Clearly, $x = 1/2\beta^n$ is such a fixed point: after $n+1$ iterations of eqn 3, x goes to zero, and stays there. Is this the only such fixed point? The answer depends on β. If β can be written in the form of $\beta^n = 2m+1$ for some integers n and m, then the values of x which can iterate down to zero in $n+1$ steps are dense in the interval $[0, \beta/2]$. (TODO: this needs a simple, non-tedious proof). Otherwise, the fixed points are isolated. Curiously, such values β are dense in the interval $[1/2, 1)$.

Similar statements can be made about the values for which the orbits are periodic.

3 Transfer operators

The discovery and study of invariant measures, as well as of decaying states can be approached via the transfer operator, or, properly named, the Ruelle-Frobenius-Perron operator. This is an operator that captures the behavior of a distribution under the action of a map. The invariant measure is an eigenstate of this operator; indeed, it provides a formal definition for what it means to be invariant under the action of the map.

Given an iterated map $g : [0,1] \to [0,1]$ on the unit interval, the transfer operator defines how distributions are acted on by this map. It is defined as

$$\left[\mathscr{L}_g f\right](y) = \sum_{x=g^{-1}(y)} \frac{f(x)}{|g'(x)|}$$

The left adjoint of the transfer operator is the composition operator (Koopman operator). This is defined as

$$\left[\mathscr{C}_g f\right](y) = f\left(g(y)\right)$$

The Koopman operator is adjoint, in the sense that $\mathscr{L}_g \mathscr{C}_g = 1$ but that, in general, $\mathscr{C}_g \mathscr{L}_g \neq 1$.

3.1 The β-transform Transfer Operator

The transfer operator for the downshift map $b_\beta(x)$ is

$$\left[\mathscr{L}_\beta f\right](y) = \begin{cases} \frac{1}{\beta}\left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y}{\beta}+\frac{1}{2}\right)\right] & \text{for } 0 \le y \le \beta/2 \\ 0 & \text{for } \beta/2 < y \le 1 \end{cases}$$

or, written more compactly

$$\left[\mathscr{L}_\beta f\right](y) = \frac{1}{\beta}\left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y}{\beta}+\frac{1}{2}\right)\right]\Theta\left(\frac{\beta}{2}-y\right) \tag{11}$$

where Θ is the Heaviside step function. The density distributions graphed in figure 1 are those functions satisfying

$$\left[\mathscr{L}_\beta \rho\right](y) = \rho(y) \tag{12}$$

That is, the $\rho(y)$ satisfies

$$\rho(y) = \frac{1}{\beta}\left[\rho\left(\frac{y}{\beta}\right) + \rho\left(\frac{y}{\beta}+\frac{1}{2}\right)\right]\Theta\left(\frac{\beta}{2}-y\right) \tag{13}$$

This is generally referred to as the Ruelle-Frobenius-Perron eigenfunction, as it corresponds to the largest eigenvalue of the transfer operator, and specifically, the eigenvalue 1.

More generally, one is interested in characterizing the eigenspectrum

$$\left[\mathscr{L}_\beta \rho\right](y) = \lambda \rho(y)$$

for eigenvalues $|\lambda| \leq 1$ and eigenfunctions $\rho(y)$. Solving this equation requires choosing a space of functions in which to work. Natural choices include any of the Banach spaces, and in particular, the space of square-integrable functions. Particularly interesting is the space of almost-smooth functions, those having discontinuities at only a countable number of locations, but otherwise being infinitely differentiable. Although the discussion so far implicitly conditions one to restrict oneself to real-valued functions, and to consider only real-valued eigenvalues, this is perhaps too sharp a restriction. As will be seen below, the complex eigenvalues seem to play some important role. At any rate, it should be obvious that, whatever the choice of function space, one must have that $\rho(y) = 0$ whenever $\beta < 2y$. This turns out to be a rather harsh condition.

A very minor simplification can be achieved with a change of variable. Let $y = \frac{\beta}{2} - \varepsilon$. Then the eigenequation becomes

$$\lambda\beta\rho\left(\frac{\beta}{2}-\varepsilon\right) = \rho\left(\frac{1}{2}-\frac{\varepsilon}{\beta}\right) + \rho\left(1-\frac{\varepsilon}{\beta}\right)$$

The second term vanishes whenever $\beta/2 < 1-\varepsilon/\beta$ or $\varepsilon < \beta(1-\beta/2)$ and so one has the simpler recurrence relation

$$\lambda\rho(y) = \frac{1}{\beta}\rho\left(\frac{y}{\beta}\right) \tag{14}$$

whenever $\beta(\beta-1) < 2y \leq \beta$.

The equations 13 and 14 can be treated as recurrence relations, defining the $\lambda = 1$ eigenstate. Recursing on these gives exactly the densities shown in figure 1. Computationally, these are much, much cheaper to compute, at least for β much larger than 1, although convergence issues present themselves as β approaches 1. The resulting density may be called the Ruelle-Frobenius-Perron eigenstate; because it can be quickly computed, it provides an alternative view of figure 1, free of stochastic sampling noise.

3.2 Almost solutions

If one ignores the Heaviside step function in the definition, one easily finds a number of "almost solutions" to the transfer operator. These are most easily discussed by defining the operator

$$\left[\mathscr{P}_\beta f\right](y) = \frac{1}{\beta}\left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y}{\beta}+\frac{1}{2}\right)\right]$$

Solving this operator is relatively straight-forward. Consider, for example, the monomial $f(y) = y^n$. Clearly, $\left[\mathscr{P}_\beta y^n\right]$ is a polynomial of degree n and that therefore, the space of polynomials is closed under the action of $\mathscr{P}_\beta$. But this result is even stronger: the monomials provide a basis in which $\mathscr{P}_\beta$ is upper-triangular, *i.e.* solvable. It's eigensolutions in this basis are polynomials. The eigenspectrum is clearly discrete, and is given by $(\beta)^{-n-1}$ for integers n corresponding to the degree of the polynomial eigensolution.

This all goes horribly wrong if one instead considers $\mathscr{L}_\beta$ and the almost-monomials $f(y) = y^n\Theta\left(\frac{\beta}{2}-y\right)$. This does not provide a basis that is closed under the action of $\mathscr{L}_\beta$. Attempting to find the closure by iterating on $\mathscr{L}_\beta$ generates a splatter of step functions. This case is examined more closely in the next section.

Attempting some guess-work, the self-similarity of the cpr function suggests an opening. Specifically, let $\mathrm{ei}_\beta(x) = \mathrm{cpr}_\beta(x) - 1/2$. The one finds that

$$\begin{aligned}\left[\mathscr{P}_\beta \mathrm{ei}_\beta\right](y) &= \frac{1}{\beta}\left[\mathrm{ei}_\beta\left(\frac{y}{\beta}\right) + \mathrm{ei}_\beta\left(\frac{y}{\beta}+\frac{1}{2}\right)\right] \\ &= \frac{\mathrm{ei}_\beta(y)}{\beta}\end{aligned}$$

This is a non-polynomial, fractal eigenfunction of $\mathscr{P}_\beta$, and, with a bit of elbow-grease, one can find many more. This includes the Takagi functions, and their higher-order analogs, which are, roughly speaking, Takagi functions constructed from polynomials. These all have interesting self-similarity properties under the dyadic monoid.

Unfortunately, one has that $\mathrm{ei}_\beta(x) \neq 0$ when $\beta < 2x$; it won't do as an eigenfunction of $\mathscr{L}_\beta$. There is no obvious, simple modification of $\mathrm{ei}_\beta(x)$ that would cause it to be a valid eigensolution of $\mathscr{L}_\beta$. Manually installing a factor of $\Theta\left(\frac{\beta}{2}-y\right)$ and then iterating to find the closure leads to the same splatter of step functions as in the case of the polynomials.

Another interesting case arises if one attempts a Fourier-inspired basis. Define

$$\mathrm{e}_{\beta;n;l}(x) = \exp i2\pi(2l+1)\beta^n x$$

for integer l. One then obtains a shift sequence

$$\left[\mathscr{P}_\beta \mathrm{e}_{\beta;n;l}\right](x) = \frac{1}{\beta}\mathrm{e}_{\beta;n-1;l}(x)\left(1+\mathrm{e}_{\beta;n;l}\left(\frac{1}{2}\right)\right)$$

This is not a viable candidate for $\mathscr{L}_\beta$, as it is again beset by the step function. As a shift sequence, it can be used to construct coherent states that are eigenfunctions of $\mathscr{P}_\beta$,

having any eigenvalue within the unit disk. Specifically, observe that $\mathrm{e}_{\beta;0;l}(1/2) = \exp i\pi(2l+1) = -1$ so that $\left[\mathscr{P}_\beta \mathrm{e}_{\beta;0;l}\right](x) = 0$ and so the shift sequence terminates after finite iteration. Given a complex value z, construct the coherent state as

$$\phi_{l;z}(x) = \sum_{n=0}^{\infty} z^n \mathrm{e}_{\beta;n;l}(x)$$

The shift is then

$$\left[\mathscr{P}_\beta \phi_{l;z}\right](x) = \frac{z}{\beta} \sum_{n=0}^{\infty} z^n \left(1 + \mathrm{e}_{\beta;n+1;l}\left(\frac{1}{2}\right)\right) \mathrm{e}_{\beta;n;l}(x)$$

This is not particularly useful, until one notices that for for certain values of β, this contains nilpotent sub-series.

Specifically, fix a value of $n = N$ and consider those values of β for which $\mathrm{e}_{\beta;N;l}(1/2) = -1$. This holds whenever β^N is an odd integer, that is, whenever $\beta = (2m+1)^{1/N}$ (and, as always, $\beta \leq 2$). For these special values of β, one has that $\left[\mathscr{P}_\beta \mathrm{e}_{\beta;N;l}\right](x) = 0$ and so the functions

$$\phi_{l;z;N}(x) = \sum_{n=0}^{N} z^n \mathrm{e}_{\beta;n;l}(x)$$

vanish after N iterations of $\mathscr{P}_\beta$. That is, these can be used to form a a basis in which $\mathscr{P}_\beta$ is nilpotent. Conversely, letting m and N be free, the values for which $\beta = (2m+1)^{1/N}$ are dense in the interval $[1,2)$ and so any β is arbitrarily close to one with a nilpotent function space. These values of β are exactly the same values for which the bit sequences given by eqn 4 eventually terminate in all zeros; i.e. become periodic fixed points with period 1.

The existence of a dense set of fixed points is dual to the the existence of nilpotent densities. That is, one "causes" or "forces" the other to happen. This idea should be further elaborated, as it establishes a duality between countable and uncountable sets, which has an element of curiosity to it.

Presumably, there are special values of β which allow a periodic orbits to originate from a dense set. Such values of β, and such periodic orbits, should then correspond to specific self-similarities of the $\phi_{l;z}(x)$ function, specifically manifesting as cyclic behavior in $\left(1 + \mathrm{e}_{\beta;n+1;l}\left(\frac{1}{2}\right)\right)^p$ for some period p. Whether there is some similar manifestation for $\mathscr{L}_\beta$ is wholly unclear.

3.3 Iterated transfer operator

To understand the nature of the steady-state solution (the Frobenius-Perron eigenstate), its is worth iterating on the recurrence relation for it, by hand, the first few times. To do this, it is convenient to write it in the form

$$\left[\mathscr{L}_\beta f\right](y) = \frac{\Theta(y)}{\beta}\left[f(\alpha(y)) + f(\omega(y))\right]$$

where $\Theta(y) = 1$ if $y \leq \beta/2$ else zero; this is a step function to denote the vanishing for the operator for $2y > \beta$. (This differs from the use of Θ as the Heaviside step function

in earlier sections; the intent is the same, but the goal is to have a briefer notation here. Which is which should be clear from context.) The functions $\alpha(y) = y/\beta$ and $\omega(y) = \frac{1}{2} + \alpha(y)$ are convenient shorthands for symbolic iteration.

Iterating once gives

$$\begin{aligned}\left[\mathscr{L}^2 f\right](y) = \quad & \tfrac{\Theta(y)}{\beta^2} \quad \left[\Theta(\alpha(y))\left[f\left(\alpha^2(y)\right) + f\left((\omega\circ\alpha)(y)\right)\right] + \right. \\ & \left. \Theta(\omega(y))\left[f\left((\alpha\circ\omega)(y)\right) + f\left(\omega^2(y)\right)\right]\right]\end{aligned}$$

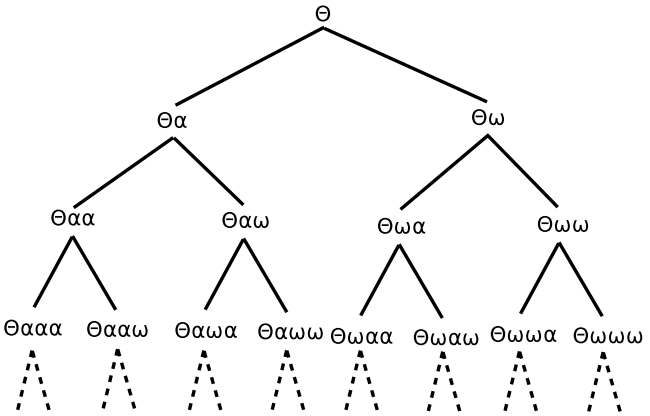
Using a simplified notation $g(y) = f(\alpha(y)) + f(\omega(y))$ allows this to be iterated a third time:

$$\begin{aligned}\left[\mathscr{L}^3 f\right](y) = \quad & \tfrac{\Theta(y)}{\beta^3} \quad \left[\Theta(\alpha(y))\left[\Theta\left(\alpha^2(y)\right) g\left(\alpha^2(y)\right) + \Theta(\omega\alpha(y))\, g(\omega\alpha(y))\right] + \right. \\ & \left. \Theta(\omega(y))\left[\Theta(\alpha\omega(y))\, g(\alpha\omega(y)) + \Theta\left(\omega^2(y)\right) g\left(\omega^2(y)\right)\right]\right]\end{aligned}$$

and a fourth time, this time omitting the argument, and the various nesting parenthesis.

$$\begin{aligned}\left[\mathscr{L}^4 f\right](y) = \quad & \tfrac{\Theta(y)}{\beta^4} \quad \left[\Theta\alpha\Theta\alpha^2\left[\Theta\alpha^3 g\alpha^3 + \Theta\omega\alpha^2 g\omega\alpha^2\right] + \right. \\ & \Theta\alpha\Theta\omega\alpha\left[\Theta\alpha\omega\alpha g\alpha\omega\alpha + \Theta\omega^2\alpha g\omega^2\alpha\right] \\ & \Theta\omega\Theta\alpha\omega\left[\Theta\alpha^2\omega g\alpha^2\omega + \Theta\omega\alpha\omega g\omega\alpha\omega\right] \\ & \left.\Theta\omega\Theta\omega^2\left[\Theta\alpha\omega^2 g\alpha\omega^2 + \Theta\omega^3 g\omega^3\right]\right]\end{aligned}$$

Notice that the primary structure is given by a product of step functions. This is more conveniently visualized as a tree:



For any given iteration, the result is the sum of the vertexes at a given level, while the product of step functions is the product of the step functions in the tree, following the path to each node. Because any particular step function might be zero, it effectively acts to cut off the tree at that location. It is therefore interesting to understand general products of the α and β functions.

It is convenient to define

$$\gamma_x(y) = \frac{x}{2} + \frac{y}{\beta}$$

so that $\alpha(y) = \gamma_0(y)$ and $\omega(y) = \gamma_1(y)$, so that a general iterated sequence of intermixed α's and ω's can be written uniformly in terms of γ. Given a sequence of bits $b_0 b_1 b_2 \cdots b_n$ with each b_k being either zero or one, the iterated sequence of functions can be written as

$$\left(\gamma_{b_0}\gamma_{b_1}\gamma_{b_2}\cdots\gamma_{b_n}\right)(y) = \frac{1}{2}\left[b_0 + \frac{b_1}{\beta} + \frac{b_2}{\beta^2} + \cdots + \frac{b_n}{\beta^n}\right] + \frac{y}{\beta^{n+1}} \tag{15}$$

So, for example:

$$\alpha^n(y) = \frac{y}{\beta^n}$$

while

$$\omega^2(y) = \frac{1}{2} + \frac{1}{\beta}\left(\frac{1}{2} + \frac{y}{\beta}\right)$$

and, in general, that

$$\omega^n(y) = \frac{1}{2}\left[1 + \frac{1}{\beta} + \frac{1}{\beta^2} + \cdots \frac{1}{\beta^{n-1}}\right] + \frac{y}{\beta^n}$$

Iterated sequences of pairs of functions, of the form $\gamma_{b_0}\gamma_{b_1}\gamma_{b_2}\cdots\gamma_{b_n}$ are reminiscent of de Rham curves, which generalize Cesaro curves and the Koch snowflake. The proper definition of a de Rham curve assumes the sequence is of infinite length, and applies a certain continuity condition, and is generally carried out on the complex plane, so that a continuous, nowhere-differentiable curve results. Here, the curve is distinctly not continuous: eqn 15 is a finite-length form of the shift series 5 which can be visualized as the expander function pdr 7, as shown in figure 8.

