

On the Beta Transformation

Linas Vepštas

December 2017 (Current version: March 2024)*

linasvepstas@gmail.com [ORCID: 0000-0002-2557-740X](https://orcid.org/0000-0002-2557-740X)

Abstract

The beta transformation is the iterated map $\beta x \pmod 1$; it generates the base- β expansion of a real number x . Every iterated piece-wise monotonic map is topologically conjugate to the beta transformation. For all but a countable subset of β , the orbits of x are ergodic; yet it is the finite orbits that determine overall behavior.

This is a large text; it splits into four parts. The first part provides a review of general concepts and properties associated with the beta shift. The second part examines the spectrum of the Ruelle–Frobenius–Perron operator, and gives explicit expressions for various classes of eigenfunctions, some forming a discrete spectrum, others forming a continuous spectrum.

The third part examines the finite and the periodic orbits. These are in one-to-one correspondence with monic integer polynomials. They are “quasi-cyclotomic” and can be counted with Moreau’s necklace-counting function; curiously, they do not have any obvious relation to other systems countable by the necklace function. The positive real roots are dense in the reals; they include the Golden and silver ratios, the Pisot numbers, the n -bonacci (tribonacci, tetranacci, *etc.*) numbers. The beta-polynomials yoke all of these together into a regular structure. An explicit bijection to the rationals is presented.

The fourth part of this text examines small perturbations. These introduce Arnold tongues, which inflate the finite orbits, a set of measure zero, to finite size.

This text assumes very little mathematical sophistication on the part of the reader, and should be approachable for any enthusiast with minimal or no prior experience in ergodic theory. Most of the development is casual. As a side effect, the introductory sections are perhaps a fair bit longer than strictly needed to present the new results.

*This text is an extract of a diary of research notes, first published in December 2017, with major updates published in February 2018, December 2018 and January 2024. The current version expunges all the unfinished, boring and speculative parts of the diary, and prettifies the rest into a presentable format. MSC2020 Classification 37E05, 37G10.

1 Introduction

Given a real number x , it can be written down in base-ten, decimal notation, as a string of digits 0-9. Each successive digit is given by an iterated formula. Starting with $y_0 = x$, let

$$y_{n+1} = (10y_n) \pmod{1}$$

The n 'th digit in the decimal expansion is $d_n = \lfloor y_n \rfloor$. Here, $\pmod{1}$ is the modulo (or remainder) function, so that $a \pmod{b}$ is the remainder of a after dividing by b . The bracket $\lfloor z \rfloor$ is the floor function, returning the largest integer less than z . The decimal expansion of x is then $d_0.d_1d_2d_3\cdots$ with all digits d_n (except for the first) running between 0 and 9. The iteration has the form of a shift; it shifts over by one digit, and repeats the process. The shift is ergodic over the reals: any and every decimal expansion will occur.

The same idea can be applied by replacing ten by any number β ; this gives the base- β expansion. The β does not need to be an integer; the rest of this text is focused on exploring the expansions coming from $1 < \beta \leq 2$, so "base two or less".

The beta transformation is the map

$$t_\beta(x) = \beta x \pmod{1}$$

iterated on the unit interval $[0, 1]$ of the reals, and β taken to be a real positive constant. For the special case of $\beta = 2$, this is the Bernoulli process, the ergodic process of random coin flips. The Bernoulli shift is solvable in more ways than one; it is well-understood, and provides a canonical model of ergodic behavior. The beta transformation is less well-known and less carefully scrutinized. A full review of previous results is presented at the end of this introduction. The goal of this text is two-fold: to expand the collection of known properties, and to present them in the simplest, most accessible fashion possible.

One primary driver of interest in iterated functions is that they generate fractals, and that they provide simple models of chaotic dynamics. The β -transform is the simplest possible map that exhibits all of the hallmarks of chaotic dynamical systems: a non-trivial, nonuniform invariant measure, with both periodic and non-periodic orbits. It is also "generic", in that any other piecewise-monotonic map of the unit interval is topologically conjugate to it. If one can understand the β -transform, one has gone a long way towards understanding iterated maps in general.

There are two broad approaches for studying iterated functions. One is to examine the point dynamics and orbits: "where does the point x go, when iterated?". The other is in terms of distributions: "how does a scattered dust of points evolve over time?". Within the context of physics, these give two broad philosophical views of reality. The first is of microscopic, time-reversible systems whose future is deterministic and known with infinite precision. The second is of macroscopic, time-irreversible thermodynamics, where time can only go forward, and the future is unknown and unknowable. Of these two approaches, the first is commonplace and inescapable; the second remains obscure, poorly-recognized and opaque. Thus, a large part of this text is devoted to this second approach.

If iterating a map $x \mapsto f(x) \mapsto f(f(x)) \mapsto \dots$ pulls a point x through f , through time, then the action of the map f on a distribution ρ is a pushforward:

$$\rho(A) \mapsto \rho(f^{-1}(A)) \mapsto \rho(f^{-1}(f^{-1}(A))) \mapsto \dots$$

The proper definition of a pushforward requires a significant development of the concepts of measurable spaces and Borel sigma algebras, topics that will be gently reviewed a bit further in this introductory section. For the present, it is enough to take $\rho : [0, 1] \rightarrow \mathbb{R}$ to be some function defined on the unit interval. In the above, $A \subset [0, 1]$ is a subset of the unit interval, so that $\rho(A) = \int_A \rho(x) dx$ is an ordinary integral, the “size” of the set A with respect to the distribution.

The challenge is to find an explicit expression for the pushforward $\rho(f^{-1}(A))$. This can be obtained as a change of variable $y = f(x)$ under integration. Start with any function h ; its integral over the set A is as above: $h(A) = \int_A h(y) dy$. Under the change of variable, this becomes

$$h(A) = \int_A h(y) dy = \int_{f^{-1}(A)} h(f(x)) |f'(x)| dx$$

Writing the integrand as $\rho(x) = h(f(x)) |f'(x)|$ and working backwards, one recovers

$$h(y) = \frac{\rho(f^{-1}(y))}{|f'(f^{-1}(y))|}$$

Plugging this back through gives the identity

$$\int_A \frac{\rho(f^{-1}(y))}{|f'(f^{-1}(y))|} dy = \int_{f^{-1}(A)} \rho(x) dx = \rho(f^{-1}(A))$$

The right-hand-side is the desired pushforward; the left-hand side is an explicit expression for it. There was a minor sleight-of-hand in the above derivation: the map $y = f(x)$ may not be one-to-one. Thus, there may be several distinct points $x = f^{-1}(y)$. In this case, the above needs to be amended as

$$h(y) = \sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|f'(x)|}$$

As h depends only on ρ and f , the sum construction on the right-hand side can be thought of as an operation \mathcal{L} , defined by f , acting on ρ ; in short-hand, $h = \mathcal{L}_f \rho$.

The symbol \mathcal{L} is used to remind that this is a linear operator: $\mathcal{L}(a\rho + b\sigma) = a\mathcal{L}\rho + b\mathcal{L}\sigma$ for any pair of real numbers a, b and any functions ρ, σ . The pushforward sequence now becomes

$$\rho(A) \mapsto [\mathcal{L}_f \rho](A) \mapsto [\mathcal{L}_f \mathcal{L}_f \rho](A) \mapsto \dots$$

Thus, we’ve defined a linear operator \mathcal{L}_f that depends only on the iterated function f , and has the property of mapping distributions to other distributions as it is iterated. It is the result of commuting with function composition: $\mathcal{L}_f \circ \rho = \rho \circ f^{-1}$; it’s a kind of a trick with function composition. Indeed, one can define an analogous operator

$\mathcal{H}_f \circ \rho = \rho \circ f$, the “composition operator” or “Koopman operator”, that acts as a kind of (one-sided) inverse to \mathcal{L}_f .

Formally, the pushforward \mathcal{L}_f is called the “transfer operator” or the “Ruelle–Frobenius–Perron operator”. As a linear operator, the full force of operator theory comes into play. The primary task is to describe its spectrum (its eigenfunctions and eigenvalues). Two aspects of this spectrum are interesting. The first is the so-called “invariant measure”, the distribution $\mu : [0, 1] \rightarrow \mathbb{R}$ that defines a density on the unit interval that is invariant under the application of the pushforward: $\mathcal{L}_f \mu = \mu$. An informal example of such an invariant measure are the rings of Saturn: an accumulation of dust and gravel, orbiting Saturn, coupled by gravitation to both Saturn and orbiting moons, yet in a stable dynamical distribution. This is the physical meaning and importance of the invariant measure; more generally, it appears as the “ground state” or “thermodynamic equilibrium state” in a vast variety of dynamical systems.

Aside from the invariant measure, there is also the question of the rest of the spectrum. These are described by the eigenfunctions ρ satisfying $\mathcal{L}\rho = \lambda\rho$. By the theorem of Frobenius–Perron, all these other solutions have eigenvalue $|\lambda| < 1$. In physics, these correspond to the decaying modes, to the distributions that disappear over time. For the example of Saturn, these are anything not orbiting in the plane of the rings: tidal forces and perturbations from the moons will force such orbits either into the ring, or crash into the planet, or possibly fly away to infinity. The other orbits are not stable. Thus, a characterization of the decaying spectrum is of general interest.

A much stronger conception is that the decaying spectrum has something to do with the irreversibility of time. In the macroscopic world, this is plainly obvious. In the microscopic world, the laws of physics are manifestly time-reversible. Somehow, complex dynamical systems pass through a region of chaotic and turbulent motion, culminating in thermodynamic equilibrium. The decaying spectrum provides a conceptual framework in which one can ponder this transition.

This is where the mathematical fun begins. The spectrum is not a “fixed thing”, but depends strongly on the space of functions on which \mathcal{L}_f is allowed to act. If one limits oneself to ρ drawn from the space of piece-wise continuous and smooth functions, *i.e.* polynomials, then \mathcal{L}_f will in general have a discrete spectrum. If instead, $\rho \in L^2 [0, 1]$ the space of functions that are square-integrable on the unit interval, then the spectrum will often be continuous, and perhaps may have a large kernel. Larger spaces exhibit even wilder behavior: if one asks only that ρ be L^1 -integrable (not square-integrable), then it is possible for continuous-nowhere functions to appear as eigenfunctions of \mathcal{L}_f . An explicit example of the latter is the Minkowski measure for the transfer operator of the Minkowski Question Mark function: it vanishes on the rationals, but can be integrated just fine; its integral is the Question Mark function. In short, a rich variety can often be found. In the present case, it seems, nothing quite this rich, but getting there.

1.1 Summary

This text is quite long and large, as it is a summary of the results from a multi-year investigation into the β -transform. It ranges over a variety of disparate topics.

Part one continues with a general overview of basic concepts, including the Bernoulli shift, the definition of shift spaces in a relatively abstract fashion, some basics about the β -shift, and a collection of pretty visualizations to anchor and motivate. Part one concludes with a short review of prior research into the β -shift.

Part two defines and works with the transfer operator \mathcal{L}_β for the β -shift. The concept of “analytic algorithmics” is introduced, by analogy to the idea of “analytic combinatorics”. The Heaviside theta function $\Theta(x)$ can be viewed as the algorithmic snippet “if $x > 0$ then 1 else 0”; eigenfunctions of the transfer operator are series summations over $\Theta(\beta t_n - 1)$ for a sequence of iterates t_n . A countable number of eigenvalues λ of \mathcal{L}_β are shown to accumulate on the circle $|\lambda| = 1/\beta$ in the complex plane.

The series summation can be extended to piecewise polynomial segments. In this representation, the spectrum of \mathcal{L}_β is reflected by the polynomial, factoring into components for each monomial term. The resulting spectrum remains discrete, and extends into the region $\beta^{-j-1} < |\lambda| \leq \beta^{-j}$ for generating polynomials of degree j . In the limiting form of this construction, the piecewise components are coherent states, of the form $\sum_k \lambda^k g_k$ for generating functions g_k . These form a continuous spectrum, valid for any $|\lambda| < 1$. There are a lot of these: for any fixed λ , there appears to be a countable number of orthogonal eigenstates.

Yet, the description is incomplete; a comprehensive understanding is elusive. For example, iterating \mathcal{L}_β shows that there are almost-resonances: distributions that are almost, but not quite stable under repeated application of \mathcal{L}_β . These remain uncharacterized. The stability of the spectrum under perturbations is an entirely opaque topic.

Part three takes a close look at those β values which are characterized by finite orbits. These β values have the curious property of being “self-describing”: they occur as roots of a polynomial generated by the orbit of $x = 1$ under the iteration of $t_\beta(x) = \beta x \bmod 1$. The orbit produces a bit-string $b_0 b_1 \dots b_{k-1}$ of binary bits. When these are arranged into a polynomial $p_n(z) = z^{k+1} - b_0 z^k - b_1 z^{k-1} - \dots - b_{k-1} z - 1$, the unique real positive root of $p_n(z)$ is then the β that generated the orbit: thus, “self-describing”. That is, β generates a bitstring for which $p_n(\beta) = 0$.

The polynomial $p_n(z)$ is an odd beast. It is “quasi-cyclotomic”, in that the (complex) roots are approximately evenly distributed in an approximately circular fashion on the complex plane. It is a generator of number sequences that generalize the Fibonacci numbers; indeed, the first polynomial in the series, $p_1(z) = z^2 - z - 1$ has the Golden mean $\varphi = (1 + \sqrt{5})/2$ as the (real, positive) root. The assortment of popular “generalized Fibonacci numbers” are all special cases of this class. There seems to be a rich associated number theory to go with this, but it does not seem to align with any known “conventional” combinatorics. For example, the $p_n(z)$ vaguely resemble polynomials over the field \mathbb{F}_2 , but that is as far as things seem to go: a vague resemblance. It is tempting to guess that concepts and theorems from Galois theory might go through, but they don’t: the polynomials don’t fit that mold in any obvious way. It seems to be a large class, but there is no existing theory that it maps to.

Such self-describing β values are dense in the reals, and can be placed in one-to-one correspondence with the dyadic rationals. This is not as easy as it sounds: not all possible bit-sequences occur, and thus much of the work is to describe the “valid” bit-sequences. An explicit formula is presented: it provides a bijection between the set of

valid bit-sequences and the set of all bit-sequences. This bijection is articulated; since the bit-sequences are countable and dense, and the reals are separable, the mapping extends to a monotonically increasing continuous, nowhere-differentiable function on the reals.

The non-terminating orbits fall into two classes: the ultimately-periodic orbits, and the ergodic orbits. The ultimately-periodic orbits correspond to rationals, but again, not all rationals are “valid”; they don’t form self-describing orbits. The above-mentioned bijection saves the day: it maps all rationals to valid rationals.

Part four explores perturbations of the β -map. The exploration is driven by the idea that the periodic orbits re-materialize as the “islands of stability” in other iterated systems. In the β -map, the periodic orbits are dense in β , but they form a set of measure zero: they are countable. Small perturbations of the β -map give rise to Arnold tongues. These are now finite-sized regions occupied by the periodic orbits: these are the “mode-locked” regions. As noted earlier, the β -map is topologically conjugate to *all* piecewise continuous maps of the unit interval; but this is a very abstract statement. Explicit examples of mode-locking and Arnold tongues illustrates exactly what this conjugacy actually looks like. Literally; this section is filled with pretty pictures.

1.2 Bernoulli shift

The Bernoulli shift, also known as the bit-shift map, the dyadic transform and the full shift, is an iterated map on the unit interval, given by

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

It can be written much more compactly as $b(x) = 2x \pmod{1}$. The symbolic dynamics of this map gives the binary digit expansion of x . That is, write

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

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

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

Attention: n is a subscript on the left, and a superscript on the right! The left is a sequence, the right is an iteration. Using the letter b on both sides is a convenient abuse of notation. Notation will be abused a lot in this text, except when it isn’t. The symbolic dynamics recreates the initial real number:

$$x = \sum_{n=0}^{\infty} b_n(x) 2^{-n-1} \quad (3)$$

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

$$x = 0.b_0b_1b_2\dots$$

is a representation of a real number with a bit-string.

1.3 Bijections

A variety of mathematical objects that can be placed into a bijection with collections of bit-strings, and much of this text is an exploration of what happens when this is done. There will be several recurring themes; these are reviewed here.

The collection of all infinitely-long bit-strings $\{0, 1\}^\omega = \{0, 1\} \times \{0, 1\} \times \dots$ is known as the Cantor space; ω denotes countable infinity, so this is a countable product of repeated copies of two things. Closely related is the Cantor set, which is famously the collection of points $y = \sum_{n=0}^{\infty} b_n(x) 3^{-n-1}$ that results from taking the binary expansion of a real number, and re-expressing it as a base-three expansion. The Cantor set can also be constructed by repeatedly removing the middle-third. If one is careful that the middle-third is always an open set, what remains after a single subtraction is a closed set. What remains after infinite repetition is a “perfect set”, and a key theorem is that this perfect set is identical to the collection of points obtained with the sum above. Bouncing between these two distinct constructions requires the definition of the product topology on Cantor space, and thence the Borel sigma algebra, so that one can work in a consistent way with set complements. These ideas will be reviewed as the need arises.

Associated with Cantor space is the infinite binary tree. Any given location in the tree can be specified by giving a sequence of left-right moves, down the tree, starting at the root. Such left-right moves can (of course) be interpreted as bit-strings. After a finite number of moves, one arrives at a node, and under that node extends another infinite binary tree, just like the original. If one has a function $f(b)$ that is defined on every node b of the binary tree, then one can compare this function to $f(Lb)$ and $f(Rb)$ that result from a left move and a right move. If these are equal to each other, or scale in some way when compared to the original, or if there exist two other functions g_L and g_R such that $f(Lb) = g_L(f(b))$ for all b , and likewise, that $f(Rb) = g_R(f(b))$, then one has fractal self-similarity. More explicitly, whenever one has a pair of commuting diagrams, $f \circ L = g_L \circ f$ and $f \circ R = g_R \circ f$, then one has a dyadic monoid self-symmetry. This is the symmetry of a large class of fractals.

1.3.1 Formalities

The last paragraph is a bit glib, and so some formalities and examples are in order. These are all very straightforward and conventional, almost trivial, belaboring the obvious. Despite the seeming triteness of the next handful of paragraphs, these formal definitions will be needed, so as to avoid future ambiguities and confusions.

Let $\mathbb{M} = \{L, R\}^{<\omega}$ denote the collection of finite-length binary strings. These can be graded by the length v of the string, so that

$$\mathbb{M} = \{L, R\}^{<\omega} = \bigcup_{v=0}^{\infty} \{L, R\}^v = \varepsilon \cup \{L, R\} \cup (\{L, R\} \times \{L, R\}) \cup \dots$$

with ε denoting the empty (zero-length) string. This can be turned into a monoid by defining multiplication as string concatenation: given $\gamma \in \{L, R\}^n$ some sequence of L, R moves of length n , and $\gamma' \in \{L, R\}^m$ some other sequence of length m , then $\gamma\gamma' \in \{L, R\}^{n+m}$ is some other string of length $m+n$.

This set is in bijection to the integers in a straight-forward way. This bijection, written as $\kappa : \{L, R\}^{<\omega} \rightarrow \mathbb{N}$, can be defined recursively, by a simple commuting diagram. Define $\kappa(\varepsilon) = 1$ and ask that $\kappa(L\gamma) = 2\kappa(\gamma)$ and that $\kappa(R\gamma) = 2\kappa(\gamma) + 1$ for every $\gamma \in \{L, R\}^n$. Thus, $\kappa(L) = 2$ and $\kappa(R) = 3$ and LL, LR, RL, RR map to 4, 5, 6, 7 respectively.

This bijection commutes with the canonical L, R moves on the natural numbers. Write these as a pair of functions $L : \mathbb{N} \rightarrow \mathbb{N}$ and $R : \mathbb{N} \rightarrow \mathbb{N}$, defined as $L : m \mapsto 2m$ and $R : m \mapsto 2m + 1$. Interpreting these as strings, one promptly has a pair of commuting diagrams $\kappa \circ L = L \circ \kappa$ and $\kappa \circ R = R \circ \kappa$. True formality would have required writing $\kappa \circ L_s = L_{\mathbb{N}} \circ \kappa$ or even that $\kappa(L_s\gamma) = L_{\mathbb{N}}(\kappa(\gamma))$ in order to remind us that L_s is a string (of length one) concatenated onto some other string γ , while $L_{\mathbb{N}}$ is a map of the natural numbers. It is convenient to drop these labels, as they mostly serve to clutter the text. The intended meaning is always clear from context.

Associated with the set $\{L, R\}^{<\omega}$ is a binary tree \mathbb{B} . It can be defined as a graph of vertexes v_j and edges e_{ij} connecting vertex v_i to vertex v_j . Formally, it is the graph $\mathbb{B} = \{v_i, e_{ij} : i \in \mathbb{N}, j \in \{2i, 2i + 1\}\}$. Every vertex $v_i \in \mathbb{B}$ can be given an integer label: it is just the integer i itself. To formalize this, there is a map $\eta : \mathbb{N} \rightarrow \mathbb{B}$ that provides this labeling. The canonical labeling gives the root node a label of 1, the left and right sub-nodes 2, 3, and so on.