3.4 The Tree Function

Given a bit sequence (b_k) and value for y, define the tree function as

$$T_\beta((b_k); y) = \Theta(y)\prod_{n=0}^{\infty}\Theta\left(\gamma_{b_0}\gamma_{b_1}\gamma_{b_2}\cdots\gamma_{b_n}(y)\right)$$

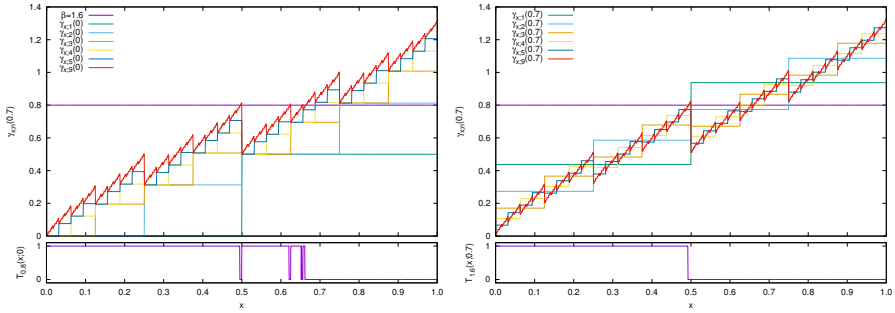
For any given fixed sequence of bits and value of y, this function is either zero or one. One way to understand this function is to ask how it varies for fixed β and y, but with the bit sequence coming from the Bernoulli shift of eqn 2, so that $b_n = b_n(x)$. This simplifies notation, so that one can write

$$T_\beta(x; y) = T_\beta((b_k(x)); y) = \Theta(y)\prod_{n=0}^{\infty}\Theta(\gamma_{x;n}(y))$$

with $\gamma_{x;n}(y) = \gamma_{b_0}\gamma_{b_1}\gamma_{b_2}\cdots\gamma_{b_n}(y)$. Its clear that the tree function has maximum support when $y = 0$. Figure 12 shows several gamma functions, and the corresponding tree function that results. Figure 13 shows the x vs. y behavior of the tree functions.

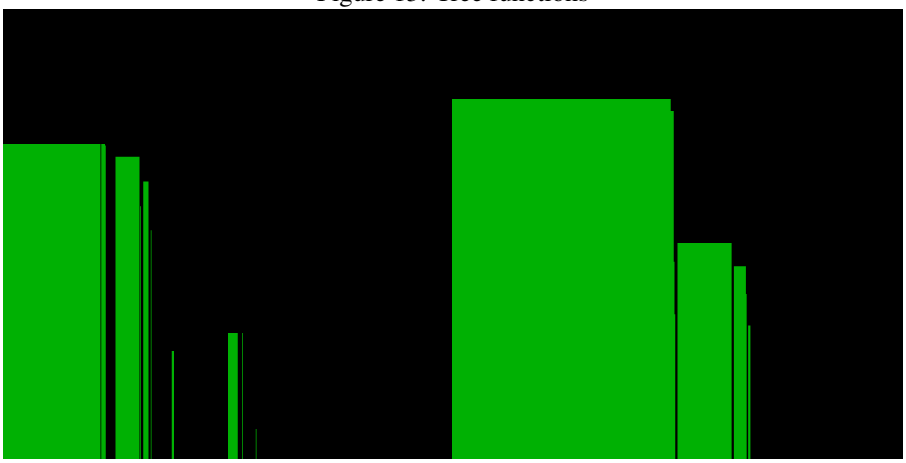
TODO: Graph the figures13 and 14 in the usual gloriously-colored-3D style.

Figure 12: Gamma functions



Examples of "typical" gamma functions. Both figures show gamma functions for $\beta = 1.6$; the one on the left shows them for $y = 0$, while the one on the right shows them for $y = 0.7$. Every gamma function is a sequence of plateaus; the zig-zag line is a high-order gamma, essentially showing the limiting case. The tree function is unity whenever all of these curves are below $\beta/2$, and is zero when above. So, for example, for the left figure, the tree function is unity, for all values of x less than about 0.4952; it drops to zero, then returns to unity above $x = 0.5$, until about 0.6221, when it briefly plunges and rises again. Then, another dip, before finally settling to zero near 0.6541. For the right figure, a high-order zig-zag rises above 0.8 somewhere near 0.4914; then $\gamma_{x;1}(0.7)$ rises above 0.8 and stays there, driving the tree function to zero, rendering all other orders irrelevant.

Figure 13: Tree functions



The above illustrate the y vs. x dependence of the tree functions; the left image is for $\beta = 1.4$, the right is for $\beta = 1.6$. Green indicates the regions where the tree function is unity, and black where it is zero. To be clear, this shows $T_\beta(x;y)$ with x and y plotted along the x and y axes. The tree functions shown in figure 12 are just two horizontal slices taken from the right image: a slice along the bottom, and a slice a bit above the middle.

3.5 Haar Basis Matrix Elements

The symmetric Haar wavelets are built from the mother wavelet

$$h(x) = \begin{cases} 1 & \text{for } 0 \le x < 1/2 \\ -1 & \text{for } 1/2 \le x < 1 \end{cases}$$

and has individual wavelets given by

$$h_{nj}(x) = 2^{n/2} h(2^n x - j) \quad \text{for } 0 \le j \le 2^n - 1$$

The matrix elements of the transfer operator are

$$\langle mi | \mathscr{L}_\beta | nj \rangle = \int_0^1 h_{mi}(x) \left[\mathscr{L}_\beta h_{nj}\right](x)\, dx$$

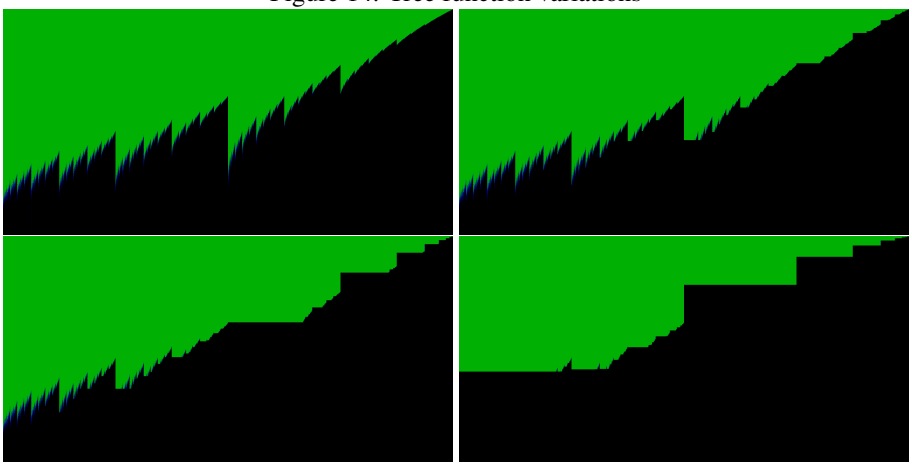
where the operator $\mathscr{L}_\beta$ is given by eqn 11. Computing these by hand promptly pushes into a big mess. One can obtain explicit expressions, just that they are tedious to obtain. Some preliminary observations include that

$$\langle mi | \mathscr{L}_\beta | nj \rangle = 0 \text{ if } \beta \le i/2^{m-1}$$

because the transfer operator vanishes above $\beta/2$. In the same vein, matrix elements vanish unless

$$\left[\frac{i}{2^m}, \frac{i+1}{2^m}\right] \cap \left[\frac{\beta j}{2^n}, \frac{\beta(j+1)}{2^n}\right] \neq \emptyset$$

Figure 14: Tree function variations



These figures illustrate the β vs. x dependence of the tree function. The upper left shows $T_\beta(x;0)$, the upper right shows $T_\beta(x;0.3)$, the lower left shows $T_\beta(x;0.5)$, the lower right shows $T_\beta(x;0.7)$. In each case, x runs from 0 to 1 along the x axis, while β runs from 1 to 2 along the vertical axis. As before, green indicates where the tree function is unity, and black where it is zero. The tree functions shown in figure 12 correspond to horizontal slices in the first and last images. Note that many (possibly all??) of the green spikes in the upper-left image reach all the way down to the bottom, although they are mostly much thinner than a pixel and thus not rendered. The vague blue hazing near the spikes is an attempt at anti-aliasing, to highlight the sharpness.

or if

$$\left[\frac{i}{2^m}, \frac{i+1}{2^m}\right] \cap \left[\beta\left(\frac{j}{2^n} - \frac{1}{2}\right), \beta\left(\frac{j+1}{2^n} - \frac{1}{2}\right)\right] \neq \emptyset$$

In all other cases, the Haar wavelets completely fail to overlap, and thus the matrix elements are zero. In addition, only three pairs of wavelets overlap in a non-zero fashion. That is, for a fixed m, n and j, there are at most six different values of i for which the matrix elements are non-vanishing: the first three of these are the values for which

$$\frac{\beta j}{2^n} \in \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right] \text{ or } \frac{\beta\left(j+\frac{1}{2}\right)}{2^n} \in \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right] \text{ or } \frac{\beta(j+1)}{2^n} \in \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right]$$

and likewise for three more. The observation is that the integral vanishes unless the first wavelet intersects an edge transition of the second wavelet.

The primary failure of this basis is that there is no obvious way to diagonalize the transfer operator in this basis. There is no obvious way of solving it, of finding it's eigenfunctions and eigenvalues, other than by brute-force numerical attack.

3.6 Julia Set

Consider the two iterators $a_0(y) = \min\left(\frac{\beta}{2}, \beta y\right)$ and $a_1(y) = \max\left(0, \beta y - \frac{\beta}{2}\right)$. Individually, they are the two arms of the downshift. Here, they have been separated from each other, so that the full domain $0 \leq y \leq 1$ is allowed. Exploring all possible inter-iterations for these gives the Julia set for the transfer operator: it indicates where a point "came from", for the iterated transfer operator. There are several related ways to visualize this. One way is to fix y and then, given a bit-sequence (b_n) to compute

$$j((b_n)) = a_{b_o} \circ a_{b_1} \circ a_{b_2} \circ \cdots (y)$$

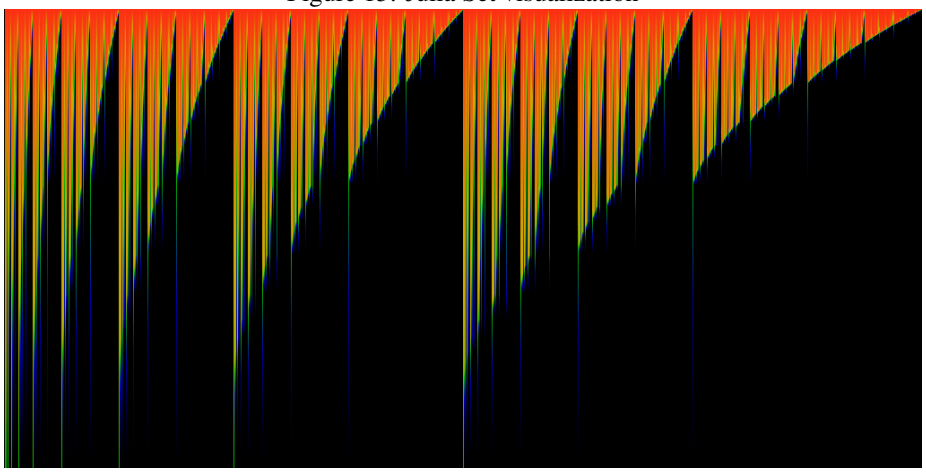
Figure 15 shows a visualization for finite bit-sequences: in essence, the very first few iterations. Although it is similar to figure 9, it is not the same.

For a related notion, consider the defintion of "laps", from Jeffrey Lagerias *etal*.[7].

4 Hessenberg basis

There is a set of Haar-like wavelets in which the transfer operator is of the form of a Hessenberg operator - that is, the operator becomes almost upper-diagonal, with only one diagonal, just below the main diagonal, that is non-zero. Explicitly, the transfer operator $\mathscr{L}_\beta$ has matrix entries $\left[\mathscr{L}_\beta\right]_{ij}$ such that $\left[\mathscr{L}_\beta\right]_{ij} = 0$ whenever $i > j+1$. A matrix having this form is called a Hessenberg matrix; such matrixes have various interesting properties, and occur naturally in spectral measure theory. In particular, they generalize the Jacobi operator for real Borel measures. This is explored in greater detail in this section.

Figure 15: Julia Set visualization



Consider the binary tree of dyadic fractions: that is, the tree whose rows are 1/2, (1/4 3/4), (1/8 3/8 5/8 7/8), ... Consider a function J on this tree. For the head of the tree, set $J(1/2) = \beta$. For the next row, set $J(1/4) = a_0(J(1/2))$ and $J(3/4) = a_1(J(1/2))$. Iterate in this fashion so that $J\left((2k-1)/2^{n+1}\right) = a_0(J(k/2^n))$ and $J\left((2k+1)/2^{n+1}\right) = a_1(J(k/2^n))$ recursively. This produces a function J taking values on every dyadic fraction $k/2^n$.

In the above figure, β runs from 1 at the bottom to 2 at the top. A single horizontal slice through the image shows a color-coded version of J, with red coding values near 1, green coding values near 1/2 and blue, fading to black coding values of 1/4 and less. Note that there are many faint blue lines that extend quite far down, but not all the way down: these form a stair-step. The image is 1024 pixels wide: it shows the first ten rows of the binary tree. Although this image is similar to figure 9, it differs in many details.

4.1 Hessenberg wavelet basis

The transfer operator can be fairly easily brought into Hessenberg matrix form. A sequence of of orthonormal functions is constructed in this section; when used as a basis, the transfer operator becomes almost upper-diagonal.

The trick to the construction is to define wavelets such that the transfer operator applied each wavelet causes the end-points of the wavelet to exactly line up with the end- or mid-points of previous wavelets, thus avoiding the nasty interval-overlap algebra required with the Haar basis. This is accomplished by carefully picking the midpoint of the next wavelet in the sequence to be located exactly at the discontinuity of the transfer operator applied to the previous wavelet.

The construction proceeds as follows. Let

$$\psi_0(x) = \begin{cases} \frac{1}{\sqrt{\beta/2}} & \text{for } 0 \le x \le \beta/2 \\ 0 & \text{for } \beta/2 < x \le 1 \end{cases}$$

Consider $\mathscr{L}_\beta \psi_0$. It is the sum of two parts: two step-functions; one which is constant for $x \le \beta/2$ and another that is constant for $\frac{x}{\beta} + \frac{1}{2} \le \frac{\beta}{2}$. Solving explicitly for the location of the step, it is $x = \beta(\beta-1)/2$. For convenience, define $m_1 = \beta(\beta-1)/2$ and $m_0 = \beta/2$. These will anchor a series of midpoints, beginning with $m_{-1} = 0$. Using the midpoint m_1, construct the wavelet

$$\psi_1(x) = \begin{cases} \frac{1}{m_1}\sqrt{\frac{m_1(m_0-m_1)}{m_0}} & \text{for } 0 \le x \le m_1 \\ \frac{-1}{m_0-m_1}\sqrt{\frac{m_1(m_0-m_1)}{m_0}} & \text{for } m_1 < x \le m_0 \\ 0 & \text{for } m_0 < x \le 1 \end{cases}$$

Note that this is normalized to unit length: $\int_0^1 |\psi_1(x)|^2 dx = 1$ and that it is explicitly orthogonal to the first: $\int_0^1 \psi_1(x)\psi_0(x)dx = 0$.

Consider $\mathscr{L}_\beta \psi_1$. As always, it is the sum of two parts. The midpoint of ψ_1 is at $m_1 = \beta(\beta-1)/2$ and this mid-point is mapped to one of two different places. If $m_1 < 1/2$ then it is mapped to $m_2 = \beta m_1$ else it maps to $m_2 = \beta(m_1 - 1/2)$. Thus, if $m_1 < 1/2$, define

$$\psi_2(x) = \begin{cases} 0 & \text{for } 0 \le x \le m_1 \\ \frac{1}{(m_2-m_1)}\sqrt{\frac{(m_2-m_1)(m_0-m_2)}{m_0-m_1}} & \text{for } m_1 \le x \le m_2 \\ \frac{-1}{(m_0-m_2)}\sqrt{\frac{(m_2-m_1)(m_0-m_2)}{m_0-m_1}} & \text{for } m_2 < x \le m_0 \\ 0 & \text{for } m_0 < x \le 1 \end{cases}$$

else define

$$\psi_2(x) = \begin{cases} \frac{1}{m_2}\sqrt{\frac{m_2(m_2-m_1)}{m_1}} & \text{for } 0 \le x \le m_2 \\ \frac{-1}{(m_1-m_2)}\sqrt{\frac{m_2(m_2-m_1)}{m_1}} & \text{for } m_2 \le x \le m_1 \\ 0 & \text{for } m_1 < x \le 1 \end{cases}$$

Because each end of the interval on which ψ_2 is non-zero lies entirely within one of the constant arms of ψ_1, one has, by construction, that $\int_0^1 \psi_2(x)\,\psi_1(x)\,dx = 0$ (and, of course, $\int_0^1 \psi_2(x)\,\psi_0(x)\,dx = 0$.)