The canonical moves on this binary tree are $L : \mathbb{B} \rightarrow \mathbb{B}$ and $R : \mathbb{B} \rightarrow \mathbb{B}$ defined by $L : v_i \mapsto v_{2i}$ and $R : v_i \mapsto v_{2i+1}$. Just as above, these commute with the left and right moves on the integers. So, $\eta \circ L = L \circ \eta$ and $\eta \circ R = R \circ \eta$.

The pattern repeats. Consider the set \mathbb{D} of dyadic rationals between zero and one. These are fractions that can be written as $(2n + 1)/2^m$ for some non-negative integers m, n . These are in one-to-one correspondence with the integers: there is a canonical bijection is $\delta : \mathbb{D} \rightarrow \mathbb{N}$ given by $\delta : (2n + 1)/2^m \mapsto 2^{m-1} + n$. There are obvious left and right moves, given by $\delta \circ L = L \circ \delta$ and $\delta \circ R = R \circ \delta$. So, for example, $L(1/2) = 1/4$ and $R(1/2) = 3/4$. Viewed as a tree, this places $1/2$ at the root of the tree, and $1/4$ and $3/4$ as the nodes to the left and right.

All of these maps were bijections: they are all invertable. They place elements of all four objects $\mathbb{M}, \mathbb{N}, \mathbb{B}, \mathbb{D}$ in one-to-one correspondence with each other. The commutation of the L, R moves guarantees that the multiplication on \mathbb{M} , i.e. string concatenation, allows elements $\gamma \in \mathbb{M}$ to act on $\mathbb{N}, \mathbb{B}, \mathbb{D}$ in the obvious, intended way.

1.3.2 Example: Julia Sets

As a practical example of the above machinery, consider the Julia set of the Mandelbrot map. Recall the Mandelbrot map is an iterated map on the complex plane, given by $z \mapsto z^2 + c$. The Julia set is the set of points of “where things came from” in the Mandelbrot map; it is the inverse “map” $z \mapsto \pm\sqrt{z - c}$. The word “map” is in scare quotes, as the plus-minus in front of the square root indicate that each z maps to either one of two distinct predecessors. The choice of the plus-minus signs can be interpreted as left-right moves, and so the Julia set can be interpreted as a representation of the binary tree, with it’s elements labeled by integers, or dyadic fractions, or nodes in the binary tree, or strings of ones and zeros.

We take a moment to make this explicit. Fix a point $c \in \mathbb{C}$ in the complex plane.

Define a function $j_c : \mathbb{N} \rightarrow \mathbb{C}$ recursively, by writing $j_c(1) = 0$ and then $j_c(2m) = -\sqrt{j_c(m) - c}$ and $j_c(2m+1) = +\sqrt{j_c(m) - c}$. Equivalently, there are a pair of moves on the complex plane, $L_c : \mathbb{C} \rightarrow \mathbb{C}$ and $R_c : \mathbb{C} \rightarrow \mathbb{C}$ given by $L_c : z \rightarrow -\sqrt{z - c}$ and $R_c : z \rightarrow +\sqrt{z - c}$ which commute with the Julia map: $j_c \circ L = L_c \circ j_c$ and $j_c \circ R = R_c \circ j_c$.

The skeleton of the Julia set is a set of points in the complex plane:

$$\mathbb{J}_c = \{z \in \mathbb{C} : z = j_c(n), n \in \mathbb{N}\}$$

The full Julia set is the closure $\overline{\mathbb{J}_c}$ which includes all of the limit points of \mathbb{J}_c . The set \mathbb{J}_c is countable; the closure $\overline{\mathbb{J}_c}$ is uncountable.

1.3.3 Example: de Rham Curves

The above construction generalizes. Given any pair of functions $f : X \rightarrow X$ and $g : X \rightarrow X$ and a point $x_0 \in X$ for any space X , the dyadic monoid induces a map $j : \mathbb{N} \rightarrow X$ given recursively as $j(1) = x_0$ and $j(2m) = f(j(m))$ and $j(2m+1) = g(j(m))$. As a commuting diagram, one has that $j \circ L = f \circ j$ and $j \circ R = g \circ j$. This in turn induces a set $\mathbb{J} = \{x \in X : x = j(n), n \in \mathbb{N}\}$. If the space X is a topological space, so that limits can be meaningfully taken, then one also has the closure $\overline{\mathbb{J}}$.

In 1957, Georges de Rham notes that if X is a topological space, so that continuity can be defined, and if there exist two points $x_a, x_b \in X$ such that $g(x_a) = f(x_b)$, then the set \mathbb{J} is a continuous curve.[1] The original proof also requires that X be a metric space, and that the two functions f, g be suitably contracting, so that the set \mathbb{J} remains compact. More precisely, so that the Banach fixed-point theorem can be applied, giving two fixed points $x_a = f(x_a)$ and $x_b = g(x_b)$. Without compactness, the curve wanders off to infinity, where conceptions of continuity break down. It is no longer a curve, “out there”.

The continuity condition $g(x_a) = f(x_b)$ and the fixed points have a direct interpretation from the viewpoint of the binary tree \mathbb{B} . Pick a point $x \in X$, any point at all, and make nothing but left moves: the infinite string $LLL \dots$. The map converts this to the iteration $f(f(\dots(x)))$ which converges onto the fixed point $x_a = f(x_a)$. Likewise, all right-moves converge onto $x_b = g(x_b)$. In between, there are all other branches in \mathbb{B} ; but there are also the “gaps” in between the branches.

Consider the two paths $RLLL \dots$ and $LRRR \dots$ down the tree. Both start at the root, but end up at different places. Yet, they are immediate neighbors: there are no other branches “in between” these two. Such immediate neighbors always lie at either end of a “gap”. Each gap is headed up by the root that sits immediately above them, so that each gap can be labeled by the node from which these two distinct branches diverged. The continuity condition asks that these gaps be closed up: the requirement that $g(x_a) = f(x_b)$ is the requirement that the two sides of the central gap converge to the same point. The curve becomes continuous at this point. By self-similarity, each gap in the tree closes up; the curve is continuous at all such gaps.

As a specific example, consider $X = \mathbb{R}$ and $f(x) = x/2$ and $g(x) = (x+1)/2$. The fixed points are $f(0) = 0$ and $g(1) = 1$ and the continuity condition is satisfied: $g(0) = f(1)$. Iteration produces a curve that is just all of the real numbers of the unit interval. This curve is just the standard mapping of the Cantor space to the unit interval:

it is one-to-one for all points that are not dyadic rationals, and it is two-to-one at the dyadic rationals, as the continuity condition explicitly forces the two-to-one mapping.

Note that Julia sets are not de Rham curves: they don't satisfy the continuity criterion.

1.3.4 Shifts

Adjoint to the left and right moves is a shift τ that undoes what the L and R moves do. It cancels them out, so that $\tau \circ L = \tau \circ R = e$ with e the identity function.

Given a string $\gamma \in \mathbb{M}$ of length v , consisting of letters $a_0 a_1 \cdots a_{v-1}$ so that each $a_k \in \{L, R\}$ is a single letter, define the shift $\tau : \mathbb{M} \rightarrow \mathbb{M}$ as the function that lops off a single letter from the front, so that $\tau : a_0 a_1 \cdots a_{v-1} \mapsto a_1 a_2 \cdots a_{v-1}$ is a string that is one letter shorter. This shift is adjoint to the moves L, R , which prepend either L or R to the string. That is, $L : \mathbb{M} \rightarrow \mathbb{M}$ which acts on strings as $L : \gamma \mapsto L\gamma$, and likewise, $R : \gamma \mapsto R\gamma$. Then, taken as functions, $\tau \circ L = \tau \circ R = e$ with e being the identity function on \mathbb{M} , the function that does nothing. The shift τ is only an adjoint, not an inverse, since there is no way to reattach what was lopped off, at least, not without knowing what it was in the first place. Thus $L \circ \tau \neq e \neq R \circ \tau$. The maps L, R were one-to-one but not onto; the map τ is onto but not one-to-one.

The shift can be composed with either κ or with δ , to have the obvious effect. It's handy to introduce a new letter and a new function $T : \mathbb{N} \rightarrow \mathbb{N}$ so that one has $T \circ \kappa = \kappa \circ \tau$ acting as $T(2m) = T(2m+1) = m$. Since τ applied to the empty string returns the empty string, so also $T(1) = 1$. Recycling the same letter $T : \mathbb{D} \rightarrow \mathbb{D}$ and defining it so that $\delta \circ T = T \circ \delta$ one can infer that $Tx = 2x \pmod{1}$ for any $x \in \mathbb{D}$. So, for example, $T(1/4) = T(3/4) = 1/2$.

On the binary tree \mathbb{B} , the shift moves back up the tree, from either the left or the right side. That is, given a vertex $v \in \mathbb{B}$, it is the map $\tau : v_{2m} \mapsto v_m$ and likewise $\tau : v_{2m+1} \mapsto v_m$.

For the Julia set example, it has a meaningful form: $\tau : z = j_c(n) \mapsto z^2 + c$. It re-does what the two Julia set maps undid. It is onto: it maps \mathbb{J}_c into all of \mathbb{J}_c , and likewise $\overline{\mathbb{J}_c}$ onto $\overline{\mathbb{J}_c}$. For the de Rham curve example, it maps the curve back onto itself. In all three examples, these sets are fixed points of τ . Taking $\mathbb{J}_c \subset \mathbb{C}$ as a subset of the complex plane, it is invariant under the action of τ , so that one has $\tau(\mathbb{J}_c) = \mathbb{J}_c$ and likewise $\tau(\overline{\mathbb{J}_c}) = \overline{\mathbb{J}_c}$. Likewise, the de Rham curve stays fixed in X . These are all examples of invariant subspaces.

1.3.5 Completions

The previous section defined a binary tree \mathbb{B} , but this tree is not the "infinite binary tree" alluded to in the opening paragraphs. It is incomplete, in that it does not go "all the way down" to its leaves. It is not compact, in the same sense that the dyadic fractions \mathbb{D} are not compact: the limit points are absent. The Cantor tree $\overline{\mathbb{B}}$ is the closure or completion or compactification of \mathbb{B} ; it contains all infinitely-long branches, all the way down to the "leaves" of the tree. The Cantor tree $\overline{\mathbb{B}}$ is in one-to-one correspondence with the Cantor space $\{0, 1\}^\omega$, and both can be mapped down to the reals on the unit interval,

using eqn 3. None of this is particularly deep, but a few paragraphs to articulate these ideas will help avoid later confusion and imprecision.

To convert letter strings to binary strings, define a function $b : \mathbb{M} \rightarrow \{0, 1\}$ such that it returns 0 if the first letter of a string is L, and otherwise it returns 1. If the string is of zero length, then one has a choice: one can take b to be undefined, or let it be 0, or 1, or introduce a wild-card character $*$ = $0 \vee 1$ denoting “either zero or one”. For the present, any of these choices is satisfactory. The wildcard is appealing when working with the product topology; but, at the moment, we have no topologies at play.

To extract the n 'th letter from a string, define $b_n : \mathbb{M} \rightarrow \{0, 1\}$ as $b_n = b \circ \tau^n$. Thus, given $\gamma \in \mathbb{M}$ of length v , one can create a bitstring $b_0 b_1 \cdots b_{v-1}$. It can be assigned the obvious numerical value:

$$[\delta^{-1} \circ \kappa](\gamma) = \sum_{n=0}^{v-1} b_n(\gamma) 2^{-n-1}$$

Comparing this to eqn 3, the completion is obvious: $\mathbb{M} = \{L, R\}^{<\omega}$ is completed as $\overline{\mathbb{M}} = \{L, R\}^\omega$ so that it contains all strings of infinite length. This is consistent with the completion of the dyadic rationals \mathbb{D} being the entire real unit interval: $\overline{\mathbb{D}} = [0, 1]$. There is no completion $\overline{\mathbb{N}}$ of the countable numbers, at least, not unless one wishes to say that it is the uncountable infinity. This could be done, but then the games gets even more circular, as this completion is just the Cantor set, and we already have that. It seems best to leave $\overline{\mathbb{N}}$ undefined, to avoid circular confusions. The completion $\overline{\mathbb{B}}$ engenders similar confusion. In the original definition, \mathbb{B} was defined as a graph with a countable number of vertexes, each labeled with an integer. This labeling must be abandoned: $\overline{\mathbb{B}}$ is a graph with an uncountable number of vertexes, each labeled by an element from $\overline{\mathbb{M}}$.

The distinction between $\overline{\mathbb{M}}$, $\overline{\mathbb{B}}$ and $\{0, 1\}^\omega$ becomes hard to maintain at this point: they are all isomorphic. The distinction between $\overline{\mathbb{M}}$ and $\{0, 1\}^\omega$ is particularly strained: they are both collections of strings in two symbols. The primary purpose of trying to maintain this distinction is to remind that $\overline{\mathbb{M}}$ should be thought of as a collection of actions that can be applied to sets, while $\{0, 1\}^\omega$ is a set, a collection of points that sometimes act as labels for things. This distinction is useful for avoiding off-by-one mistakes during calculations; it is a notational convenience. This is a variant of common practice in textbooks: after showing that two things are isomorphic, only rarely is the notation collapsed into one big tangle. One maintains a Rosetta Stone of different ways of writing the same thing. And so here: a distinction without a difference.

1.4 Shift space

The shift τ was defined above as an operator that takes a sequence of letters, and lops off the left-most symbol, returning a new sequence that is the remainder of the string. A shift space S is any subset of the set of all infinite strings that remains invariant under the shift: $\tau S = S$.

In general settings, one considers a vocabulary of N letters, and the set of infinite sequences N^ω , so that a shift space $S \subseteq N^\omega$ is a subset of the “full shift” N^ω (which is

trivially invariant under τ). Shifts that are proper subsets of a full shift will be called subshifts. For the Bernoulli shift, there are $N = 2$ letters, and the Bernoulli shift is by definition the full shift $2^\omega = \{0, 1\}^\omega$. A trivial example of a subshift that is not a full shift is the set $S = \{0^\omega, 1^\omega\}$; it has two elements, and is obviously invariant; both 0^ω and 1^ω are fixed points of τ . Another example is $S = \{(01)^\omega, (10)^\omega\}$, where $(01)^\omega = 0101 \dots$ is a repeating periodic string. Any collection of such periodic strings forms a subshift.

Clearly, the union of two subshifts is a subshift, and so, to classify subshifts, one wants to find all of the indecomposable pieces, and characterize those. Factors include periodic strings of fixed period; but not all of these are unique: so, the period-4 string $(0101)^\omega$ is really just the period-two string in disguise.

Subshifts consisting entirely of periodic strings can be characterized in terms of Lyndon words. Lyndon words are fixed length strings that are not decomposable into shorter sequences. Thus, each one characterizes a periodic subshift. Cyclic permutations of a Lyndon word give the same subshift; for example, both $(01)^\omega$ and $(10)^\omega$ belong to the same subshift. The number of distinct, unique subshifts of length ν is given by Moreau's necklace counting function: it counts the number of distinct sequences of a given length, modulo cyclic permutations thereof.

Characterizing the subshifts that do not consist of periodic orbits is considerably harder. For example, consider the string $s = 010^\omega$. It has an orbit: $\tau s = 10^\omega$ and $\tau^2 s = 0^\omega$ and so one can write down a set $\langle s \rangle = \{010^\omega, 10^\omega, 0^\omega\}$ which has the property that $\tau \langle s \rangle \subset \langle s \rangle$. However, it is not a subshift, because $\tau \langle s \rangle \neq \langle s \rangle$. The first two points "wander away" under the application of the shift; they are part of the "wandering set". What remains is the fixed point $\tau 0^\omega = 0^\omega$. The ergodic decomposition theorem states that all such sets X having the property that $\tau X \subset X$ can be decomposed into two pieces: $X = S \cup W$ with S a subshift, $\tau S = S$ and W the wandering set or dissipative set, that eventually dissipates into the empty set: $\lim_{n \rightarrow \infty} \tau^n W = \emptyset$. Subshifts are fixed-points; everything else disappears.

Given some real number $x \in [0, 1]$, and its binary expansion $x = 0.b_0b_1 \dots$, defined in eqn 2, what is the nature of $\langle s \rangle$ for $s = b_0b_1 \dots$? That is, defining

$$\langle s \rangle = \{\gamma = \tau^n s : n \in \mathbb{N}\}$$

what portion of $\langle s \rangle$ is wandering, and which part is a subshift? If $x = p/q$ is a rational number, the answer is easy: rational numbers have binary expansions that are eventually periodic. They consist of some finite-length prefix of non-repeating digits, followed by an infinite-length cyclic orbit. The finite-length prefix is the wandering set; the cyclic part is a subshift. If the period of the cyclic part is ν , then the subshift contains precisely ν elements.

For the Bernoulli shift $\tau x = 2x \pmod{1}$, for the real numbers, the answer is provided by the ergodic theorem. For all real numbers $x \in [0, 1] \setminus \mathbb{Q}$, that is, the unit interval excluding the rationals, the orbit of x is ergodic: given any real number $\varepsilon > 0$ and any $y \in [0, 1]$ there exists some $n \in \mathbb{N}$ such that $|y - \tau^n x| < \varepsilon$. Iteration takes x arbitrarily close to any location on the unit interval. In terms of symbolic dynamics, the binary expansion of y and the binary expansion of $\tau^n x$ will have $m = \lfloor \log_2 \varepsilon \rfloor$ digits in common. The number m can be made arbitrarily large; the subsequence will occur somewhere in

the expansion. Put differently, every finite-length string $\gamma \in \{0, 1\}^{<\omega}$ occurs as a prefix of (uncountably many) of the strings in $\langle x \rangle$.

In essence, that takes care of that, for the Bernoulli shift, at least, if one is looking at it from the point of view of point dynamics. As long as one thinks with the mind-set of points and their orbits, there is not much more to be said. The above is, in a sense, a complete description of the Bernoulli shift. But the introduction to this text gave lie to this claim. If one instead looks at the shift in terms of its transfer operator acting on distributions, then much more can be said. The spectrum of the transfer operator is non-trivial, and the eigenfunctions are fractal, in general. This will be examined more carefully, later; but for now, the topic of point dynamics in the Bernoulli shift is exhausted. This is the end of the line.

The Bernoulli shift is not the only shift on Cantor space. And so, onward ho.

1.5 Beta shift

The beta shift 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

$$T_\beta(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} \\ \beta(x - \frac{1}{2}) & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (4)$$

This map, together with similar maps, is illustrated in figure 6 below.

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

$$k_n = \begin{cases} 0 & \text{if } 0 \leq T_\beta^n(x) < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq T_\beta^n(x) \leq 1 \end{cases} \quad (5)$$

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

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

Written this way, the $T_\beta(x)$ clearly acts as a shift on this sequence:

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

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 so-called “ β -expansion” of a real number x , namely the series

$$x = \frac{1}{2} \sum_{n=0}^{\infty} \frac{k_n}{\beta^n} \quad (6)$$

The bit-sequence that was extracted by iteration can be used to reconstruct the original real number. Setting $\beta = 2$ in eqn 5 gives the Bernoulli shift: $T_2(x) = b(x)$.

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 . The structure of this shift is explored in detail in a later section.

1.6 Associated polynomial

The iterated shift can also be written as a finite sum. A later section will be devoted entirely to the properties of this sum. Observe that

$$T_\beta(x) = \beta \left(x - \frac{k_0}{2} \right)$$

and that

$$T_\beta^2(x) = \beta^2 x - \frac{\beta}{2} (\beta k_0 + k_1)$$

and that

$$T_\beta^3(x) = \beta^3 x - \frac{\beta}{2} (\beta^2 k_0 + \beta k_1 + k_2)$$

The general form is then:

$$T_\beta^p(x) = \beta^p x - \frac{\beta}{2} \sum_{m=0}^{p-1} k_m \beta^{p-m-1} \quad (7)$$

Since the k_m depend on both β and on x , and are only piece-wise continuous functions, this is not a true polynomial. It does provide a polynomial-like representation with a range of interesting properties.

1.7 Density Visualizations

The long-term dynamics of the β -shift can be visualized by means of a bifurcation diagram. The idea of a bifurcation diagram gained fame with the Feigenbaum map (shown in figure 5). The same idea is applied here: track orbits over long periods of time, and see where they go. This forms a density, which can be numerically explored by histogramming. This is shown in figure 2.

When this is done for the β -shift, one thing becomes immediately apparent: there are no actual “bifurcations”, no “islands of stability”, no extended period-doubling regions, none of the stuff normally associated with the Feigenbaum map. Although there are periodic orbits, these form a set of measure zero: the iteration produces purely chaotic motion for almost all values of x and all values of $\beta > 1$. In this sense, the beta transform provides a clean form of “pure chaos”,¹ 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. The unit interval is divided 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 $j/N \leq b_\beta^n(x) < (j+1)/N$, then increment the count for the j 'th bin. After a total of M iterations, let $c(j;M)$ be the count in the

¹Formal mathematics distinguishes between many different kinds of chaotic number sequences: those that are ergodic, those that are weakly or strongly Bernoulli, weakly or strongly mixing. The beta transform is known to be ergodic,[2] weakly mixing[3] and weakly Bernoulli.[4]

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:

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

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. By definition, $\rho(y)$ is a probability distribution:

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

Probability distributions are the same thing as measures; they assign a density to each point on the unit interval. It can be shown that 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 for a selection of fixed β . 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 . A related visualization is in 3, which highlights the discontinuities in 2. Periodic orbits appear wherever the tracteries in this image intersect. A characterization of these orbits occupies a large portion of this text.