The rest of the basis can be constructed iteratively, based on these examples. The midpoints are given by iterating 3 on $m_0 = \beta/2$, so that $m_p = T_\beta(m_{p-1}) = T_\beta^p(m_0)$ is the p'th iterate of $\beta/2$. Let m_l be largest midpoint smaller than m_p (and $l < p$); let m_u be the smallest midpoint larger than m_p (and $l < p$). Let $m_{-1} = 0$ initiate the sequence by providing the smallest-possible "midpoint"; $m_0 = \beta/2$ already provides the largest possible.

Then define

$$\psi_p(x) = \begin{cases} 0 & \text{for } 0 \le x \le m_l \\ \frac{C_p}{(m_p - m_l)} & \text{for } m_l \le x \le m_p \\ \frac{-C_p}{(m_u - m_p)} & \text{for } m_p < x \le m_u \\ 0 & \text{for } m_u < x \le 1 \end{cases} \tag{16}$$

By construction, this has the property that $\int_0^1 \psi_{p+1}(x)\,\psi_n(x)\,dx = 0$ for any $n < p+1$. The normalization constant is

$$C_p = \sqrt{\frac{(m_p - m_l)(m_u - m_p)}{m_u - m_l}}$$

which is determined by requiring that $\int_0^1 \left|\psi_p(x)\right|^2 dx = 1$.

4.2 The (Generalized) Golden Ratio

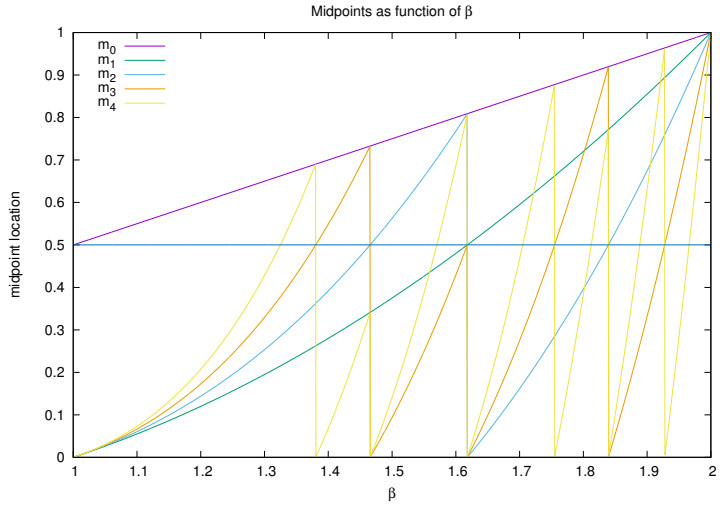
The above wavelet basis seems to be well-behaved, except when $\beta = \varphi = \left(1+\sqrt{5}\right)/2$ the Golden Ratio. In this situation, one has that $m_0 = \varphi/2$ and $m_1 = 1/2$. At this location, further iteration breaks down. That is, $m_2 = T_\varphi(m_1)$ can either be taken to be $m_2 = 0$ or $m_2 = m_0$. In the former case, iteration simply stops; in the later case, it repeats, again, without generating new midpoints that can provide a workable basis.

Working backwards, this issue re-appears whenever the p'th iterate $m_p = T_\beta^p(m_0)$ lands at the discontinuity, so that one may take either $m_p = 0$ or $m_p = m_0$. For $p = 3$, there are two trouble spots, which occur when either $\beta^3 - \beta^2 - 1 = 0$ or when $\beta^3 - \beta^2 - \beta - 1 = 0$. These correspond to the values of $\beta = 1.465571231876768\cdots$ and $\beta = 1.839286755214161\cdots$.

Where are the trouble spots located? Consider, for example, $m_4 = T_\beta^4(m_0)$, and consider the movement of m_4 as β is swept through the range $1 < \beta < 2$. This is shown in figure 16. As made clear in the image, three new trouble spots appear. These are located at $\beta = 1.380327757\cdots$ and $\beta = 1.754877668\cdots$ and $\beta = 1.927561978\cdots$, which are the real roots of $\beta^4 - \beta^3 - 1 = 0$ and $\beta^4 - \beta^3 - \beta^2 - 1 = 0$ and $\beta^4 - \beta^3 - \beta^2 - \beta - 1 = 0$ respectively.

Following a similar suggestion by Dajani[3], numbers of this kind may be called "generalized golden means". Unfortunately, the term "generalized golden mean" is in common use, and is applied to a variety of different systems. Not all are relevant; one

Figure 16: Location of Midpoints



This rather busy image illustrates the location of the first five midpoints, $m_0, m_1, \cdots, m_4$ as a function of β. The locations of the discontinuities are termed "trouble spots"; the first trouble spot occuring for m_2 at $\beta = \varphi$. The midpoint m_3 has two new trouble spots at $\beta = 1.465\cdots$ and $\beta = 1.839\cdots$; the trouble spot at $\beta = \varphi$ being disallowed, as it already lead to a termination of midpoint iteration. The midpoint m_4 has three new troublespots.

that is, is given by Hare *et al.*[11] who provide series expansions for the real roots of $\beta^n - \sum_{k=0}^{n-1} \beta^k = 0$. Stakhov[12] considers $\beta^{n+1} - \beta^n - 1 = 0$ in general settings. Some, but not all of these numbers are known to be Pisot numbers or Salem numbers[5].

How many trouble spots are there? The table below shows the count n_p of the number of "new" trouble spots, as a function of the midpoint index p.

p	n_p
2	1
3	2
4	3
5	6
6	9
7	18
8	30
9	56
10	99

This appears to be Sloane's OEIS A001037 which has a number of known relationships to roots of unity, Lyndon words, and the number of orbits in the tent map. The trouble spots are the positive real roots of polynomials of the form $\beta^p - \beta^{p-1} - b_2\beta^{p-2} - b_3\beta^{p-3} - \cdots - 1 = 0$. There is just one such root for each such polynomial. These polynomials must clearly be irreducible, as otherwise, the root would have been seen earlier. Since the digits b_k must be zero or one, this implies that the polynomils can be taken to be the irreducible polynomials over the field $\mathbb{Z}_2$ which is precisely the definition of OEIS A001037. (These last few sentences can be taken as a proof).

The coefficients b_k are not immediately obvious, but a stunningly bad conjecture is that they form Lyndon words, viz, that $0\cdots b_3 b_2 1$ is a Lyndon word, of length p. This clearly holds up to $p = 4$ but surely cannot hold any farther?

Clearly, iteration paths that lead to the trouble points can be formulated as left-right moves on the infinite binary tree. Is the resulting subtree a subshift? What, precisely, are it's self-symmetries? A relationship to OEIS A000002 Oldenburger-Kolakoski sequence is observed; that sequence has several fractal self-similarities which presumably describe the subshift.

There are several ways in which these sequences of troublesome β values can be visualized, and doing so is a worthwhile exercise.

4.3 Islands of Stability as Arnold Tongues

These trouble-spots can be placed in one-to-one accordance with the "islands of stability" seen in the iterated logistic map. They are, in essence, locations where periodic orbits "could be pasted", or where they "naturally would appear", if the map supported periodic attractors. That is, the downshift only supports a single attractor, of period-one at $x = 0$; there is no "room" for anything more. This is analogous, in a way, to the phase locked loop, at zero coupling constant. At finite coupling strength, these "trouble spots" expand out as Arnold tongues, to have a finite size, visible on the Feigenbaum diagram for the logistic map as regions where period-doubling is occuring.

The idea here can be illustrated explicitly. Given some "small", real $\varepsilon > 0$, define the ε-generalization of the map 3 as

$$T_{\beta,\varepsilon}(x) = \begin{cases} \beta x & \text{for } 0 \le x < \frac{1}{2} - \varepsilon \\ \beta\left(\frac{1-4\varepsilon}{4\varepsilon}\right)\left(\frac{1}{2} - x\right) + \frac{\beta}{4} & \text{for } \frac{1}{2} - \varepsilon \le x < \frac{1}{2} + \varepsilon \\ \beta\left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} + \varepsilon \le x \le 1 \end{cases}$$

This simply connects the two endpoints in the middle of the downshift, "widening" it so that the has finite, not infinite slope, and converting the iterated function from a discontinuous to a continuous one. In the limit, one regains the earlier downshift: $\lim_{\varepsilon\to 0} T_{\beta,\varepsilon} = T_\beta$, as the slope of the middle bit becomes infinite. For finite ε, however, one gains a critical point in the region $\frac{1}{2} - \varepsilon \le x < \frac{1}{2} + \varepsilon$. To arrange for a critical point at a specifically chosen trouble spot, one need only choose an ε small enough, such that $\varepsilon < \left|m_k - \frac{1}{2}\right|$ for each $k < p$ prior to the trouble-spot $m_p = m_0$.

To summarize: the "trouble spots" don't "just" break the ability to create a Hessenberg basis at certain values of β: they are more "fundamental" than that: they indicate the regions where ("phase-locked") periodic orbits can be made to appear.

Note to reader: I presume that the above observations are generically "well known", and presented in some pop-lit on fractals, but I am not aware of any references discussing this topic. If you, dear reader, know of such references, please drop me a line at the posted email address.

Exercise left to the reader: the above arguments should be sufficient to fully demonstrate that the circle map, which is well-known to exhibit phase locking regions called Arnold tongues, is topologically conjugate to the fattened downshift $T_{\beta,\varepsilon}$. Or something like that. In a certain sense, this can be argued to be a "complete" solution, via topological conjugacy, of the tent map, the logistic map and the circle map. This is a worthwhile exercise to actually perform, i.e. to give explicit expressions mapping the various regions, as appropriate. XXX Perhaps I can get around to this, someday, eh?

4.4 Matrix Elements

The above-defined basis provides the Hessenberg representation for the transfer operator. Defining

$$\left\langle n \left| \mathscr{L}_\beta \right| m \right\rangle = \int_0^1 \psi_n(x) \left[\mathscr{L}_\beta \psi_m\right](x)\, dx \tag{17}$$

this has the expected Hessenberg form, in that

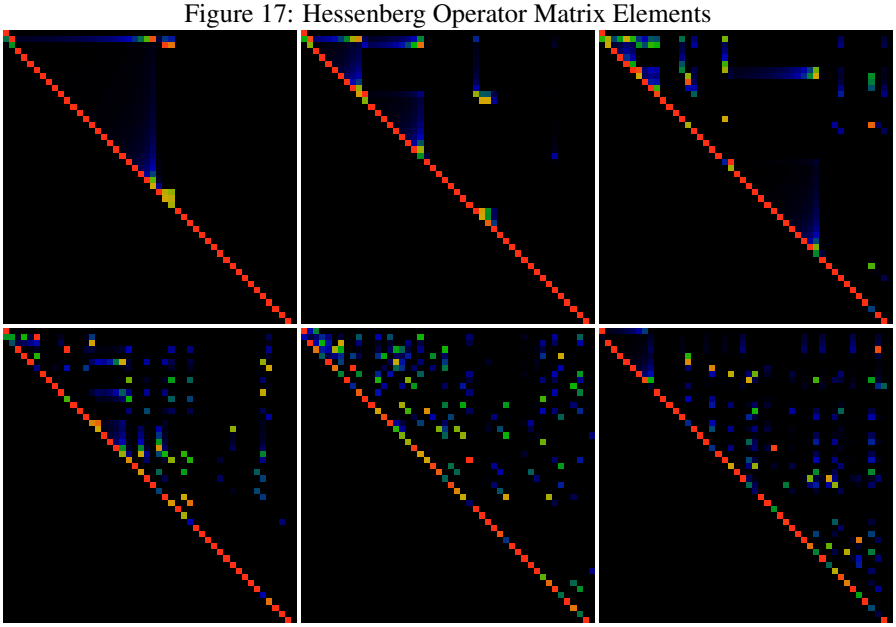
$$\left\langle n \left| \mathscr{L}_\beta \right| m \right\rangle = 0 \quad \text{for } n > m+1$$

This is just one diagonal short of being actually solvable. A visualization of the matrix elements is shown in figure 17.

4.5 Completeness

The Hessenberg basis construction gives a countable set of ψ_n that is an orthonormal basis on the unit interval: $\int_0^1 \psi_m(x)\,\psi_n(x)\,dx = \delta_{mn}$. Are they complete? Obviously

Figure 17: Hessenberg Operator Matrix Elements



Six illustrations of the absolute value of the matrix elements $\langle n | \mathscr{L}_\beta | m \rangle$ for the transfer operator $\mathscr{L}_\beta$ for (left to right, top to bottom) β =1.1, 1.2, 1.3, 1.6, 1.90, 1.998 and $0 \le n,m < 48$ in the Hessenberg basis. The red color represents values of 0.66 or larger, green represents values of 0.33 and blue and darker correspond to 0.16 or less. Almost all matrix elements are in fact precisely zero; black pixels in these images correspond to matrix elements that are zero. Note that the almost all of the diagonal matrix elements are exactly zero: that is $\langle n | \mathscr{L}_\beta | n \rangle = 0$ for most n. The bright-red pixels are just below the diagonal: for most n, one has that $\langle n+1 | \mathscr{L}_\beta | n \rangle \gtrsim 0.5$ with the occasional blueish pixel suggesting a smaller value. These two, taken together, suggests that the eigenvalue spectrum is rapidly decreasing. The first few images suggests a regular pattern that gets increasingly compressed and chaotic as β increases. More-or-less the sane structure previals if one instead zooms out to look at the 600x600 submatrix; animating with fine-grained steps in β does not result in an interesting animated movie.

not when β is equal to a "trouble spot", as discussed above. But the trouble spots, although dense in the interval $1 < \beta < 2$, are countable and so "most" values of β are not problematic. So what about these other values?

Obviously the $\{\psi_n\}$ cannot be complete on the unit interval, as they all vanish for $\beta/2 < x$. Perhaps they are complete on the interval $[0, \beta/2]$, where they are already orthonormal: $\int_0^{\beta/2} \psi_m(x)\, \psi_n(x)\, dx = \delta_{mn}$.

Clearly, the ψ_n span some subspace; do they span the Hilbert space $L_2[0, \beta/2]$ of square-integrable functions on the interval $[0, \beta/2]$? To what degree can one legitimately write

$$\delta(y-x) = \sum_{n=0}^{\infty} \psi_n(y)\, \psi_n(x)$$

as a resolution of the identity? A numerical exploration shows that the midpoints m_p are dense in the interval $(0, \beta/2)$, and so this suggests that the basis should be considered to be "sufficiently complete" on the interval $[0, \beta/2]$. The distribution of the m_p follow exactly the distribution of the invariant measure. Convergence is uniform to the same degree that the midpoints "eventually" fill in and become dense in some interval.

The question of the completeness of states dogs some "obvious" assumptions one wants to make. For example, if the set of states is complete, and the resolution of the identity hiolds, then one expects that the transfer operator resolves to the iterated function:

$$\delta(y - (\beta x \mod 1)) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \psi_n(y) \left\langle n \left| \mathscr{L}_\beta \right| m \right\rangle \psi_m(x)$$

It is fun to verify that the world works as one expects it to work: the above can be verified to hold numerically, for sums limited to a finite cutoff.

4.6 Numerical Eigenvalues

Given the apparent sparsity visible in figure 17, one might think that the eigenvalue problem is fairly stable, numerically. It is not all that much. Numerical exploration suggests that the spectrum is a circle lying in the complex plane[1], of radius $|\lambda| = 1/\beta$ (ignoring, that is, the leading eigenvalue of 1, which is easily found).

To be clear, this is a numerical exploration of the $N \times N$ principle submatrix of $\left\langle n \left| \mathscr{L}_\beta \right| m \right\rangle$. The eigenvalue problem being posed is to find a vector $\vec{v} = (v_k)_{k=0}^N$ that solves

$$\sum_{m=0}^{N} \left\langle n \left| \mathscr{L}_\beta \right| m \right\rangle v_m = \lambda v_n$$

for some constant λ (with the set of possible λ depending on N, of course).