1.8 Tent Map

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

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

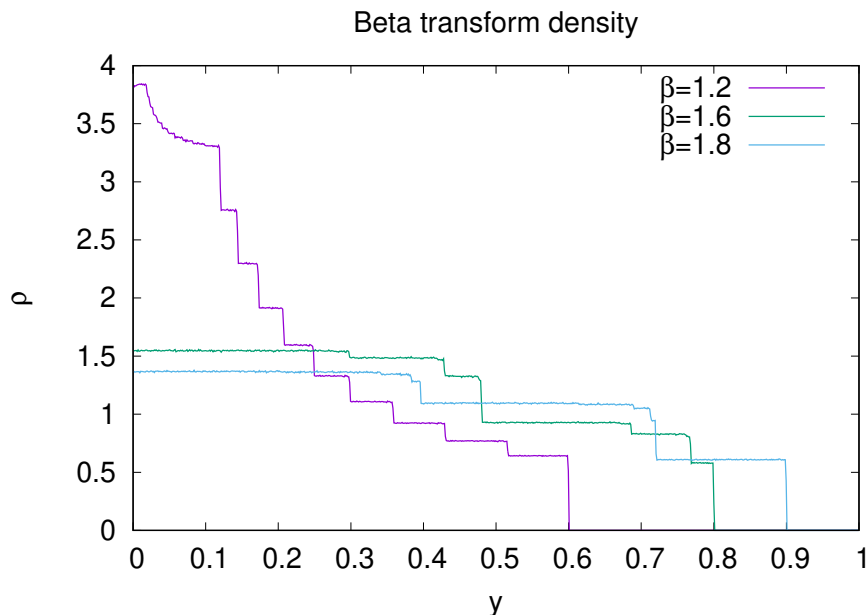
Its similar to the beta shift, except that the second arm is reflected backwards, forming a tent. The bifurcation diagram is shown in figure 4. Its is worth contemplating the similarities between this, and the corresponding beta shift diagram. Clearly, there are a number of shared features.

1.9 Logistic Map

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

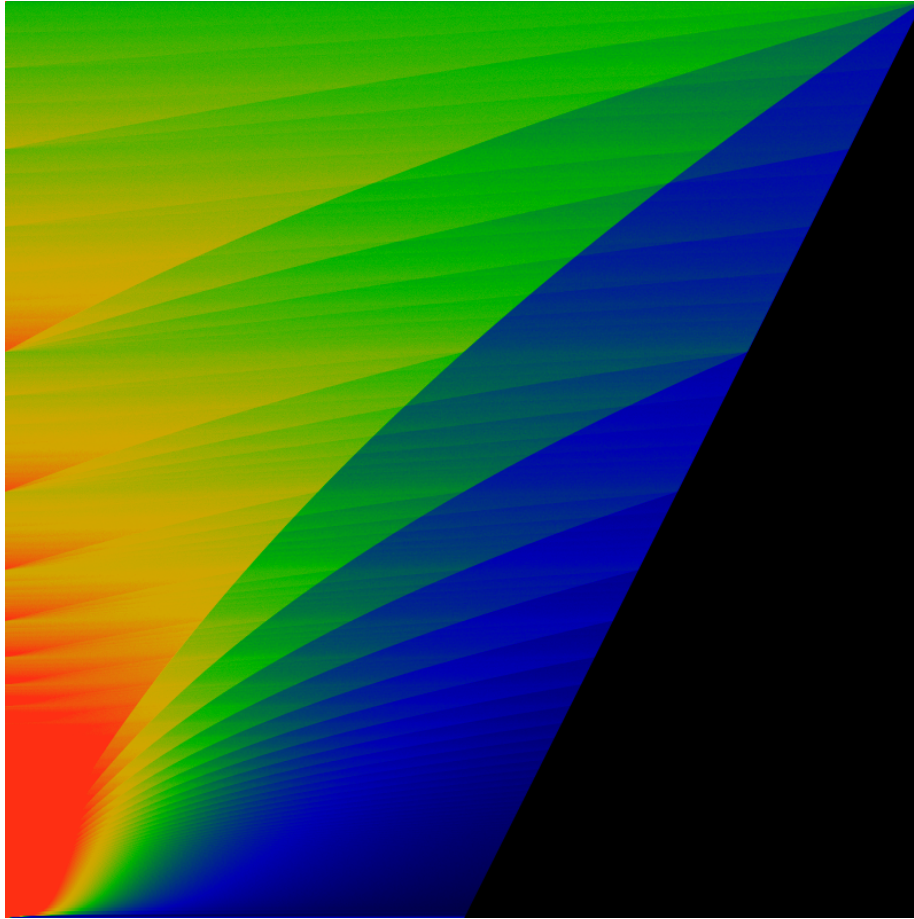
$$f_\beta(x) = 2\beta x(1-x)$$

Figure 1: Beta-shift Density Distribution



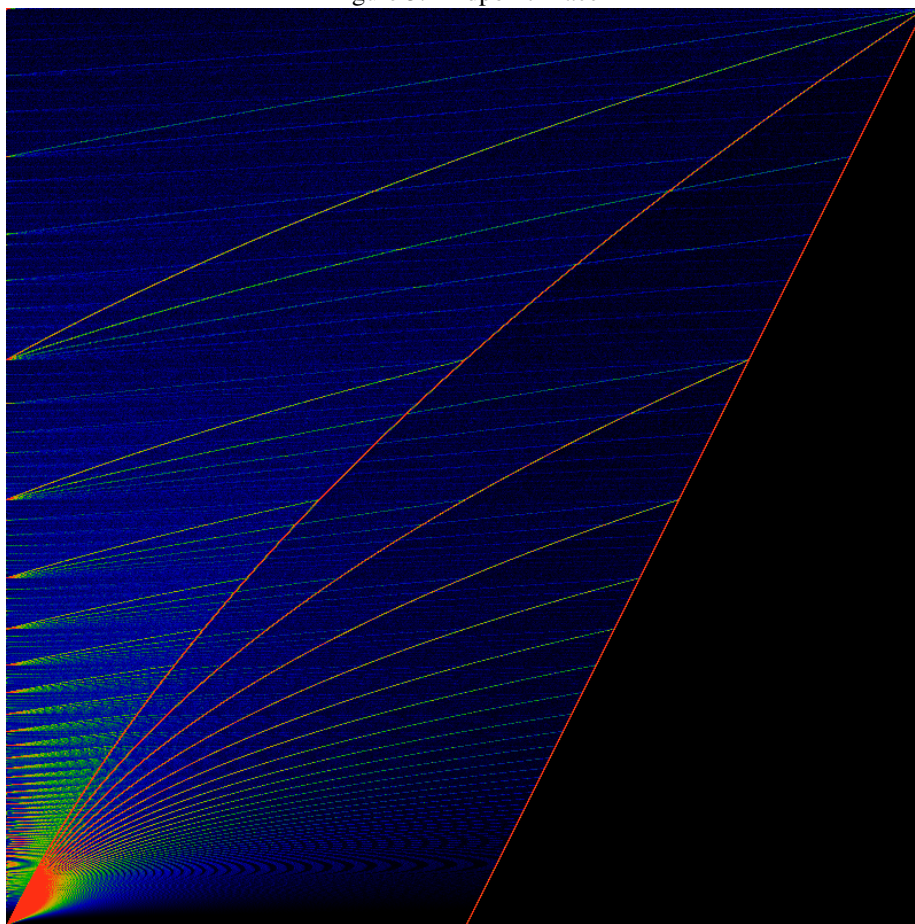
The above figure shows three different density distributions, for $\rho_{1.2}(y)$, $\rho_{1.6}(y)$ and $\rho_{1.8}(y)$, calculated numerically. These are obtained by histogramming a large number of point trajectories, as described in the text. The small quantities of jitter are due to a finite number of samples. To generate this figure, a total of $M = 4000$ iterations were performed, using randomly generated arbitrary-precision floats (using the Gnu GMP package), partitioned into $N = 800$ bins, and sampled 24000 times (or 30 times per bin) to perform the averaging integral. It will later be seen that the discontinuities in this graph occur at the “iterated midpoints” $m_p = T_\beta^p(\beta/2)$. The flat plateaus are not quite flat, but are filled with microscopic steps. There is a discontinuous step at every p ; these are ergodically distributed, *i.e.* dense in the interval, so that there are steps everywhere. This is the general case; for special cases, when the midpoint has a finite orbit, then there are a finite number of perfectly flat plateaus. The first such example occurs at $\beta = (1 + \sqrt{5})/2 = \varphi$ the Golden Ratio. In this case, there are only two such plateaus.

Figure 2: Beta-shift Bifurcation Diagram



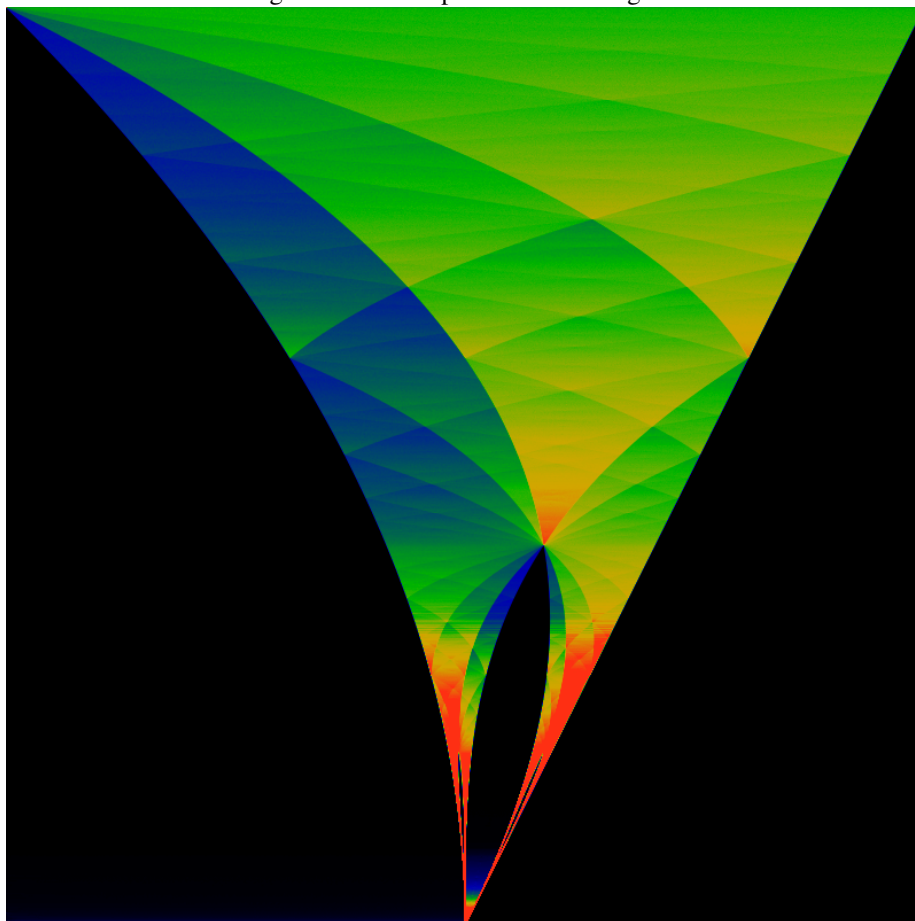
This figure shows the density $\rho_\beta(y)$, rendered in color. The constant β is varied from 1 at the bottom to 2 at the top; whereas y runs from 0 on the left to 1 on the right. Thus, a fixed value of β corresponds to a horizontal slice through the diagram. The color green represents values of $\rho_\beta(y) \approx 1.0$, while red represents $\rho_\beta(y) \gtrsim 2$ and blue-to-black represents $\rho_\beta(y) \lesssim 0.5$. The lines forming the fan shape are not straight. The precise form is examined in a later section, given by a variant of the polynomial in eqn 7. The discontinuities in this figure are more clearly highlighted in the next figure, 3.

Figure 3: Midpoint Trace



Traces of midpoint iteration. Each horizontal line corresponds to a fixed β , with β running from 1 at the bottom, to 2 at the top. At each fixed β , the midpoint $x = 1/2$ is iterated to generate $T_\beta^n(1/2)$. At each such location (from left to right, of 0 to 1), the corresponding pixel is given a color assignment, fading from red, through a rainbow, to black, as n increases. This is a variant of 2, highlighting the edges. A formal analysis of the tracers begins with figure 11

Figure 4: 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 right, the straight boundary indicates that after one iteration, points in the domain $\beta/2 \leq x \leq 1$ are never visited. To the left, points in the domain $0 \leq x \leq \beta(1 - \beta/2)$ are never visited more than a finite number of times.

The figure can be imagined to be a superposition of a countable number of copies of figure 2, each drawn out so as to terminate in a point, but separated into distinct arms. Each copy is recapitulated in the Feigenbaum bifurcation diagram.

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 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 5. 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 the β -transform.

The general visual similarity between the figures 2, 4 and 5 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: all three of these maps are topologically conjugate to one-another. This is perhaps surprising, but is based on the observation that the “islands of stability” in the logistic map are countable, and are in one-to-one correspondence with certain “trouble points” in the iterated beta transformation. These are in turn in one-to-one correspondence 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 (Arnold tongues) appear in the circle map. This is examined in a later section; it is mentioned here only to whet the appetite.

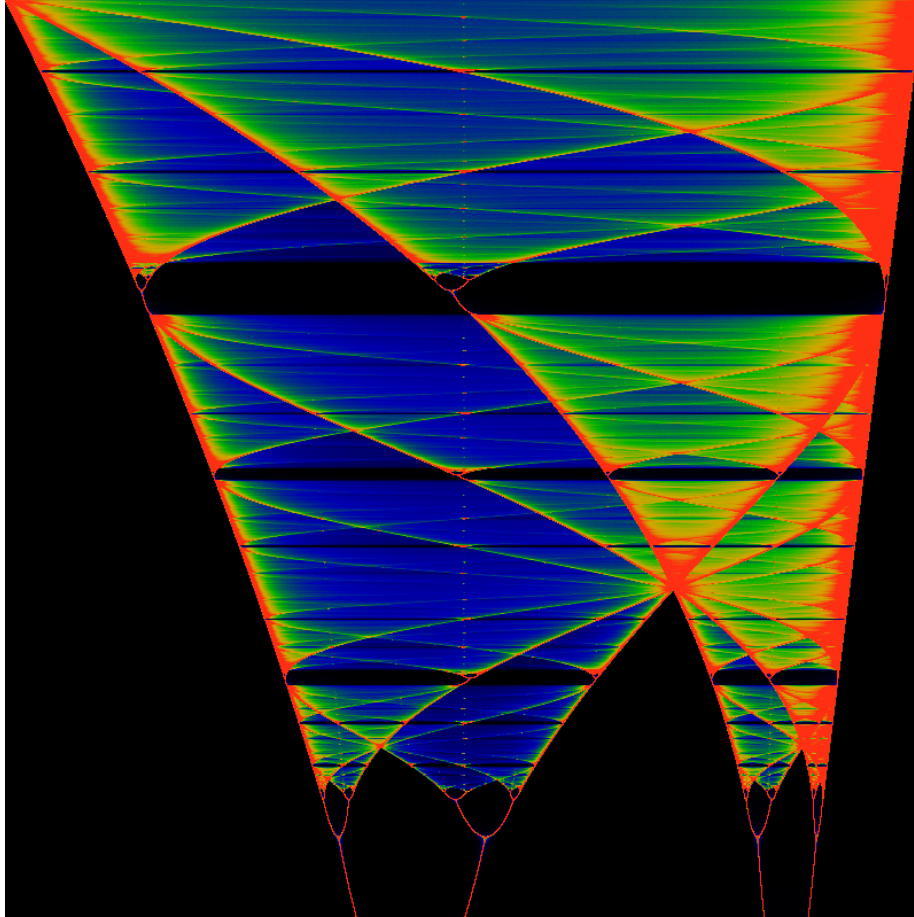
1.10 Beta Transformation

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

$$t_\beta(u) = \begin{cases} \beta u & \text{for } 0 \leq u < \frac{1}{\beta} \\ \beta u - 1 & \text{for } \frac{1}{\beta} \leq u \leq 1 \end{cases} \quad (8)$$

This can be written more compactly as $t_\beta(x) = \beta x \pmod{1}$. In this form, the function is named “the beta-transform”, written as the β -transformation, presenting a typesetting challenge to search engines when used in titles of papers. The orbit of a point x in the beta-shift is identical to the orbit of a point $u = 2x/\beta$ in the beta-transformation.

Figure 5: 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.

This is a very widely reproduced diagram. The goal here is not to waste space reproducing it yet again, but to draw attention to the similarities between this diagram, and figure 4. The tent can be smoothly deformed into the parabola; doing so separates the multiple legs visible at the bottom of 4, with each leg becoming a bifurcation branch. The central point, where all legs come together, is preserved in both images. The curving arcs, highlighted in figure 3, are retained as well. The legs of 4 are chaotic regions, and remain so as the tent is deformed into a parabola. They are merely separated from one-another. The legs do *not* shrink to zero width.

Explicitly comparing to the beta-shift of eqn 4:

$$T_{\beta}^n(x) = \frac{\beta}{2} t_{\beta}^n\left(\frac{2x}{\beta}\right) \quad (9)$$

The beta-shift and the β -transformation are essentially the same function; this text works almost exclusively with the beta-shift, and is thus idiosyncratic, as it flouts the more common convention of working with the β -transformation. The primary reason for this is a historical quirk, as this text was started before the author became aware of the β -transformation.

After a single iteration of the tent map, a similar situation applies. After one iteration, all initial points $\beta/2 \leq x \leq 1$ are swept up into the domain $0 \leq x < \beta/2$. After a finite number of iterations, all points $0 < x \leq \beta(1 - \beta/2)$ 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

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

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

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

The tent map (and the flipped tent) exhibits fixed points (periodic orbits; mode-locking) 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”:

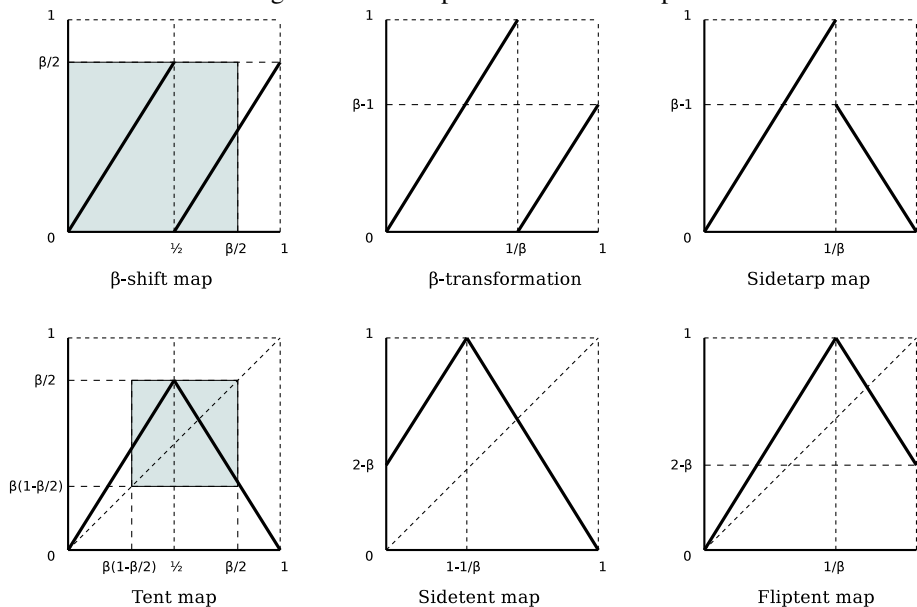
$$a_{\beta}(u) = \begin{cases} \beta u & \text{for } 0 \leq u < \frac{1}{\beta} \\ \beta(1-u) & \text{for } \frac{1}{\beta} \leq u \leq 1 \end{cases}$$

The six different maps under consideration here are depicted in figure 6. It is interesting to compare three of the bifurcation diagrams, side-by-side. These are shown in figure 7.

1.11 Dynamical Systems

A brief review of dynamical systems is in order, as it provides a coherent language with which to talk about and think about the beta-shift. The technical reason for this is that a subshift $S \subset \{0, 1\}^{\omega}$ provides a more natural setting for the theory, and that a lot of the confusion about what happens on the unit interval is intimately entangled with the homomorphism 3 (or 6 as the case may be). Disentangling the subshift from the homomorphism provides a clearer insight into what phenomena are due to which component.