[1]This was confirmed with both GSL gsl_eigen_nonsymmv() and Lapack DHSEQR solvers, exploring the principle sumatrix of various sizes, up to about 2000×2000 entries. Both systems agree to at least six decimal places, if not more. Both show sporadic eigenvalues off the circle, but these are not numerically stable; ergo, the only valid eigenvalues are thos on the circle. The matrix entries were constructed using the midpoint aglorithm, described in the last section. To verify that they are accurate, several techiques were used: numerical integration to confirm orthogonality, and the use of the GnuMP multi-precision library to push up accuracy.

There are various pitfalls in extrapolating from this to the $N \to \infty$ limit. For the next few paragraphs, consider only some notion of a "minimal" extension from finite N to the limit. That is, for each finite N, one has a finite set of eigenvalues and eigenvectors. In the limit, there may be accumulation points: points where the eigenvalues accumulate to a limit point, in a standard topological sense. What should that that topological space be? For finite N, all eigenvectors are explicity summable, and thus can be taken to belong to any Banach space ℓ_p. One may as well take $p = 2$ the Hilbert space, and normalize the eigenvectors $\vec{v}$ so that $1 = \sum_{m=0}^{N} v_m^2$.

For finite N, it appears that most eigenvalues λ are "near" the circle $|\lambda| = 1/\beta$, and that they seem to be very uniformly distributed around this circle. The numerical results indicate that in hte limit $N \to \infty$, that "most" becomes "almost all" in the usual sense. Similarly, "near" appears to mean that for any given λ at finite N , one has that $|\lambda| - 1/\beta \sim \mathcal{O}(1/N)$. As to uniformity, it seems that the spacing between nearest neighbors is also $\mathcal{O}(1/N)$, and that there are no "premature" accumulation points: egenvalues never get any closer than $\mathcal{O}(1/N)$, either.

Thus, the minimal closure, the minimal extrapolation to limit points strongly suggests that the limit points really do lie, uniformly distributed, on the circle $|\lambda| = 1/\beta$. Then, writing a given accumulation point as $\lambda = \beta^{-1} \exp 2\pi i\phi$, what the numerics do not reveal, or, at least, do not easily reveal, is whether the allowed values of ϕ are always rational, irrational or might assume arbitrary real values. The numerical exploration does suggest that the eigenvalues are dense on the circle. Certainly it is the case Hessenberg basis is countable, an so one would expect the eigenvalue spectrum obtained in this way to be at least countable, as well. Whether it is also uncountable seems unknownble in this naive sense.

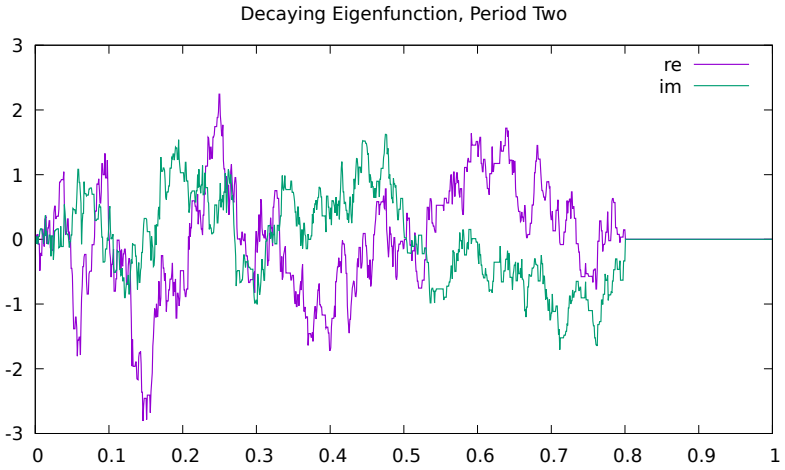
This question is interesting because if only rational ϕ are allowed, then the decaying eigenfunctions belong to a cyclic group, and exhibit an exact form of Poincaré recurrence as they decay. If irrational ϕ are allowed, then the decaying eigenfunctions are more throughly chaotic.

For $\beta = 2$, the β-transform is the Bernoulli shift, the transfer operator is solvable, and the spectrum is exactly known. This has been explored by various authors[13]. I've written extensively about this spectrum and the eigenvalues in other texts[14, 15, 16]. To recap, it takes several forms, depending on the function space that one chooses to work in. If one restricts oneself to polynomial eigenfunctions, then the spectrum is real, non-negative (it has an extensive kernel) and has eigenvalues of 2^{-n} for all n. The eigenfunctions are the Bernoulli polynomials. Restricting to square-integrable eigenfunctions, the spectrum continuous, having eigenvalues on the unit disk in the complex plane. The continuous-spectrum eigenfunctions (for eigenvalues other than 2^{-n}) can be understood in several ways: if forced to be differentiable, then they are not bounded (they diverge) at the endpoints of the interval. If forced to be bounded, then they are fractal (non-smooth) over the entire interval. The unitary spectrum corresponds to differentiable-nowhere eigenfunctions (wait, or continuous-nowhere? I forget.)

A pair of plausible, numerically-extracted eigenfunctions are shown in image 18.

Presumably, the spectrum can be related to the lap-counting function, given by Lagarias[7].

Figure 18: Decaying Eigenfunction, Period Two



This shows a numerically-computed decaying eigenfunction of period two, for $\beta = 1.6$. It is period two, in that it corresponds to an eigenvalue of $\lambda = -1/\beta = -0.625$, so that after one iteration of $\mathscr{L}_\beta$, the sign flips. This can be confirmed, numerically: after one iteration, the sign really does flip, to within numerical errors. This was computed by numerically diagonalizing the 861×861 matrix given by the lowest terms of eqn 17, and then graphing the eigenvector closes to $\lambda = -0.625$ (The GnuMP library was used to provide the required level of precision in the calculations.)

Although this figure is drawn with curves labeled "real" and "imaginary", this is a bit fantastic, and is a numeric artifact. For any period-two eigenfunction, the real and imaginary parts would have no coupling, and would be independent of each other; either one could be set to zero and one would still have a valid eigenfunction. This differs from the case of period-three and higher, where the real and imaginary parts are expected to mix. (Nor are the two components orthogonal, as one might expect.) The eigenfunction is also fantastic in that only slightly different numerics result in a completely different eigenfunction being computed. Even the functions resulting from diagonlizing the 863×863 matrix differ fair amount from those arising from the 861×861 matrix; there's only a general resemblance. This is not entirely surprising: the magnitude of the basis coefficients decays very slowly; even at 861, that are still on the order of 10^{-3}, and thus contribute strongly.

Computed eigenfunctions for period-three are not dissimilar, nor are the ones for other values of β. They do seem to start having the general oscillatory character of $\sin(1/x)$ as $\beta \to 1$, but its not clear if this is a numeric artifact, or something real. The wildeness of these functions contrast sharply with the seemingly tame $\lambda = 1$ eigenfunctions shown in figure 1.

4.7 (Non-)Unitarity

The numerical results suggest a hypothesis that some fragment of $\mathscr{L}_\beta$ is unitary, as it is ordinarily the case that when eigenvalues appear on the unit circle, its because the operator is unitary. That does not seem to be the case here. Specifically, define the Frobenius-Perron eigenvector ρ as the one satisfying $\mathscr{L}_\beta \rho = \rho$ and normalizing it to unit length, so that $\|\rho\| = 1$ in the Hilbert (mean-square) norm. Define the reduced operator $\mathscr{R}_\beta$ in terms of the matrix elements

$$\frac{1}{\beta} \langle n | \mathscr{R}_\beta | m \rangle = \langle n | \mathscr{L}_\beta | m \rangle - \langle \rho | n \rangle \langle \rho | m \rangle$$

That is, it is just the downshift operator, with the Frobenius-Perron eigenvector removed, so that $\mathscr{R}_\beta \rho = 0$. Its rescaled, so that the remaining eigenvectors of $\mathscr{R}_\beta$ lie on the unit circle. Is this operator unitary in any way? That is, might either $\mathscr{R}_\beta \mathscr{R}_\beta^\dagger$ or $\mathscr{R}_\beta^\dagger \mathscr{R}_\beta$ be the identity? Here, the dagger $\dagger$ is just the transpose, as $\mathscr{R}_\beta$ is purely real. Numerical exploration clearly shows that $\mathscr{R}_\beta$ is neither unitary on the left nor on the right. Not a surprise, but does leave the understanding of $\mathscr{L}_\beta$ in a curious situation.

4.8 Generating Function

Let v_n be the Ruelle-Frobenius-Perron eigenvector in the Hessenberg basis. That is, let v_n be the vector that solves

$$\sum_{m=0}^{\infty} \langle n | \mathscr{L}_\beta | m \rangle v_m = v_n \tag{18}$$

This is readily computed numerically, and it is straightforward to verify the numerics by confirming that

$$\rho(x) = \sum_{m=0}^{\infty} v_m \psi_m(x)$$

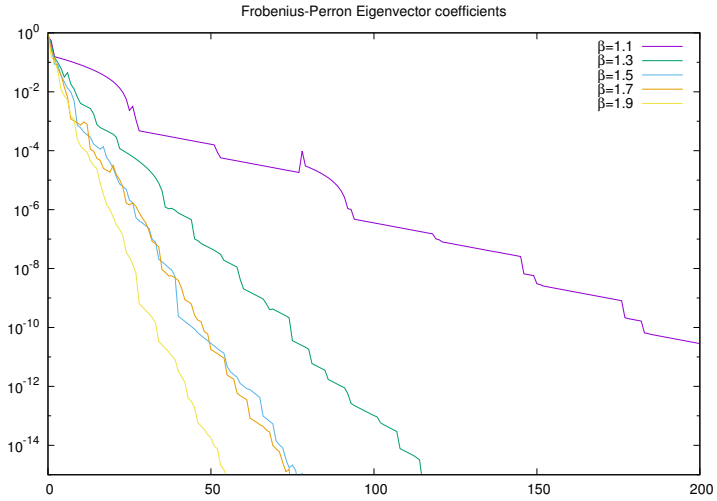
is the invariant measure of equations 12,13. The truncated ordinary generating function associated with this eigenvector is

$$G_N(z) = \sum_{m=0}^{N} v_m z^m$$

with the ordinary generating function being the limit $N \to \infty$. A numerical study of this function indicates that most of the N zeroes of G_N are arranged approximately on a circle of radius β. The arrangement appear to be quite uniform, with more-or-less equidistant spacing of the zeros. As N increases, it seems that more of the zeroes get closer to the circle, although the numerical instabilities associated with double-precision math make this difficult to control; an arbitrary-precision eigenvalue solver would be needed to confirm this behavior.

If this behavior persists, then the limit $N \to \infty$ cannot really be taken, and the ordinary generating function doesn't "really exist", per-se. The figure 19 depicts the values of v_n as a function of n for selected values of β. Note that these are all real and positive.

Figure 19: Frobenius-Perron Eigenvector Coefficients



The coefficients v_n solving eqn 18 as a function of n, for various values of β. Note that the coefficeints are all real and positive. When $\beta > \left(1+\sqrt{5}\right)/2$ (the Golden ratio), the general but rather loose trend for the slopes seems to be about $\log_2 v_n \sim \mathcal{O}((1-\beta)n)$. For β less than the Golden ratio, the falloff is sharper. The Golden ratio seems to appear unexpected and unexplainedly in the proceedings; later sections give a firmer standing for this. In this figure, the two slopes for $\beta = 1.5$ and $\beta = 1.7$ unexpectedly overlap. Apperently, this is not an accident or a numerical defect.

4.9 Bergman Space

Given a matrix operator in Hessenberg form, it can be interpreted as a right-shift on the space of polynomials. That is, given an unreduced Hessenberg matrix with matrix entries A_{ij}, one can write a recurrance relation that defines a sequence of polynomials as

$$zp_n(z) = \sum_{k=0}^{n+1} A_{kn} p_k(z) \tag{19}$$

with $p_0(z) = 1$. This relation is easily solvable in closed form, as the recurrance relation terminates in a finite number of steps.

One important property of these polynomials is that the zeros of $p_n(z)$ correspond to the eigenvalues of the $n \times n$ principle submatrix of A. Numeric exploration of these polynomials confirms the previous results on eigenvalues obtained from direct diagnoalization: the zeros of the $p_n(z)$ seems to lie mostly near the circle of radius $1/\beta$, distributed uniformly over all angles.

If all of the subdiagonal entries obey $A_{n+1,n} > 0$, then the polynomials form an orthonormal basis for Bergman space. That is, there exists a domain in the complex plane on which the polynomials provide a basis for a Hilbert space of holomorphic functions on that domain[17, 18, 19]. That is, one has the orthogonality relation

$$\delta_{mn} = \int_D p_m(\bar{z})\, p_n(z)\, d\mu(z)$$

for some domain $D \subset \mathbb{C}$ of the complex plane, and some (Borel) measure $d\mu$ on that domain.

The matrix A can be interpreted as an operator with a continuous spectrum. To do this, fix a certain, specific value of $z = c$ a constant, and then notice that $\vec{p} = (p_n(z))_{n=0}^{\infty}$ is a vector having the property that $A^T \vec{p} = z\vec{p}$. That is, $\vec{p}$ is a left-eigenvector of A; equivalently, a right-eigenvector of its transpose A^T. Clearly, the spectrum is continuous on the domain D.

The matrix operator A can also be interpreted as a right-shift on Bergman space. To do this, define

$$\mathscr{A}(w,z) = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_k(w) A_{kn} p_n(\bar{z})$$

Then, given some holomorphic function $f(z)$ decomposed in terms of the polynomials, so that $f(z) = \sum_n a_n p_n(z)$, one has that

$$\begin{aligned}
[\mathscr{A} f](w) &= \int \mathscr{A}(w,z)\, f(z)\, d\mu(z) \\
&= \sum_k \sum_n p_k(w) A_{kn} a_n \\
&= w \sum_n a_n p_n(w) \\
&= w f(w)
\end{aligned}$$

That is, given a sequence $(a_0, a_1, a_2, \cdots)$, the Hessenberg matrix acts as a right-shift, mapping it to the sequence $(0, a_0, a_1, \cdots)$.

This is perhaps a bit silly, as one could instead just perform the same manipulation without the $f(z)$, by observing that, formally,

$$\mathscr{A}(w,z) = w \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_k(w)\, p_n(\bar{z})$$

The above treatment is breezy and "formal", paying no heed to summability, convergence or responding to any questions about what spaces the various vectors may live in. This is as appropriate, since the task here is to discover which spaces are the appropriate ones, when the Hessenberg matrix arises from the downshift.

Notice that the word "operator" is a bit mis-used, here, as a vague synonym for "infinite-dimensional matrix". Properly, the word "operator" should be reserved for an infinite-dimensional matrix acting on some given space, having general properties that are independent of the basis chosen for that space. So far, that might not be the case here: the infinite-dimensional matrixes here might not be bounded operators; they might not even be continuous, viz. we have not ruled out the possibility that the space of interest is some Fréchet space or some more general topological vector space. It is well known that operators on such spaces can have "unexpected" discontinuities, unexpcected in that they are not seen in ordinary Banach spaces.

At any rate, if polynomials obtained from the downshift are orthogonal on some domain $D \subset \mathbb{C}$ that is the support of some measure $d\mu$, it is not at all clear what this measure might be. They are certainly not orthogonal on the unit disk, with uniform measure.

Notice also that the above treatment seems to be a special case of a more general principle: when an operator has a continuous spectrum, it can sometimes be interpreted as a right-shift. That is, given some arbitrary operator $\mathscr{H}$, then if one has that $\mathscr{H} f = \lambda f$ and λ takes arbitrary values $\lambda \in D \subset \mathbb{C}$, then $\mathscr{H}$ can be taken to be a right-shift opertator, provided that $f = f(\lambda)$ can be decomposed into a set of orthogonal polynomials in λ.

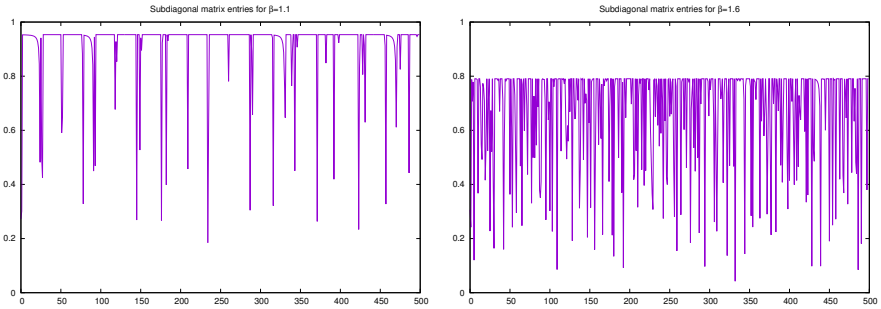
4.10 Beta Bergman Shift

The primary question for this section is whether the β-transform transfer operator, in the Hessenberg basis, can be considered to be a Bergman shift.