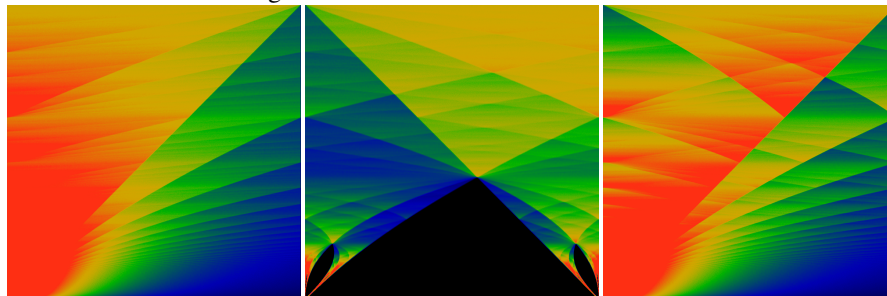
Figure 6: Iterated piece-wise linear maps



The beta shift map, shown in the upper left, generates orbits that spend all of their 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 beta shift, except that one arm is flipped to make a tent-shape. After a finite number of iterations, orbits move entirely 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 prominent fixed points and periodic orbits, essentially because the diagonal intersects the map. The top three maps have periodic orbits, but these occur only for a countable number of β values. General orbits are purely chaotic, essentially because the diagonal does not intersect them. Note that the slopes and the geometric proportions of all six maps are identical; they are merely rearrangements of the same basic elements.

Figure 7: Beta transform and Side-tent



The left figure shows the bifurcation diagram for the β -transform, as it is normally defined as the $\beta x \bmod 1$ map. It is the same map as the beta shift, 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 4, 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.

The review of dynamical systems here is more-or-less textbook-standard material; it is included here only to provide a firm grounding for later discussion.

The Cantor space $\{0, 1\}^{\omega}$ can be given a topology, the product topology. The open sets of this topology are called “cylinder sets”. These are the infinite strings in three symbols: a finite number of 0 and 1 symbols, and an infinite number of * symbols, the latter meaning “don’t care”. Set union is defined location-by-location, with $0 \cup * = 1 \cup * = *$ and set intersection as $0 \cap * = 0$ and $1 \cap * = 1$. Set complement exchanges 0 and 1 and leaves * alone: $\bar{0} = 1$, $\bar{1} = 0$ and $\bar{*} = *$. The topology is then the collection of all cylinder sets. Note that the intersection of any finite number of cylinder sets is still a cylinder set, as is the union of an infinite number of them. The product topology does *not* contain any “points”: strings consisting solely of just 0 and 1 are *not* allowed in the topology. By definition, topologies only allow finite intersections, and thus don’t provide any way of constructing “points”. Of course, points can always be added “by hand”, but doing so tends to generate a topology (the “box topology”) that is “too fine”; in particular, the common-sense notions of a continuous function are ruined by fine topologies. The product topology is “coarse”.

The Borel algebra, or sigma-algebra, takes the topology and also allows set complement. This effectively changes nothing, as the open sets are still the cylinder sets, although now they are “clopen”, as they are both closed and open.

Denote the Borel algebra by \mathcal{B} . A shift is now a map $T : \mathcal{B} \rightarrow \mathcal{B}$ that lops off the leading symbol of a given cylinder set. This provides strong theoretical advantages over working with “point dynamics”: confusions about counting points and orbits and defining densities go away. This is done by recasting discussion in terms of functions $f : \mathcal{B} \rightarrow \mathbb{R}$ from Borel sets to the reals (or the complex numbers \mathbb{C} or other fields, when

this is interesting). An important class of such functions are the measures. These are functions $\mu : \mathcal{B} \rightarrow \mathbb{R}$ that are positive, and are “compatible” with the sigma algebra, in that $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$ and (for product-space measures) that $\mu(A \cap B) = \mu(A)\mu(B)$ for all $A, B \in \mathcal{B}$. The measure of the total space $\Omega = \{0, 1\}^\omega$ is by convention unity: $\mu(\Omega) = 1$.

The prototypical example of a measure is the Bernoulli measure, which assigns probability p to any string containing a single 0 and the rest all *’s. By complement, a string containing a single 1 and the rest all *’s has probability $1 - p$. The rest follows from the sigma algebra: a cylinder set consisting of m zeros and n ones has measure $p^m(1 - p)^n$. It is usually convenient to take $p = 1/2$, the “fair coin”; the Bernoulli process is a sequence of coin tosses.

The map given in equation 3 is a homomorphism from the Cantor space to the unit interval. It extends naturally to a map from the Borel algebra \mathcal{B} to the algebra of intervals on the unit interval. It is not an isomorphism: cylinder sets are both open and closed, whereas intervals on the real number line are either open, or closed (or half-open). It is convenient to take the map as a map to closed intervals, so that it’s a surjection onto the reals, although usually, this detail does not matter. What does matter is if one takes $p = 1/2$, then the Bernoulli measure is preserved: it is mapped onto the conventional measure on the real-number line. Thus, the cylinder set $0****\dots$ is mapped to the interval $[0, 1/2]$ and $1****\dots$ is mapped to $[1/2, 1]$ and both have a measure of $1/2$ and this extends likewise to all intersections and unions. Points have a measure of zero. That is, the homomorphism 3 preserves the fair-coin Bernoulli measure.

Much of what is said above still holds for subshifts. Recall, a subshift S is a subspace $S \subset \{0, 1\}^\omega$ that is invariant under the shift T , so that $TS = S$. The space S inherits a topology from $\{0, 1\}^\omega$; this is the subspace topology. The Borel algebra \mathcal{B} is similarly defined, as are measures. One can now (finally!) give a precise definition for an invariant measure: it is a measure μ such that $\mu \circ T^{-1} = \mu$, or more precisely, for which $\mu(T^{-1}(\sigma)) = \mu(\sigma)$ for almost all cylinder sets $\sigma \in S$. This is what shift invariance looks like. Note carefully that T^{-1} and not T is used in the definition. This is because T^{-1} is a surjection while T is not: every cylinder set σ in the subshift “came from somewhere”; we want to define invariance for all σ and not just for some of them.

The T^{-1} is technically called a “pushforward”, and it defines a linear operator \mathcal{L}_T on the space \mathcal{F} of all functions $f : \mathcal{B} \rightarrow \mathbb{R}$. It is defined as $\mathcal{L}_T : \mathcal{F} \rightarrow \mathcal{F}$ by setting $\mathcal{L}_T : f \mapsto f \circ T^{-1}$. It is obviously linear, in that $\mathcal{L}_T(af + bg) = a\mathcal{L}_T(f) + b\mathcal{L}_T(g)$. This pushforward is canonically called the “transfer operator” or the “Ruelle-Frobenius-Perron operator”. Like any linear operator, it has a spectrum. The precise spectrum depends on the space \mathcal{F} .

The canonical example is again the Bernoulli shift. For this, we invoke the inverse of the mapping of eqn 3 so that $f : [0, 1] \rightarrow \mathbb{R}$ is a function defined on the unit interval, instead of $f : \mathcal{B} \rightarrow \mathbb{R}$. When \mathcal{F} is the space of real-analytic functions on the unit interval, that is, the closure of the space of all polynomials in $x \in [0, 1]$, then the spectrum of \mathcal{L}_T is discrete. It consists of the Bernoulli polynomials $B_n(x)$ corresponding to an eigenvalue of 2^{-n} . That is, $\mathcal{L}_T B_n = 2^{-n} B_n$. Note that $B_0(x) = 1$ is the invariant measure on the full shift. For the space of square-integrable functions $f : [0, 1] \rightarrow \mathbb{R}$,

the spectrum of \mathcal{L}_T is continuous, and consists of the unit disk in the complex plane; the corresponding eigenfunctions are fractal. Even more interesting constructions are possible; the Minkowski question mark function provides an example of a measure on $\{0, 1\}^\omega$ that is invariant under the shift defined by the Gauss map $h(x) \mapsto \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$. That is, as a measure, it solves $\mathcal{L}_T ?' = ?'$ with $?$ the Minkowski question mark function, and $'$ its derivative; note that the derivative is “continuous nowhere”. This rather confusing idea (of something being “continuous nowhere”) can be completely dispelled by observing that it is well-defined on all cylinder sets in \mathcal{B} and is finite on all of them – not only finite, but less than one, as any good measure must obey.

These last examples are mentioned so as to reinforce the idea that working with \mathcal{B} instead of the unit interval $[0, 1]$ really does offer some strong conceptual advantages. They also reinforce the idea that the Bernoulli shift is not the only “full shift”. In the following text, we will be working with subshifts, primarily the beta-shift, but will draw on ideas from the above so as to make rigorous statements about measurability and invariance, without having to descend into either *ad hoc* hand-waving or provide painfully difficult (and confusing) reasoning about subsets of the real-number line.

1.12 Beta Transformation Literature Review and References

The β -transformation, in the form of $t_\beta(x) = \beta x \bmod 1$ has been well-studied over the decades. The beta-expansion 5 was introduced by A. Renyi[2] in 1957, who demonstrates the existence of the invariant measure. The ergodic properties of the transform were proven by W. Parry[3] in 1960, who also shows that the system is weakly mixing.

An explicit expression for the invariant measure was obtained independently by A.O. Gel'fond[5] in 1959, and by W. Parry[3], as a summation of step functions

$$v_\beta(y) = \frac{1}{F} \sum_{n=0}^{\infty} \frac{\varepsilon_n(y)}{\beta^n} \quad (10)$$

where ε_n is the digit sequence

$$\varepsilon_n(y) = \begin{cases} 1 & \text{if } y < t_\beta^n(1) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

and F is a normalization constant. By integrating $\varepsilon_n(y)$ under the sum, the normalization is given by

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

Analogous to the way in which a dyadic rational $p/2^n$ has two different binary expansions, one ending in all-zeros, and a second ending in all-ones, so one may also ask if and when a real number x might have more than one β -expansion (for fixed β). In general, it can; N. Sidorov shows that almost every number has a continuum of such expansions![6] This signals that the beta shift behaves rather differently from the Cantor set in its embedding.

Conversely, the “univoque numbers” are those values of β for which there is only one, unique expansion for $x = 1$. These are studied by De Vries.[7]

The β -transformation has been shown to have the same ergodicity properties as the Bernoulli shift.[4] The fact that the beta shift, and its subshifts are all ergodic is established by Climenhaga and Thompson.[8]

An alternative to the notion of ergodicity is the notion of universality: a β -expansion is universal if, for any given finite string of bits, that finite string occurs somewhere in the expansion. This variant of universality was introduced by Erdős and Komornik[9]. It is shown by N. Sidorov that almost every β -expansion is universal.[10] Conversely, there are some values of β for which rational numbers have purely periodic β -expansions;[11] all such numbers are Pisot numbers.[12]

The symbolic dynamics of the beta-transformation was analyzed by F. Blanchard[13]. A characterization of the periodic points are given by Bruno Maia[14]. A discussion of various open problems with respect to the beta expansion is given by Akiyama.[15]

When the beta expansion is expanded to the entire real-number line, one effectively has a representation of reals in a non-integer base. One may ask about arithmetic properties, such as the behavior of addition and multiplication, in this base - for example, the sum or product of two β -integers may have a fractional part! Bounds on the lengths of these fractional parts, and related topics, are explored by multiple authors.[16, 17, 18]

Certain values of β - generally, the Pisot numbers, generate fractal tilings,[19, 20, 21, 11, 15] which are generalizations of the Rauzy fractal. An overview, with common terminology and definitions is provided by Akiyama.[22] The tilings, sometimes called (generalized) Rauzy fractals, can be thought of as living in a direct product of Euclidean and p -adic spaces.[23]

The set of finite beta-expansions constitutes a language, in the formal sense of model theory and computer science. This language is recursive (that is, decidable by a Turing machine), if and only if β is a computable real number.[24]

The zeta function, and a lap-counting function, are given by Lagarias[25]. The Hausdorff dimension, the topological entropy and general notions of topological pressure arising from conditional variational principles is given by Daniel Thompson[26]. A proper background on this topic is given by Barreira and Saussol[27].