To obtain the orthogonal polynomial basis, one must satisfy the constraint that $A_{n+1,n} > 0$ for the matrix elements $A_{kn} = \langle k | \mathscr{L}_\beta | n \rangle$ of eqn 17. Numeric exploration indicates that this is indeed the case, with the subdiagonal entries all positive (none are zero), and all tend to have the same value, with sporadic exceptions. These are shown in figure 20.

Can one find a domain on the complex plane that would have such Bergman polynomials? The references[17, 19] provide a technique for doing so, provided that the matrix is asymptotically Toeplitz. That is, if the diagonals of A_{ij} have nice limits, that $\lim_{n\to\infty} A_{n-k,n}$ exists for fixed k, then a Jordan arc bounding a domain on the complex plane can be found. The figure 20 indicates that this limit does not exist, in the strict sense: the values bounce away from an obvious limit point indefinitely. Exactly what

Figure 20: Subdiagonal Entries



These charts show the subdiagonal matrix entries $\langle n+1 | \mathscr{L}_\beta | n \rangle$ for the first $n < 500$. Representative values of β are as labelled; other values behave similarly. The relation of the supremum to β is not clear.

---

this implies is unclear. Perhaps it is possible to extend the results of [17, 19] to matrixes that arewhere the diagonals merely have an accumulation point, as opposed to a well-defined limit?

Based on numeric exploration, it appears that the domain is the unit disk. That is, $A^T \vec{p} = z\vec{p}$ holds for $|z| \leq 1$.

TODO: Graph figure20 in the usual gloriously-colored-3D style.

4.11 Bergman Alternative

The Bergman polynomials of eqn 19 define an orthonormal basis for some region of the complex plane. For the square-integrable norm, this basis is the basis of a Hilbert space, and specifically, that of a reproducing kernel Hilbert space.

Yet, something funny happens on the unit disk. Let $p_m(z)$ be the polynomials, and for some sequence of coefficients $\{a_n\}$, consider a generic function

$$f(z) = \sum_{k=0}^{\infty} a_k p_k(z)$$

Consider the case where the $\{a_n\}$ are a right-eigenvector of the Hessenberg operator, that is, where

$$\sum_{m=0}^{\infty} A_{km} a_m = \lambda a_k$$

Substituting into the above, one has

$$f(z) = \sum_{k=0}^{\infty} \frac{1}{\lambda} \sum_{m=0}^{\infty} A_{km} a_m p_k(z) = \frac{z}{\lambda} \sum_{m=0}^{\infty} a_m p_m(z) = \frac{z f(z)}{\lambda}$$

There are two alternatives to solving this; either $f(z) = 0$ or $z = \lambda$. Since this is a reproducing kernel Hilbert space, then if $z = \lambda$ is part of the domain of the Bergman

space, then one must conclude that $f(z) = 0$ everywhere. That is, right-eigenvalues of A correspond to functions $f(z)$ that are vanishing. To invent a new name, by analogy to the Fredholm alternative, perhaps this can be called the Bergman alternative.

Numerical exploration indicates that, for the matrix elements of eqn, 17, the function $f(z)$ vanishes inside the unit disk $|z| < 1$, and is undefined (infinite) outside of it.

4.12 Left Factorization

Suppose one is given an (arbitrary) sequence of polynomials $(p_n(z))_{n=0}^{\infty}$, such that the order of p_n is n. Then each individial polynomial can be expanded as or $\beta > \varphi$,

$$p_n(z) = \sum_{k=0}^{n} p_{nk} z^k$$

This defines an infinite matrix $\mathscr{P} = [p_{nk}]$, provided that the coefficients are extended so that $p_{nk} = 0$ whenver $k > n$. This matrix is manifestly lower-triangular. Writing vectors $\vec{z} = (z^n)_{n=0}^{\infty}$ and $\vec{p} = (p_n(z))_{n=0}^{\infty}$ as before, the above is just the matrix equation

$$\vec{p} = \mathscr{P}\vec{z}$$

Consider now the case where the polynomials were constructed from some irreducible Hessenberg matrix A. The earlier observation that A^T is a shift, namely, that $A^T\vec{p} = z\vec{p}$ can now be written as

$$A^T \mathscr{P}\vec{z} = z\mathscr{P}\vec{z} = \mathscr{P}z\vec{z} = \mathscr{P}\mathscr{K}\vec{z}$$

In the above, the z without the vector notation is just a scalar, and thus commutes (trivially) with $\mathscr{P}$. Its eliminated by explicitly making use of the right-shift (Koopman) operator, which, in this basis, is

$$\mathscr{K} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \ddots \\ 0 & 0 & 0 & 0 & \ddots \end{bmatrix}$$

Since $\mathscr{P}$ is lower-triangular, it is invertible on the right, that is, the inverse $\mathscr{P}^{-1}$ exists, and so one is left with

$$\mathscr{P}^{-1} A^T \mathscr{P} = \mathscr{K}$$

The irreducibility of A is important, here; non-zero entries on the subdiagonal are required, else trouble ensues.

Rearranging, this provides an explicit decomposition of A into triangular matrixes:

$$A^T = \mathscr{P}\mathscr{K}\mathscr{P}^{-1}$$

Taking the transpose, this gives

$$A = \left[\mathscr{P}^{-1}\right]^T \mathscr{K}^T \mathscr{P}^T$$

with $\mathscr{P}^T$ and $\left[\mathscr{P}^{-1}\right]^T$ both being upper-trangular, and $\mathscr{K}^T$ being the left-shift.

This system is solvable. Given some matrix A in Hessenberg form, the matrix elements of $\mathscr{P}$ can be computed resursively, in a finite number of steps (i.e. in closed form), directly from 19. The explicit expression is

$$A_{n+1,n} p_{n+1,j} = p_{n,j-1} - \sum_{k=0}^{n} A_{kn} p_{kj}$$

The starting conditions are $p_{00} = 1$. To handle the $j = 0$ case in the above, set $p_{n,-1} = 0$.

Because $\mathscr{P}$ is lower triangular, its inverse $\mathscr{P}^{-1} \equiv \mathscr{R} = [r_{kn}]$ can be obtained explicitly. Along the digaonal, one has $r_{nn} = 1/p_{nn}$ while the lower triangular form means $r_{kn} = 0$ for $k < n$. For the remaining entries $m < n$, one has

$$0 = \sum_{k=m}^{n} p_{nk} r_{km}$$

This can be solved in a finite number of iterations on

$$p_{nn} r_{nm} = -\sum_{k=m}^{n-1} p_{nk} r_{km}$$

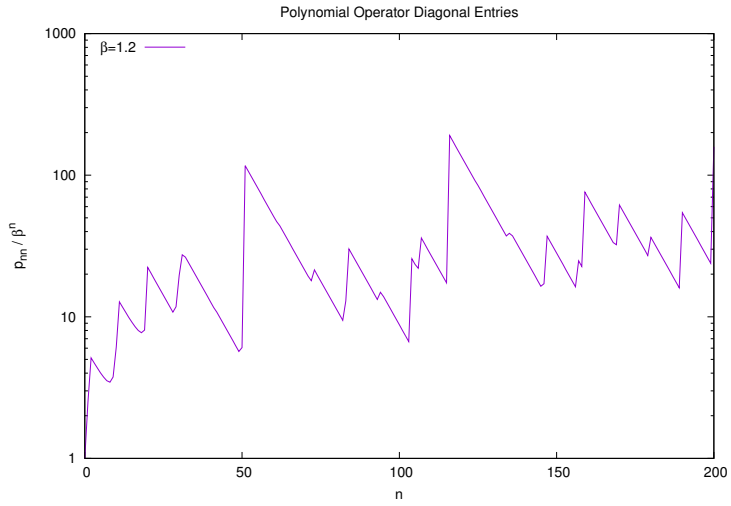
The above avoids questions of convergence, or any notion of the spaces on which the matrixes or operators might act. The norm to be used for $\vec{z}$ and $\vec{p}$ is not specified. This is appropriate at this stage: it is the algebraic manipulations that are interesting, at this point, rather than the spaces on which the matrixes/operators might act. One can invent several kinds of norms that might be applicable, but there is no particular reason to beleive that $\vec{p}$ might have a finite norm. Likewise, $\mathscr{P}$ may not have a finite norm. For the case of the Hessenberg operator originating with the downshift operator, it does not; the individual matrix elements p_{nm} increase without bound. That is, $\mathscr{P}$ is an infinite matrix, but it is not clear that it is also an operator. If it is, it is certainly not a compact operator.

Some of the poor behavior can be brought under control by factoring $\mathscr{P} = \mathscr{D}\mathscr{N}$ with $\mathscr{N}$ being unitriangular (all ones on the diagonal) and $\mathscr{D}$ a diagonal matrix, with entries $[\mathscr{D}]_{nk} = p_{nn}\delta_{nk}$. With this factorization, one may then write

$$\mathscr{N}^{-1} A^T \mathscr{N} = \mathscr{D}\mathscr{K}\mathscr{D}^{-1}$$

so that $\mathscr{D}\mathscr{K}\mathscr{D}^{-1}$ has off-diagonal matrix entries $\left[\mathscr{D}\mathscr{K}\mathscr{D}^{-1}\right]_{nk} = \delta_{n+1,k} p_{nn}/p_{kk}$. This is a rescaling of the shift $[\mathscr{K}]_{nk} = \delta_{n+1,k}$. The scaling factor is exactly the sub-diagonal of the Hessenberg. That is, $p_{nn}/p_{n+1,n+1} = A_{n+1,n}$. The polynomials $\mathscr{N}\vec{z}$ are monic.

Figure 21: Polynomial Operator Diagonal Entries



This depicts the ratio p_{nn}/β^n of the diagonal matrix entries p_{nn} of the Bergman polynomial matrix operator $\mathscr{P}$ for the downshift with value $\beta = 1.2$. Other values of β are not dissimilar, although the spikes are pushed more closely together. The height of the spikes seems to be roughly the same, for all β. This is another way of visualizaing the same information as in figure 20, as the ratio $p_{nn}/p_{n+1,n+1}$ is just given by the subdiagonal entries $A_{n+1,n}$ of the Hessenberg matrix. In particular, the straight edges correspond to usually-constant values on the subdiagonal.

4.13 Beta-transform factoids

An assortment of observations follow, for the case of the downshift.

First, the matrix entries of $\mathscr{P}$ grow in an unbounded fashion. It appears that $p_{nn} \sim \mathscr{O}(\beta^n)$; the ratio p_{nn}/β^n is depicted in figure 21.

Experimentation reveals two different regimes of behavior, depending on whether or not $\beta < \varphi = \left(1+\sqrt{5}\right)/2$ the Golden ratio. Exactly why there are two different regimes is unclear. Earlier sections motivated the reason for the appearance of the golden mean; why this shows up dramatically, as it does here, is unclear (to me).

One such result is that when $\beta < \varphi$, then the sum over columns of the Bergman operator vanishes. That is,

$$\sum_{k=0}^{\infty} p_{nk} = \delta_{n0}$$

This implies that every polynomial $p_n(z)$ has a zero at $z = 1$ (except for $p_0(z) = 1$) when $\beta < \varphi$.

TODO: Graph the figure 21 in the usual gloriously-colored-3D style.

4.14 Decaying Eigenfunctions

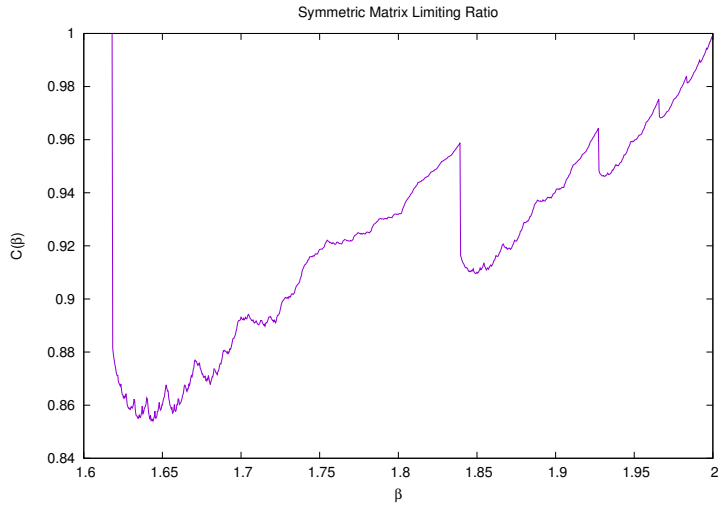
The matrix mechanics developed in the previous sections can be used to perform asymptotic expansions that rapidly converge to decaying eigenfunctions. This works most simply for the case of $\varphi < \beta$. TODO Write these down. TODO flesh out. Basically, write a vector $\vec{w}$ with elements $w_n = \omega^n$ for $1 < |\omega|$ so that this is divergent. Then write the formal vector $\vec{a} = \left[\mathscr{P}^T\right]^{-1}\vec{w}$ which is formally divergent, but can be trancated in finite dimensions, and renormalized to be of unit length. Doing so provides an eigenfunction of A. The associated eigenvalue is 1 when $\beta < \varphi$ but is less than 1 when $\varphi < \beta$ (and in fact, the eigenvalue is exactly that depicted in figure 22). TODO graph some of these, explore more thoroughly, adress the issues of formal divergence.

4.15 Moment Matrix

When the Hessenberg matrix is derived from measures on the complex plane, it takes the form of $\mathscr{M} = \mathscr{R}\mathscr{R}^T$ with $\mathscr{R} = \mathscr{P}^{-1}$, so that $\mathscr{R}$ is the Cholesky decompostion of $\mathscr{M}$. This matrix is manifestly symmetric: $\mathscr{M} = \mathscr{M}^T$. Direct observation shows that it is almost positive-definite: one finds that $[\mathscr{M}]_{ij} > 0$ for all i, j except for $[\mathscr{M}]_{00} = 0$. This result can be strengthened: when $\beta < \varphi$, then $[\mathscr{M}]_{ij} > 1$ for all i, j except for $[\mathscr{M}]_{00} = 0$ and $[\mathscr{M}]_{0n} = [\mathscr{M}]_{n0} = 1$. But, for $\beta > \varphi$, one finds that $[\mathscr{M}]_{00} = 0$ and $[\mathscr{M}]_{01} = [\mathscr{M}]_{10} = [\mathscr{M}]_{11} = 1$, while all the rest obey $0 < [\mathscr{M}]_{ij} < 1$.

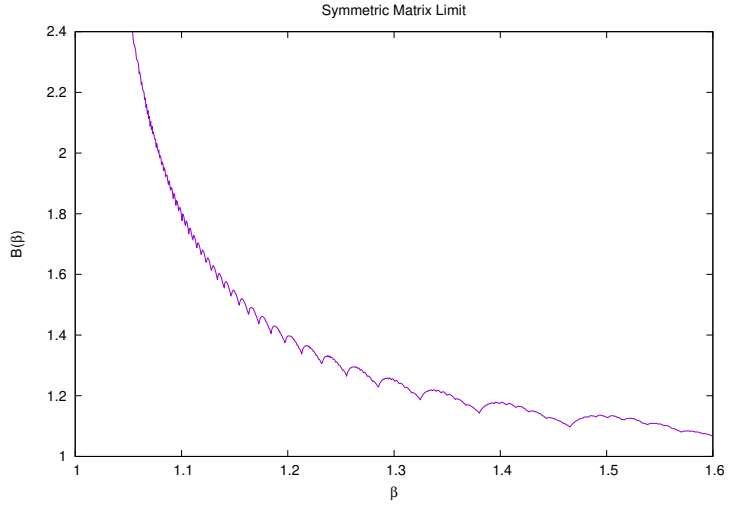
In the standard literature, $\mathscr{M}$ is usually obtained from some moment matrix, viz, for the integral $\int \bar{z}^m z^n d\mu(z)$ for some measure $d\mu(z)$. Might that be the case here? Taking the time to numerically characterize the matrix, one finds that the ratio of successive rows (or columns as its symmetric) very quickly approaches a limit $\lim_{n\to\infty,} [\mathscr{M}]_{nm} / [\mathscr{M}]_{n-1,m} = C(\beta)$ for some constant C that depends ony on β but not on m. The limit $C(\beta)$ is graphed in figure 22.

Figure 22: Symmetric Matrix Limit Ratio



This figure shows the limit $C(\beta)$ defined in the text. Note that $C(\beta) = 1$ for $\beta < \varphi$. The jump is at about $\kappa = 1.83928676\cdots$. Note this is one of the "troublesome midpoints" for the Hessenberg basis expansion: that $T_\kappa^3(\kappa/2) = 0$ or $\kappa/2$, that is, $\kappa^2(\kappa-1) - \kappa - 1 = 0$. The remaining steps presumably correspond to higher iterates p that satisfy $T_\beta^p(\beta/2) = 0$.

Figure 23: Symmetric Matrix Limit



This figure shows the limit $B(\beta)$ defined in the text. The limit is approched fairly quickly for the larger values of β, but convergence proves difficult for $\beta \lesssim 1.1$. The overall shape is that of a hyperbola, but doesn't seem to actually be hyperbolic for either small or large β. The right-most nick in the curve appears to be at $\beta = 1.465571231876768\cdots$ which solves $\beta^2(\beta-1)-1=0$; that is, it occurs at $T_\beta^3(\beta/2)=0$. The remaining nicks are presumably located at $T_\beta^p(\beta/2)=0$ for higher iterates p.

For $\beta < \varphi$, it appears that $\lim_{n\to\infty,} [\mathscr{M}]_{nm} = B(\beta)$ a constant, independent of m. This limiting value $B(\beta)$ is graphed in figure 23.

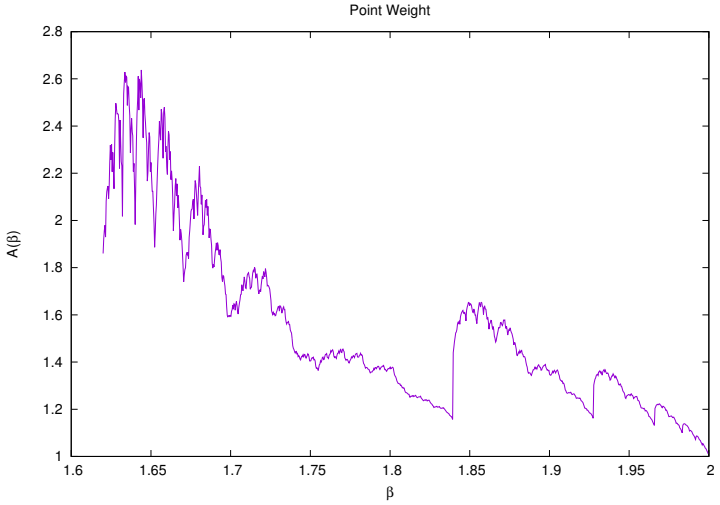
The asymptotic behavior of the matrix $[\mathscr{M}]_{ij}$ can be obtained as a momement matrix on point sources. A delta function located at $z = C$ for real C has the moments

$$\begin{aligned} C_{mn} &= \int \overline{z}^m z^n \delta(z-C)\, dz \\ &= \int r^m r^n \delta(r-C)\, r dr \int \delta(\theta)\, e^{-im\theta} e^{in\theta} d\theta \\ &= C^{m+n+1} \end{aligned}$$

Thus, for $\varphi < \beta$, the asymptotic behavior of $[\mathscr{M}]_{ij}$ is given by the distribution $A(\beta)\,\delta(z-C(\beta))$. What is $A(\beta)$? This is graphed in figure 24.

What about $\beta < \varphi$? A limiting constant distribution can be obtained from a deriva-

Figure 24: Point Weight



This figure shows the value of $A(\beta)$ that gives the point weight of the moment matrix. That is, the asymptotic behavior of $\mathscr{M}$ is given by $[\mathscr{M}]_{mn} \to \int \overline{z}^m z^n \rho(z)\, dz$ with the measure given by a point mass $\rho(z) = A(\beta)\,\delta(z - C(\beta))$. Clearly, there is a strong resemblance to figure 22.

---

tive point mass located at $z = 1$. That is,

$$\begin{aligned} D_{mn} &= \int \overline{z}^m z^n \delta'(z-1)\, dz \\ &= \int r^m r^n \delta'(r-1)\, r dr \int \delta(\theta)\, e^{-im\theta} e^{in\theta} d\theta \\ &= 1 \end{aligned}$$

so that the asymptotic behavior of $[\mathscr{M}]_{ij}$ for $\beta < \varphi$ is given by the distribution $B(\beta)\,\delta'(z-1)$. The prime superscript here means derivative, viz, in colloqial language, $\delta'(z) = d\delta(z)/dz$.

4.16 Givens rotations

A Hessenberg matrix can be brought to solvable form by applying a sequence of Givens rotations. Is the seuqnce of angles that appear in these rotations meaningful in any way, or are they just some form of uninteresting junk?

5 The Jacobi Operator

Given a Borel measure on the real number line, one can find a sequence of polynomials that are orthonormal with respect to that measure. These polynomials $p_n(x)$ are

coubled together by a three-term recurrence equation

$$xp_n(x) = a_{n+1}p_{n+1}(x) + b_n p_n(x) + a_n p_{n-1}(x)$$

with $p_0(x) = 1$ and $p_{-1}(x) = 0$. This recurrence relation can be taken to be an operator, known as the Jacobi operator $\mathscr{J}$, acting on vectors consisting of the polynomials $p(x) = \{p_n(x)\}$ so that

$$[\mathscr{J}\, p](x) = xp(x)$$

so that p is an eigenvector of $\mathscr{J}$ with eigenvalue x. The two sequences of coefficients $\{a_n\}$ and $\{b_n\}$ form three diagonals of the operator, with $\{a_n\}$ running down the center, and $\{b_n\}$ the two diagonals on either side[20].

Given that the invariant measure for the β-transform, given by eqn 12 and visualized in figure 1 is a Borel measure, it seems reasonable to ask: what is the corresponding Jacobi operator? How can the sequence of polynomials be understood?

Szegő polynomials w.r.t. $d\mu$ are a set of orthogonal polynomials on the unit circle. Applying a Cayley transform gives the Schur functions, obeying a rational recurrence relation solvable via continued fractions. Hmmm.

And then there is Favard's theorem...

5.1 Moments

Construction of the polynomial sequences require moments. Since the invariant measures (and all of the eigenfunctions) are linear combinations of the Hesenberg basis functions, it suffices to compute the moments for these. Since the basis functions are piece-wise constant, and have an explicit expression given by eqn 16, the moments can also be given explicit expression:

$$\int_0^1 x^{n-1}\psi_p(x)\,dx = \frac{C_p}{n}\left[\frac{m_p^n - m_l^n}{m_p - m_l} - \frac{m_u^n - m_p^n}{m_u - m_p}\right]$$

with the midpoint m_p and the lower and upper midpoints $m_l < m_p < m_u$ defined just as before. Clearly, the moments rapidly get small as $n \to \infty$. Likewise, for fixed n, these also rapdily get small as $p \to \infty$.

6 The Multiplication Operator

The difficulties presented in the previous section suggests that studying the multiplication operator might be simpler. Multiplication by β is given by

$$M_\beta(x) = \beta x \tag{20}$$

The corresponding transfer operator is

$$[\mathscr{M}_\beta f](y) = \frac{1}{\beta} f\left(\frac{y}{\beta}\right)$$

The multiplication operator, superficially, in itself, is not terribly interesting; it simply rescales things. It does not generate fractals, at least, not if one confines oneself to real numbers and the canonical topology on the real-number line. If instead one works with the product topology on 2^ω, then the multiplication operator becomes rather complicated and difficult to analyze. In this sense, it is promising: it avoids the overt complexity of the logistic map, the tent map and the downshift, yet still has a complicated behavior in the product topology. In particular, the multiplication of two numbers appear to involve chaotic dynamics of the carry bit.

6.1 Downshift, Revisited

The downshift of eqn 3 takes a simple form when reinterpreted on bit-strings: it is the concatenation of multiplication, followed by a left-shift. Given a bit-string $(b_n) = 0.b_0b_1b_2\cdots$ denote its left-shift by U given by

$$U(0.b_0b_1b_2\cdots) = 0.b_1b_2\cdots$$

which, for real numbers, corresponds to

$$U(x) = \begin{cases} 2x & \text{for } 0 \le x < \frac{1}{2} \\ 2x-1 & \text{for } \frac{1}{2} \le x \le 1 \end{cases}$$

which is none-other than the Bernoulli shift of eqn 1 with a change of notation. The downshift is then

$$T_\beta(x) = M_\beta(U(x))$$

so that the iterated downshift is an alternation between a left-shift and a multiplication. The act of discarding the most significant bit (the MSB) with each left-shift alters the dynamics of iterated multiplication.

This suggests that studying multiplication and the multiplication operator might provide fruitful insight into the downshift.

6.2 Monomial Eigenfunctions

Some properties of the multiplication operator can be guessed at directly. Obviously, $f = \text{const.}$ is a decaying/growing eigenfunction, depending on whether $\beta > 1$ or not. That is, one should imagine $f = \text{const.}$ as a uniform distribution of dust; with each iteration, it is spread either farther apart ($\beta > 1$) or bunched closer together ($\beta < 1$).

Clearly, $f(x) = x^n$ is an eigenfunction, with eigenvalue $1/\beta^{n+1}$. If one considers multiplication only to operate on the positive real-number line, then n need not be an integer. In other words, the multiplication operator has a continuous spectrum in this situation.

If the domain of the operator is extended to functions on the non-negative real-number line, then n must be positive, as otherwise $f(0)$ is ill-defined. But if n is positive, then (for $\beta < 1$) the multiplication operator only has eigenvalues greater than one, which is not, in general, very desirable.

If the domain of the multiplication operator is extended to the entire real-number line, then n is forced to be an integer, in order to avoid issues due to multi-valued functions. Extending the domain to the complex plane leads us astray, and so we will not go there.

6.3 A Fractal Eigenfunction

The compressor function is also an eigenfunction. It was previously observed in eqn 9 that

$$\text{cpr}_\beta\left(\frac{x}{\beta}\right) = \frac{1}{2}\text{cpr}_\beta(x)$$

whenever $1 < \beta \leq 2$ and $0 \leq x < 1$ and so, cpr_β is potentially be an eigenfunction of $\mathscr{M}_\beta$ with eigenvalue $1/2\beta$, provided that it is extended to arguments $1 < x$. This can be done as follows. Define the extended function, valid for $0 \leq x < \infty$ and for $1 < \beta \leq 2$ as

$$\text{ecpr}_\beta(x) = \begin{cases} \text{cpr}_\beta(x) & \text{if } 0 \leq 2x < \beta \\ 2\text{cpr}_\beta\left(\frac{x}{\beta}\right) & \text{if } \beta \leq 2x < \beta^2 \\ 4\text{cpr}_\beta\left(\frac{x}{\beta^2}\right) & \text{if } \beta^2 \leq 2x < \beta^3 \\ 2^n\text{cpr}_\beta\left(\frac{x}{\beta^n}\right) & \text{if } \beta^n \leq 2x < \beta^{n+1} \end{cases}$$

The extension is performed simply by treating the self-similarity as a recurrence relation, which can be iterated to move the argument into a region where the original definition was sufficient. In essence, one applies a right-shift operator to reduce the argument. Since the multiplication operator is odd about $x = 0$, on can trivially extend this to negative x by defining $\text{ecpr}_\beta(-x) = -\text{ecpr}_\beta(x)$.

Note that the original $\text{cpr}_\beta(x)$ also had a translation symmetry: the upper half was equal to the lower half. This translation symmetry has been lost, since after all, multiplication does not preserve translation.

The ecpr function is not square integrable; it does not have an L_p-norm for any p; and this is no surprise, as its hard to imagine how it could be otherwise, for a function to be self-similar under scaling.

6.4 A Generic log-periodic Eigenfunction

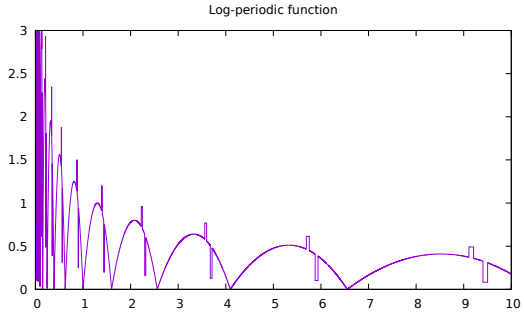
Inspired by the above, its should be clear how to build a generic eigenfunction. Let $g(x)$ be some arbitrary function, defined on the interval $1 \leq x < \beta$ (given some fixed $1 < \beta$). Define its extension as

$$g'_w(x) = w^n g\left(\frac{x}{\beta^n}\right) \quad \text{if } \beta^n \leq x < \beta^{n+1}$$

This has, by construction, the self-similarity relation $g'_w(\beta x) = w g'_w(x)$ and so is an eigenfunction with eigenvalue w/β:

$$\left[\mathscr{M}_\beta g'_w\right] = \frac{w}{\beta} g'_w$$

This function is merely log-periodic; its not fractal. Perhaps its silly to illustrate this; it should be obvious, but just in case its not, the figure below shows such a function, for $\beta = 1.6$ and $w = 0.8$. It is an eigenfunction of $\mathscr{M}_{1.6}$ with eigenvalue of 1/2.



There doesn't seem to be anything particularly interesting with this particular game. There's a simple explanation for this: The multiplication operator is generating a free monoid in one generator (the iteration itself), whereas fractals require at least two generators of self-symmetry. The (usually) free interaction of multiple generators is what forces the fractal to appear.

Note that the cpr_β function constructed above is a special case of this: It's self-similar, but the property that made it interesting, as a fractal, was erased in the construction. As before, note that $g'_w(x^n)$ is an eigenfunction with eigenvalue $1/\beta w^n$ (for integer n).

6.5 Haar Basis Matrix Elements

The Haar basis matrix elements for the downshift proved to be a bit unwieldy and not terribly useful. The corresponding matrix elements for the multiplication operator have the same general essence, but are slightly simpler and shorter to write down. In all other respects, they still have the same tractability issues.

The multiplication operator $\mathscr{M}_\beta$ has matrix elements in the standard Haar basis:

$$\begin{aligned}\left\langle mi \left| \mathscr{M}_\beta \right| nj \right\rangle &= \int_{-\infty}^{\infty} h_{mi}(x) \left[\mathscr{M}_\beta h_{nj}\right](x)\, dx \\ &= \frac{2^{(m+n)/2}}{\beta} \int_{-\infty}^{\infty} h(2^m x - i)\, h\left(\frac{2^n x}{\beta} - j\right) dx\end{aligned}$$

Instead of confining oneself to the unit interval, here it is convenient to consider the entire real-number line, and thus that is the range of the integral. Likewise, i and j an be any integers, positive or negative. As before, matrix elements vanish unless

$$\left[\frac{i}{2^m}, \frac{i+1}{2^m}\right] \cap \left[\frac{\beta j}{2^n}, \frac{\beta(j+1)}{2^n}\right] \neq \emptyset$$

This holds in three cases: where one of the intervals contains an edge transition (left, middle or right) of the other interval, without also containing the other two.

6.6 The Shift and Add algorithm

One can model the multiplication of real numbers with a number of different algorithms applied to bit strings. One of the simplest such algorithms is the shift-and-add algorithm, described here. Its just elementary-school long-form multiplication, applied to the binary expansions of the numbers.

There's a point worth laboring on: a bit string representing a real number is not the same thing as the real number. There are more bit-strings than there are real numbers. Most famously, the two bit strings $0.0111\cdots$ and $0.1000\cdots$ are two obviously distinct bit-strings, but they represent the same real number: one-half. All real numbers of the form $j/2^n$ (the "dyadic rationals") will always have dual representations; all other real numbers have a single, unique representation. These correspond to the "gaps" in the Cantor set, or, equivalently, neighboring infinite branches in the finite binary tree. Bit-strings are not real numbers. They're just a usable model of them. The usability is somewhat limited; its OK for working with individual points, but fails miserably for the topologies: the canonical topology on the reals is sharply different than the product topology on 2^ω.

The goal is to compute the product Kx with $0 \leq K \leq 1$ and $0 \leq x \leq 1$ so that the product is $0 \leq Kx \leq 1$. Both K and x are represented by their binary expansions. Let the binary expansions be

$$x = 0.b_0b_1b_2\cdots = \sum_{n=0}^{\infty} b_n 2^{-n-1}$$

and

$$K = 0.c_0c_1c_2\cdots = \sum_{n=0}^{\infty} c_n 2^{-n-1}$$

where the b_n and c_n are either 0 or 1, always.

Define $s_0 = 0$ and s_{n+1} to be the non-negative integer

$$s_{n+1} = b_nc_0 + b_{n-1}c_1 + \cdots + b_0c_n = \sum_{k=0}^{n} b_k c_{n-k} \tag{21}$$

Note that $0 \leq s_n \leq n$. It is useful to visualize this in terms of the elementary school shifted tabular form:

$$\begin{array}{llllll} 0 & c_0b_0 & c_0b_1 & c_0b_2 & c_0b_3 & \cdots \\ & & c_1b_0 & c_1b_1 & c_1b_2 & \cdots \\ & & & c_2b_0 & c_2b_1 & \cdots \\ + & & & & c_3b_0 & \cdots \\ \hline s_0 & s_1 & s_2 & s_3 & s_4 & \cdots \end{array}$$

This makes clear the shift-and-add form. The value of each individual s_n can be visualized as a stack of blocks. For the special case of $K = 0.111\cdots = 1$ one has that $s_{n+1} = \sum_{k=0}^{n} b_k$, that is, it is simply the total number of one-bits in the first n locations.

The final step is to reduce the the sum series s_n to a bit-string. This is accomplished recursively, by performing a carry operation:

$$d_n = s_n + \left\lfloor \frac{d_{n+1}}{2} \right\rfloor \tag{22}$$

where $\lfloor d \rfloor = d \mod 1$ denotes the floor of d (the integer part of d). The desired bit sequence is then

$$a_n = d_n \mod 2 \tag{23}$$

Equivalently, a_n is the remainder, the part of d_n that was not propagated to the next location. Explicitly, is is $a_n = d_n - 2\lfloor d_n/2 \rfloor$. The carry-sum propagation can be imagined as a kind of bulldozer, razing the towers d_n until they are one block high, pushing the razed bits off to the next location. The resulting sequence (a_n) is then the bit-string for the product Kx. That is,

$$Kx = 0.a_0a_1a_2\cdots = \sum_{n=0}^{\infty} a_n 2^{-n-1}$$

The problem with this algorithm is that the relation 22 for the d_n is infinitely recursive, and in general is not guaranteed to terminate. One has to start at $n = \infty$ and move backwards from there. There are two plausible scenarios for computing the a_n in practice. One is to search the n until one finds that spot where $\lfloor d_{N+1}/2 \rfloor = 0$; one can then obtain the a_n for all $n < N$ without issue. The problem here is to find such an N.

The other way to compute is to observe that the iteration is convergent. The recursion 22 only depends on a finite and fixed number of bits "behind it", roughly equal to $\log_2 n$ bits that come after this. As noted earlier, $0 \le s_n \le n$ and likewise, $0 \le d_n \le 2n+1$. To write down d_n, one needs at most $C = 1 + \lfloor \log_2(2n+1) \rfloor$ bits. This implies that a given d_n can only perturb at most $C-1$ bits downstream of it. That is, d_{n-C+1} depends on d_n but d_{n-C} does not. Thus, in order to correctly compute all bits a_k for $0 \le k \le n-C$, it is sufficient to set d_n to some arbitrary value (less than $2n+2$) and then iterate (using the correct values for s_k when $k < n$). At the end, discard all d_k and a_k for $n - C < k$, as they are incorrect.

6.7 Tree-view

Points:

1) adding one bit is like shifting the tree over sideways.
2) multiplying by one bit is like shifting the tree down-left.
3) adding a number to itself is like shifting tree up (since its just 2x)

7 Simplified Models of Multiplication

The shift-and-add algorithm is obviously rather complex; can it be replaced by something simpler? The particular question to ask is how much of the chaotic dynamics of

the downshift is due to the propagation of the carry bit, and how much of it is due to other parts of the algorithm? Specifically, the addition of two numbers, which requires a carry bit, can be replaced by a bit-wise XOR of their bit strings: this generates "almost" the same results as addition, when the number of 1-bits in the strings are sparse, but are wrong when 1-bits appear in the same location: the XOR discards the carry bits. Thus, a simplified model of multiplication would the the shift-and-XOR model: it proceeds the same way as shift-and-add, but replaces addition with XOR. What does this look like, and how does the equivalent of the downshift behave under this operation?

7.1 Shift-and-XOR

The shift-and-XOR algorithm must like the shift-and-add algorithm, except that it drops the carry bits. Starting from the same spot, let $0 \le K \le 1$ and $0 \le x \le 1$ and represent both by their binary expansions:

$$x = 0.b_0 b_1 b_2 \cdots = \sum_{n=0}^{\infty} b_n 2^{-n-1}$$

and

$$K = 0.c_0 c_1 c_2 \cdots = \sum_{n=0}^{\infty} c_n 2^{-n-1}$$

where the b_n and c_n are either 0 or 1.

Define $s_0 = 0$ and s_{n+1} to be the result of XOR-ing instead of adding the bits.

$$s_{n+1} = b_n c_0 \oplus b_{n-1} c_1 \oplus \cdots \oplus b_0 c_n = \bigoplus_{k=0}^{n} b_k c_{n-k}$$

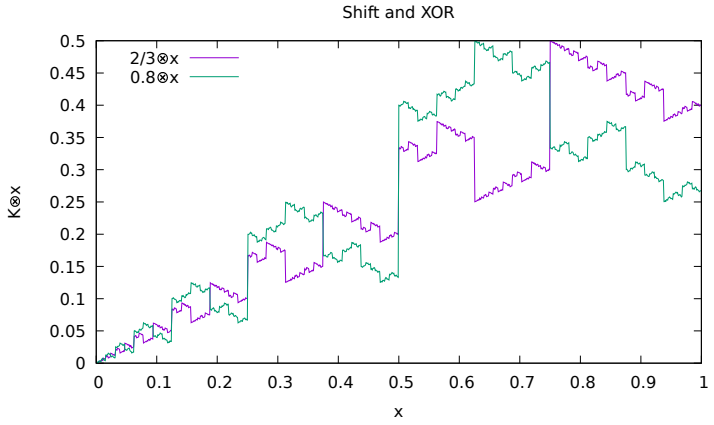
Here, the oplus symbol $\oplus$ denotes the XOR operation. Note that each s_n is either zero or one. Reconstructing a real number from this, one defines

$$K \otimes x = 0.s_0 s_1 s_2 \cdots$$

where the otimes symbol $\otimes$ is pressed into service to indicate the shift-and-XOR product. Note that it is symmetric: $K \otimes x = x \otimes K$ and so behaves at least a bit like ordinary multiplication. Its is not distributive over ordinary addition: $(a+b) \otimes x \neq a \otimes x + b \otimes x$ but it is distributive over XOR: $(a \oplus b) \otimes x = (a \otimes x) \oplus (b \otimes x)$. It is illustrated in figure 25.

The range of the shift-and-XOR operation is fundamentally different from multiplication. First, because the carry bit is dropped, one has that $s_0 = 0$ always, and so that $K \otimes x \le 1/2$ always, even when both $K \to 1$ and $x \to 1$. Next, for any value of $1/2 < K \le 1$, the range of $K \otimes x$ runs over the entire interval $[0, 1/2]$ as x runs over the interval $[0,1]$. The measure is not compressed (other than by a factor of 2) , as there is in ordinary multiplication. That is, if $S \subset [0,1]$ is a measurable subset of the unit interval, with measure $\mu(S)$, then one has $\mu(K \otimes S) = \mu(S)/2$. There are several ways to prove this. One formal approach is to consider the correspondence between the natural

Figure 25: Shift and XOR Algorithm



This figure shows two functions, $(2/3)\otimes x$ and $(4/5)\otimes x$ as a function of x.

---

measure on the reals, and the measure of cylinder sets on the product topology. That is, the Cantor space $\{0,1\}^{\omega}$ is endowed with a natural topology, the product topology. The open sets of this topology are called "cylinder sets". Their measure is uniformly distributed over unit interval, precisely because the Bernoulli shift is ergodic: the one implies the other.

Indeed, the shift-and-XOR algorithm can be best thought of as a formula for shuffling the bit-strings around, without actually altering them: re-ordering them, not changing them. The intuitive key to this is to observe that subtracting x from 1 just re-orders the unit interval, top to bottom, and that this is the same as flipping all zero bits to one, and v.v. That is, $1-x = x \oplus 0.111\cdots$.

Another way to see this shuffling is to note that $a \oplus a = 0$ and that $0 \oplus x = x$. Thus, for a fixed value of a, the string x and the string $a \oplus x$ are paired together, in a unique way, so that either can be gotten from the other. The function $a \oplus [0,1] \to [0,1]$ sending $x \mapsto a \oplus x$ is an exchange of these unique pairings of strings. It is not just a bijection, it is an involution. If the strings are given their natural lexicographic sort order, the mapping $x \mapsto a \oplus x$ is just a certain kind of shuffle of the sort order; it neither adds new strings, nor deletes any, nor changes their number. The function is one-to-one and onto. The multiply-and-XOR algorithm is just a repeated sequence of XOR's:

$$K \otimes x = \left(\frac{c_0 x}{2}\right) \oplus \left(\frac{c_1 x}{4}\right) \oplus \left(\frac{c_2 x}{8}\right) \oplus \cdots$$

and so $K \otimes x$ is nothing more than a reshuffling of strings (along with a right-shift equal to the number of leading zero-bits in the binary expansion of K; the right-shift commutes with the measure on the product topology.) Thus, $K \otimes x$ preserves the measure on the unit interval (up to a factor of 2^{-n} due to the above-mentioned right-shift). That is, for $1/2 < K \leq 1$, this discussion shows that $\mu(K \otimes S) = \mu(S)/2$.

7.2 Self-similarity

There are several self-similarity properties of the shift-XOR worth noting. It behaves very much like a classic dyadic fractal. Thus, one has that

$$K \otimes \left(\frac{x}{2}\right) = \frac{1}{2}\left(K \otimes x\right) = \frac{1}{2} K \otimes x$$

In addition... TODO: illustrate the other symmetry.

7.3 Similarity Transformations

The shift-and-XOR algorithm acts as a permutation on bit-strings. As a result, the XOR-analogs of the downshift and the tent map become uniformly ergodic, behaving exactly as the Bernoulli shift. The Frobenius-Perron solution to these is just the uniform distribution, which is featureless. All of the structure visible in figures 2 and 3 is entirely due to the dynamics of the carry bit. Effectively, the carry-bit algorithm alters the uniform distribution of the Bernoulli shift (equivalently, the uniform distribution associated with the natural measure on Cantor space.)

Define the XOR-analog of the downshift as

$$c_\beta(x) = \begin{cases} 2\beta \otimes x & \text{for } 0 \le x < \frac{1}{2} \\ 2\beta \otimes \left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} \le x < 1 \end{cases}$$

The factor of 2 makes up for the fact that shift-XOR effectively drops the top bit; thus the goal is to map each half of the unit interval into the entire interval $[0,1]$.

Given a fixed β, define $\boxtimes_\beta : [0,1] \to [0,1]$ as

$$\boxtimes_\beta(x) = \beta \otimes x$$

As observed previously, $\boxtimes_\beta$ is an automorphism of the unit interval, and more: it is a permutation on Cantor space. Let $b(x)$ be the Bernoulli shift of eqn 1; then one has that $c_\beta = \boxtimes_\beta \circ b$. Taken together, this implies that the ergodic properties of iterating on c_β follow directly from the ergodic properties of the Bernoulli shift; a shuffle, any shuffle on the Cantor set should not alter these ergodic properties.

TODO: similarity transforms on the transfer operator... and the non-alteration of the eigenspectrum, even as the eigenfunctions are altered.

7.4 Multiplication on the Cantor Space

The previous set of results indicates that all of the structure in the bifurcation diagrams of 2 and 3 is entirely due to the dynamics of the propagation of the carry sum. To explore this, the notation needs to be improved on.

The downshift can be decomposed into multiple distinct stages. First, there is a conversion from the unit interval to the Cantor space; this was defined at the very start, but now we need a less awkward notation for it. Let

$$\begin{array}{ccc} \pi : 2^\omega & \to & [0,1] \\ 0.b_0 b_1 b_2 \cdots & \mapsto & x \end{array}$$

be the projection from the Cantor space to the real-number unit interval, given by eqn 2. Note that it is a surjection: dyadic rationals (rationals of the form $m/2^n$) correspond to two distinct bit strings. For example, $1/2$ can be represented as both $0.1000\cdots$ and as $0.0111\cdots$. Cantor space covers the unit interval. Write the inverse mapping as

$$\begin{array}{ccc} \pi^{-1} : [0,1] & \to & 2^\omega \\ x & \mapsto & 0.b_0b_1b_2\cdots \end{array}$$

As a function, it is injective but not surjective. It is usually convenient to ignore this, and to pretend that both π and π^{-1} are bijections, even though they are not. This rarely leads to practical difficulties, as long as one stays conceptually tidy. Better yet, just perform all work on the Cantor space, and project to the unit interval only when needed.

Next, turn to multiplication. This has three parts. First, the summation of the carry bits:

$$\begin{array}{ccc} S_\beta : 2^\omega & \to & \mathbb{N}^\omega \\ 0.b_0b_1b_2\cdots & \mapsto & (s_0,s_1,s_2,\cdots) \end{array}$$

where the summation is given by eqn 21. Here, $\mathbb{N}^\omega$ is Baire space, the space of all infinite-length sequences of non-negative integers. In number theory, this would be called the space of arithmetic functions. The second part of multiplication is the propagation of the carry bits. Denote this as

$$\begin{array}{ccc} C : \mathbb{N}^\omega & \to & \mathbb{N}^\omega \\ (s_0,s_1,s_2,\cdots) & \mapsto & (d_0,d_1,d_2,\cdots) \end{array}$$

which is defined in eqn 22. Finally, one extracts the remainder, after propagation:

$$\begin{array}{ccc} A : \mathbb{N}^\omega & \to & 2^\omega \\ (d_0,d_1,d_2,\cdots) & \mapsto & (a_0,a_1,a_2,\cdots) \end{array}$$

which is given by eqn 23. Of the three parts into which we've decomposed multiplication, only the first part is parameterized by K. Thus, multiplication, on Cantor space, can be written as $M_\beta = A \circ C \circ S_\beta$. The shift-and-XOR algorithm omits the propagation of the carry sum. On Cantor space, it is just $\boxtimes_\beta = A \circ S_\beta$: the XOR is just modulo-2 of the carry sum.

To obtain multiplication on the real-number unit interval, we need merely to re-project from Cantor space to the reals. Thus, multiplication, given in eqn 20, decomposes into

$$M_\beta = \pi \circ A \circ C \circ S_\beta \circ \pi^{-1}$$

The downshift of eqn 3 is then

$$T_\beta = \pi \circ A \circ C \circ S_\beta \circ \pi^{-1} \circ b$$

where b is the Bernoulli shift. To simplify notation, it is convenient to go ahead and provide a symbol for the shift operator:

$$\begin{array}{ccc} B : 2^\omega & \to & 2^\omega \\ (b_0,b_1,b_2,\cdots) & \mapsto & (b_1,b_2,\cdots) \end{array}$$

so that $b = \pi \circ B \circ \pi^{-1}$. The corresponding downshift on the Cantor space is

$$B_\beta = A \circ C \circ S_\beta \circ B$$

which eliminates the pesky projection π. It should be clear that S_β is an injection, the propagation operation C and the remainder A are both surjections.

As noted, the shift-and-XOR algorithm can be written as $\boxtimes_\beta = A \circ S_\beta$; the step where the carry bits are propagated is dropped. The XOR-version of the downshift is

$$c_\beta = \boxtimes_\beta \circ B = A \circ S_\beta \circ B$$

Thus, in this new notation, it reaffirms that B is the true source of ergodicity, and that $A \circ S_\beta$ being a permutation does not alter the basic ergodic property of B. All of the structure in the bifurcation diagrams can be blamed on the propagation operator C.

7.5 Propagation games

Pinning the "blame" of complex dynamical structure on the propagation of the carry bits seems to be an open invitation to replace the propagation operator C by just about anything, to see what happens. Figure 26 illustrates some of the things that can happen.

Reviewing the images there makes it clear that although fiddling with the carry bit fundamentally alters point trajectories, it completely fails to open any doors that would provide insight into the structure of the transfer operator. The pictures are pretty, but appear to be meaningless.

8 Sci-fi day-dreaming

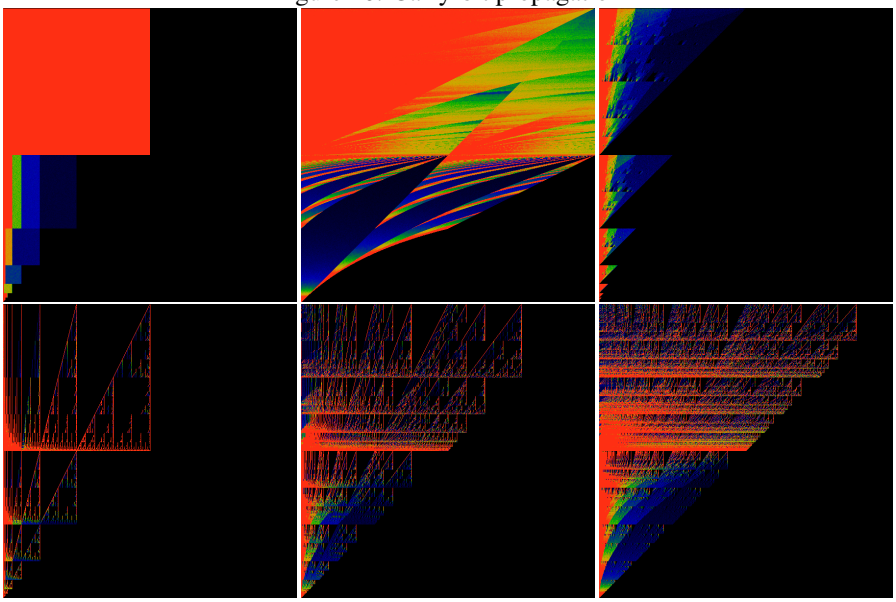
This section provides two day-dreams inspired by this material. They are just that: daydreams. If you don't like fictional daydreaming, you won't like the material here. Sorry about that.

8.1 Limits to computation

There are many limits to computation. One limit is the speed of light. In current generation CPU chips, clock rates in the vicinity of 3 gigahertz= 3×10^9 cycles per second. By comparison, the speed of light in a vacuum is about 3×10^8 meters per second. Dividing, one finds that light can travel about $3 \times 10^8 / 3 \times 10^9 = 10^{-1}$ meters, or about four inches: a bit bigger than the actual physical dimensions of a chip (typically around half-an-inch on a side), but not by much. Of course, the speed of light in a metal conductor is lower – about half the speed in a vacuum. And transistors are small – more than twenty-thousand times smaller. So, measured in terms of the size of the transistor, the speed of light is about ten or twenty transistor-widths per clock-cycle. So, OK, its still fast, at that length scale. But not really all that fast. The point here is that the speed of light is a potential limit to the speed of computation, and it is not all that far away.

In this setting, one can imagine the situation where the speed of propagating the carry bit during multiplication becomes a limiting factor. The above work hints at

Figure 26: Carry-bit propagation



Two triptychs of different carry-bit behaviors. Define $F : \mathbb{N}^\omega \to \mathbb{N}^\omega$ by $F = f \times f \times f \times \cdots$ and then iterate on $A \circ C \circ F \circ S_\beta \circ B$. For $f(n) = n$ one obtains, of course, the standard downshift of figure 2. The top-left image shows $f(n) = n \mod 2$, which is the same as iterating on the shift-XOR function c_β. Here, β runs from 0 at the bottom, to 2 at the top; x runs from 0 to 1, left to right. The uniform red square simply indicates that the iteration is completely independent of β when $1 < \beta \leq 2$: it is fully uniform and ergodic in the same way that the Bernoulli shift is. The top-middle image shows $f(n) = n + 1$, that is, pretending that there is one carry bit too many. The top-right shows $f(n) = \max(0, n-1)$, that is, having one carry-bit too few.

The bottom three shows a progression of $f(n) = \max(n, 1)$, $f(n) = \max(n, 2)$ and $f(n) = \max(n, 3)$, allowing more and more carry bits to propagate. In the limit, this becomes figure 2 once again. Except for the top-left image, the rest seem pointlessly goofy.

---

a somewhat boggling idea: can multiplication be effectively parallelized by working with transfer operators instead? That is, the multiplication of two numbers corresponds to point-wise particle dynamics: a discrete particle following a chaotic path through a complex numerical computation. By contrast, the transfer operator describes how a distribution propagates through a computation: it effectively performs "an infinite number" of multiplications at the same time, in parallel. That is, rather than asking how single values propagate, one could, and perhaps should, ask how distributions propagate – parallelize multiplication (for example) to an "infinite" degree. It is this rather ridiculous idea that suggests that the above explorations are not purely abstract, but have a potentially practical application. As I suggested – its a bit of science-fiction day-dreaming at this point. But it does hint at an alternate model of computation.

Variants of this model have already been explored, for decades. For example, Crutchfeld defined "geometric state machines" as generalizations of finite state machines, where, instead of having a finite matrix (a "transition matrix") act on a finite vector (the "state vector"), one instead considers operators acting on homogeneous spaces – that is, applying a sequence of such operators on homogeneous space. The most famous and celebrated such space would the $\mathbb{CP}^n$ – complex projective space, with the operators that act on it being the the unitary ones: $U(n)$ – such a system defining the n-qubit quantum state machine. Distributions on $\mathbb{CP}^n$ are mixed states – and the idea of quantum computing is to evolve such states through a set of operations.

The point here is that computation, by means of the time-like evolution of distributional densities, is already being explored, but in a rather different context than the one explored here. Here, it seems like we are bowled over by the complexities of a seemingly much simpler system.

8.2 Wave function collapse

There is also a different, bizarrely hypothetical way in which all of this apparatus could manifest itself. Currently, in order to avoid the rather severe issues associated with the concept of quantum-mechanical wave-function collapse, the (vast?) majority of practicing physicists believe in the many-worlds hypothesis. Clearly, this belief is entirely correct for microscopic systems, isolated from the usual thermodynamic hustle and bustle (chlorophyll, rhodopsin and the magnetically sensitive cryptochromes notwithstanding). But it seems to fly in the face of daily experience, where we are aware of just one reality. One of my favorite hypotheses is that this is the result of the (rapid) decay of macroscopic quantum states down to a probability of zero. The mechanism is presumably that of decaying subshift measures. Penrose argues that this has something to do with gravity; but we can go one better: the natural setting for shift spaces are hyperbolic spaces, as that is where there is enough room to "fit everything" in a uniform way consistent with a metric. Curiously, the world we live in – Minkowski space, is hyperbolic. This suggests that the Many Worlds interpretation is exactly right, as long as one truly is in Minkowski space, but that gravitation, which essentially bends or distorts it, squeezes down the room available for multiple quantum states, effectively forcing the collapse in this way.

Put another way: the standard treatment for quantum field theory is the Feynman functional integral; it can be viewed as an integral over all possible paths that a "par-

ticle" might take. The daydream is to equate a specific path with the idea of point-dynamics in an iterated function. As long as one considers only points, and there movement, one can be completely unaware of either the invariant measure, or of the decaying eigenstates of the shift operator. In a standard QFT textbook, all equations appear microscopically time-reversible. There's almost no idea of a measure, except for the $\exp -i\hbar S$ in the Feynman integral. The incorporation of gravity into this is famously difficult. The daydream here is that gravity manifests itself as eigenfunctions that live off of the shell of unitary evolution.

There is some practical hope of bringing this daydream to fruition: the theory of subshifts has seen dramatic advances over the last few decades, getting increasingly abstract, and gaining a firm footing in very general settings: viz not just in metric spaces, but even in more general topological vector spaces, and specifically in stereotype spaces, where most of the devices used in analysis can be exercised in reasonably safe manner. The point here is that most of QFT can be formulated using these more-or-less conventional tools and notations. The trick is to locate and extract those parts that renormalize to zero, not unlike some of the formally divergent sums explored above, which can none-the-less be regulated and made to give reasonable answers. Or at least, that's the daydream. Clearly, got far to got before it can be reality.

9 Topological Push-Forward

The transfer operator is most generally and correctly defined as an operator acting on the topology of a space, and specifically, as the push-forward of the (uniform) measure by the iterated function. That is, given any open set belonging to the topology, the transfer operator assigns a different open set of the topology: it is a map of sets to sets. For iterated maps on the unit interval, it is essentially a map of cylinder sets, the open sets of the product topology. The shift-XOR experiment shows that the ergodic properties arise from the Bernoulli shift, and that all other properties, commonly called "chaotic", are really the side effect of something else, entirely: the internal structure of the transfer operator. Fiddling with the carry-bits cannot reveal this structure; instead, they just define other, pointlessly goofy iterated functions. Point trajectories fail to reveal the internal structure of the transfer operator, and at best point in a misleading direction. To understand the transfer operator, it must be tackled for what it is: one must look at how intervals are mapped to intervals, and what sort of symmetries can be discovered in this mapping. (I've given one sketch of a proof of the transfer operator as a push-forward in this reference:[21]. There are must surely be better, more accessible and more formal and mathematically refined presentations; if you, reader, know osf such, please drop me a line.)

The action of the transfer operator on the sets belonging to the topology of the reals reveals several distinct kinds of actions. The topology on the reals can be generated from a basis consisting of connected sets. The transfer operator will map some connected sets to other connected sets, simply moving them around, shrinking or expanding them. In other cases, a connected set will be split into two disjoint parts. For maps that are continuous, there must be regions that have fixed-points and period-doubling routes to chaos: these correspond to the (countable number of) "trouble spots" illus-

trated in section 4.3.

It seems reasonable to argue that each of these different kinds of moves creates a distinct group (or monoid) of transformations: in a certain sense, those transforms that do not change th connectivity, nor do any folding, are all similar to one-another. It should be possible to write down exactly which sets belong to this type, and then give explicit transformation properties between them. Likewise, those connected sets which are split in two are all similar. It seems like there should be a prototype: a generic split, followed by some re-arrangement of the two parts. How can this classification be written in an insightful, useful way?

I beleive that there has been a sufficient number of advances in the theory of subshifts so that the above vague sketch can be presented in a fairly concrete way. Unfortunately, most of the relevant material remains rather arcane and abstract, lacking in direct accessibility to casual students. I am not currently aware of any adequate yet accessible treatment.

10 Conclusion

What, exactly, is the point of mathematics, especially in the computational age? Can't one just get a fast computer, iterate on the logistic map, and find out everything there is to find? Well, of course, yes, and no: these questions can be taken as either silly or as deeply philosophical, and it is worth the effort to understand them and address them properly.

First, lets dispose of some obvious mis-perceptions. If one carefully scrutinizes figure 1, one will find signs of a slight unevenness in the horizontal bars. These are numerical artifacts due to statistical under-sampling: they smooth out and fade away with additional sampling of the iterated equations. There is a way to obtain this same figure, far more rapidly, and without this particular form of numerical noise: one can instead iterate on equation 13. This suggests one philosophical answer: the goal of mathematics is to find faster ways of computing things; to discover better algorithms.

A uniting theme between this, and the other text that I have written on fractal issues, is that they are all explorations of the structure of the Cantor set, the structure of the space of infinite sequences of symbols, and the structure of the continuum. That is, we know the continuum in two different ways: one way is by means of the natural topology on the real number line; the other is the product topology on the space of binary strings. The former is suggested by the physical universe that we actually live in: a continuum with spatial extents. The latter is suggested by the notion of time and repetition: the making of choices naturally leads to a tree structure; tree structures necessarily embed in hyperbolic spaces; the Minkowski space that we live in is hyperbolic, and this is why, every day, as time passes on, we get to make new choices precisely because the amount of room for possibilities is ever-increasing as time flows forward.

What, exactly, do the words "exactly solvable" really mean? So, for example, equation 15 involves summation and multiplication, which has this aura of comfortable preciseness that an iterated function somehow does not. Where does this sensation come from? When performing arbitrary-precision numerical computations, it should be clear that neither addition nor multiplication are simple or easy: they both require fairly com-

plex algorithms to implement, and have not-insignificant runtimes. To be more precise: the algorithms are non-trivial because one is using a binary digit expansion to provide a model for a real number. Different representations of the real numbers potentially offer different algorithms and performance profiles: one could represent reals by rationals, but then one is left with the issue of trying to add or multiply two integers. To accomplish this, one has to represent the integers as sequences of bits, which only takes us back to where we started. There is no computational oracle that automatically knows the sum or product of integers: it has to be computed. The analysis being done in this text is a kind of a game, where not only is one algebraic arrangement of symbols is being compared to another, but also one computational algorithm is being compared for another. Unfortunately, this latter comparison is very nearly opaque and hidden. If only it could be made visible in some simple fashion.

The situation here is more easily illustrated in a different domain. The hypergeometric series was presented ad studied by Gauss; Gauss, Kummer, Pfaff and Euler observed various identities yoking together different series. By the 1950's, thousands of relations were known, along with some algorithms that can enumerate infinite series of relations. The curious situation is that there is no known algorithm that can enumerate all such relations; there is no systematic way to classify them. The situation does seem to make clear that there is an interplay between infinite series and algorithmic relationships between them. Stated a different way: hypergeometric series are self-similar, and the identities relating them are expressions of that self-similarity.

To further sharpen this idea: the dyadic monoid is the generator of self-symmetry in many common fractals; this is "well-known", and I have explored this in other texts. A more general setting for fractal self-similarities is given by tilings of the hyperbolic surface: to each tiling, there are corresponding fractals, the self-similarity of which are given by the tiling. The figures 2, 3 and 4 are clearly self-similar in some obscure way: it is visually clear, but providing a simple algebraic expression describing the similarity is difficult; I have not been successful in this. None-the-less, it seems self-evident that it will be the dyadic monoid that is somehow responsible for, or underlying the symmetries (unless, of course, there is some other, as yet undiscovered structure).

The meta-question is: what is the correct framework by which one can best understand the interplay between symmetries, infinite series and algorithms? The current tool-set seems impoverished: it does not "solve" the systems in this text. Worse, the current system reifies addition and multiplication into oracular operations that magically obtain "the right answer", when it is clear from numerical methods that addition and multiplication are necessarily algorithmic operations performed on finite truncations of infinite series. It would be nice to place these operations on equal footings, so as to expose the true nature of this beast.

11 Bibliography

The references below provide a very minimal bibliography. Its heavy on self-citations. Search engines exist to help you find the things you don't know, and want to find out more about.

References

[1] A. Renyi, "Representations for real numbers and their ergodic properties", *Acta Math Acad Sci Hungary*, 8, 1957, pp. 472–493.

[2] W. Parry, "On the beta-expansion of real numbers", *Acta Math Acad Sci Hungary*, 11, 1960, pp. 401–416.

[3] Karma Dajani, et al., "The natural extension of the beta-transformation", *Acta Math Hungary*, 73, 1996, pp. 97–109, URL https://www.researchgate.net/publication/2257842.

[4] F. Blanchard, "Beta-expansions and Symbolic Dynamics", *Tehoretical Comp Sci*, 65, 1989, pp. 131–141.

[5] Bruno Henrique Prazeres de Melo e Maia, "An equivalent system for studying periodic points of the beta-transformation for a Pisot or a Salem number", , 2007, URL http://repositorio.ual.pt/bitstream/11144/471/1/bmaia_thesis.pdf.

[6] Vaughn Climenhaga and Daniel J. Thompson, "Intrinsic ergodicity beyond specification: beta-shifts, S-gap shifts, and their factors", *Israel Journal of Mathematics*, 2010, URL https://arxiv.org/abs/1011.2780.

[7] Leopold Flatto, et al., "The Zeta Function of the Beta Transformation", *Ergodic Theory and Dynamical Systems*, 14, 1994, pp. 237–266.

[8] Daniel J. Thompson, "Irregular sets and conditional variational principles in dynamical systems", , 2010, URL https://people.math.osu.edu/thompson.2455/thesis_thompson.pdf.

[9] L. Barreira and B. Saussol, "Variational Principles and Mised Multifractal Spectra", *Transactions of the American Mathematical Society*, 353, 2001, pp. 3919–3944, URL http://www.math.univ-brest.fr/perso/benoit.saussol/art/mixed.pdf.

[10] Lyndsey Clark, "The beta-transform with a hole", *arXiv*, arXiv:1412.6384 [math.DS], 2014, URL https://arxiv.org/abs/1412.6384.

[11] Kevin Hare, et al., "Three Series for the Generalized Golden Mean", *Arxiv*, 2014, URL https://arxiv.org/abs/1401.6200.

[12] A.P. Stakhov, "The Generalized Principle of the Golden Section and its applications in mathematics, science, and engineering", *Chaos, Solitons and Fractals*, 26, 2005, pp. 263–289, URL http://www.student.oulu.fi/~taneliha/Phi6/1/The%20Generalized%20Principle%20of%20the%20Golden%20Section%20and%20its%20applications%20in%20mathematics,%20science,%20and%20engineering.pdf.

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

[14] Linas Vepstas, "The Bernoulli Map", , 2004, URL https://www.linas.org/math/bernoulli.pdf.

[15] Linas Vepstas, "The Minkowski Question Mark, PSL(2,Z) and the Modular Group", , 2004, URL https://www.linas.org/math/chap-minkowski.pdf.

[16] Linas Vepstas, "Symmetries of Period-Doubling Maps", , 2004, URL https://www.linas.org/math/chap-takagi.pdf.

[17] Edward B. Saff and Nikos Stylianopoulos, "Asymptotics for Hessenberg matrices for the Bergman shift operator on Jordan region", *arXiv*, arXiv:1205.4183 [math.CV], 2012, URL https://arxiv.org/abs/1205.4183.

[18] E. Torrano V. Tomeo, "Two applications of the subnormality of the Hessenberg matrix related to general orthogonal polynomials", *Linear Algebra and its Applications*, Volume 435, Issue 9, Pages 2314-2320, 2011, URL http://oa.upm.es/id/eprint/8725/contents.

[19] Carmen Escribano, et al., "The Hessenberg matrix and the Riemann mapping", *arXiv*, arXiv:1107.603 [math.SP], 2011, URL https://arxiv.org/abs/1107.6036.

[20] Gerald Teschl, *Jacobi Operators and Completely Integrable Nonlinear Lattices*, American Mathematical Society, 2000, URL https://www.mat.univie.ac.at/~gerald/ftp/book-jac/jacop.pdf.

[21] Linas Vepstas, "On the Minkowski Measure", *ArXiv*, arXiv:0810.1265, 2008, URL http://arxiv.org/abs/0810.1265.