None of the topics or results cited above are made use of, or further expanded on, or even touched on in the following. This is not intentional, but rather a by-product of different goals.

2 Transfer operators

Given any map from a space to itself, mapping points to points, the pushforward maps distributions to distributions. The pushforward is a linear operator, called the transfer operator or the Ruelle–Frobenius–Perron operator. The spectrum of this operator, broken down into eigenfunctions and eigenvalues, can be used to understand the time evolution of a given density distribution. The invariant measure is an eigenstate of this operator, it is the eigenstate with eigenvalue one. There are other eigenstates; these are explored in this section.

Restricting to the unit interval, given an iterated map $f : [0, 1] \rightarrow [0, 1]$, the transfer operator acting on a distribution $\rho : [0, 1] \rightarrow \mathbb{R}$ is defined as

$$[\mathcal{L}_f \rho](y) = \sum_{x=f^{-1}(y)} \frac{\rho(x)}{|f'(x)|}$$

The next subsection gives an explicit expression for this, when f is the β -transform. After that, a subsection reviewing the invariant measure, and then a discussion of some other eigenfunctions.

2.1 The β -shift Transfer Operator

This text works primarily with the β -shift, instead of the more common β -transform. These two are more-or-less the same thing, differing only by scale factors, as given in eqn. 9. The transfer operators are likewise only superficially different, being just rescalings of one-another; both are given below.

The transfer operator the beta-shift map $T_\beta(x)$ is

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

or, written more compactly

$$[\mathcal{L}_{\beta,f}](y) = \frac{1}{\beta} \left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y}{\beta} + \frac{1}{2}\right) \right] \Theta\left(\frac{\beta}{2} - y\right) \quad (12)$$

where Θ is the **Heaviside step function**. The transfer operator for the beta-transform map $t_\beta(x)$ is

$$[\mathcal{M}_{\beta,f}](y) = \frac{1}{\beta} \left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y}{\beta} + \frac{1}{\beta}\right) \right] \Theta(\beta - 1 - y)$$

The density distributions graphed in figure 1 are those functions satisfying

$$[\mathcal{L}_{\beta}\mu](y) = \mu(y) \quad (13)$$

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

$$\mu(y) = \frac{1}{\beta} \left[\mu\left(\frac{y}{\beta}\right) + \mu\left(\frac{y}{\beta} + \frac{1}{2}\right) \right] \Theta\left(\frac{\beta}{2} - y\right) \quad (14)$$

Likewise, the Gelfond-Parry measure of eqn 10 satisfies

$$[\mathcal{M}_\beta v_\beta](y) = v_\beta(y)$$

Recall that $\mu(y) = \frac{2}{\beta} v_\beta\left(\frac{2y}{\beta}\right) \Theta\left(\frac{\beta}{2} - y\right)$; the two invariant measures are just scaled copies of one-another. Both are normalized so that $\int_0^1 \mu(y) dy = \int_0^1 v_\beta(y) dy = 1$.

Both of these invariant measures are the Ruelle-Frobenius-Perron (RFP) eigenfunctions of the corresponding operators, as they correspond to the largest eigenvalues of the transfer operators, in each case being the eigenvalue one.

More generally, one is interested in characterizing the spectrum

$$[\mathcal{L}_\beta \rho](y) = \lambda \rho(y)$$

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 piece-wise continuous smooth functions (piece-wise polynomial functions), any of the Banach spaces, and in particular, the space of square-integrable functions. In general, the spectrum will be complex-valued: eigenvalues will be complex numbers.

If a distribution $\rho(y)$ is nonzero on the interval $[\beta/2, 1]$, the operator \mathcal{L}_β will map it to one that is zero on this interval. Thus, it makes sense to restrict oneself to densities that are nonzero only on $[0, \beta/2]$. When this is done, eqn 12 has the slightly more convenient form

$$[\mathcal{L}_\beta f](y) = \frac{1}{\beta} \left[f\left(\frac{y}{\beta}\right) \Theta(m_0 - y) + f\left(\frac{y}{\beta} + \frac{1}{2}\right) \Theta(m_1 - y) \right]$$

with $m_0 = \beta/2$ and $m_1 = \beta(\beta - 1)/2 = T_\beta(m_0)$. It is always the case that $m_1 < m_0$ and so the second term above vanished on the interval $[m_1, m_0]$. This can be gainfully employed in a variety of settings; typically to write $\mathcal{L}_\beta f$ on $[m_1, m_0]$ as a simple rescaling of $\mathcal{L}_\beta f$ on $[0, m_1]$.

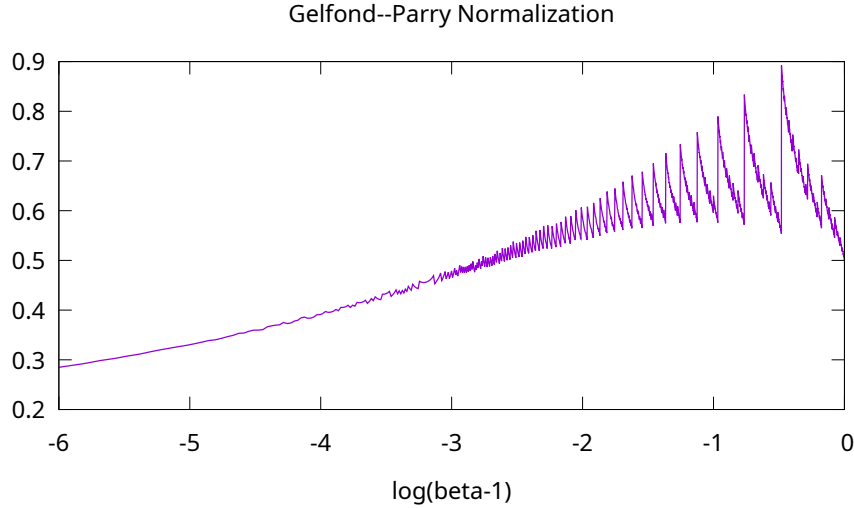
This equation can be treated as a recurrence relation; setting $\mathcal{L}_\beta f = f$ gives the $\lambda = 1$ eigenstate. Performing this recursion gives exactly the densities shown in figure 1. Computationally, these are much cheaper to compute than trying to track a scattered cloud of points; the result is free of stochastic sampling noise. This density is the Ruelle-Frobenius-Perron eigenstate; an explicit expression was given by Gelfond and by Parry, as described in the next section.

2.2 The Gelfond-Parry measure

An explicit expression for the solution to $\mathcal{M}_\beta v = v$ was given by Gelfond[5] and by Parry[3]. It is the expression given by eqn 10. Unfortunately, I find the Russian original of Gelfond's article unreadable, and Parry's work, stemming from his PhD thesis, is not available online. Therefore, it is of some interest to provide a proof suitable for the current text. A generalization of this proof, stated in terms of a Borel algebra, is used in the subsequent section to construct general eigenfunctions.

There are two routes: either a direct verification that eqn 10 is correct, or a derivation of eqn 10 from geometric intuition. The direct verification is useful for practical

Figure 8: Gelfond–Parry Normalization



The above figure shows $1/F$ for the normalization constant $F = \sum_{n=0}^{\infty} t_n \beta^{-n}$ as a function of β . The horizontal axis is stretched out using $\log(\beta - 1)$ so as to amplify the behavior as $\beta \rightarrow 1$. One has that $F \rightarrow \infty$ in this limit (so $1/F \rightarrow 0$); the curve suggests just how catastrophic that limit is. A graph of F vs. β , without the rescaling of the horizontal axis, is shown in figure 13.

purposes; the geometric construction, as a stretch-cut-stack map, provides insight. Both are given.

The Gelfond–Parry measure includes a normalization factor. It will be of recurring interest, and so a graph of it is presented in figure 8.

2.2.1 Direct verification

A direct verification of correctness is done below, explicitly showing all steps in laborious detail. It's not at all difficult; just a bit hard on the eyes.

As before, let $t(x) \equiv t_{\beta}(x) = \beta x \pmod{1}$ be the β -transformation of eqn 8, and $t^n(x)$ the iterated transformation. Let $\Theta(x)$ be the Heaviside step function as always, and to keep notation brief, let $t_n \equiv t^n(1)$. The Gelfond-Parry measure is then

$$v(y) = \frac{1}{F} \sum_{n=0}^{\infty} \frac{\Theta(t_n - y)}{\beta^n}$$

where the normalization F is given by

$$F = \sum_{n=0}^{\infty} \frac{t_n}{\beta^n}$$

The transfer operator \mathcal{M} for the beta-transformation is slightly more convenient to work with than \mathcal{L} for this particular case. It is given by

$$[\mathcal{M}f](y) = \frac{1}{\beta} \left[f\left(\frac{y}{\beta}\right) + f\left(\frac{y+1}{\beta}\right) \Theta(\beta-1-y) \right]$$

and we wish to verify that $\mathcal{M}v = v$. Plugging in directly,

$$\begin{aligned} P &= v\left(\frac{y}{\beta}\right) + v\left(\frac{y+1}{\beta}\right) \Theta(\beta-1-y) = \\ &= \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} \left(\Theta\left(t_n - \frac{y}{\beta}\right) + \Theta\left(t_n - \frac{y+1}{\beta}\right) \Theta(\beta-1-y) \right) \\ &= \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} (\Theta(\beta t_n - y) + \Theta(\beta t_n - 1 - y)) \end{aligned}$$

The second line follows from the first, since $\Theta(ax) = \Theta(x)$ for all constants a , and the $\Theta(\beta-1-y)$ can be safely dropped, since $t_n \leq 1$ for all n . These terms simplify, depending on whether t_n is small or large. Explicitly, one has

$$\begin{aligned} \Theta(\beta t_n - 1 - y) &= 0 \text{ if } \beta t_n - 1 < 0 \\ \Theta(\beta t_n - y) &= 1 \text{ if } \beta t_n - 1 > 0 \end{aligned}$$

and so

$$P = \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} (\Theta(1 - \beta t_n) \Theta(\beta t_n - y) + \Theta(\beta t_n - 1) (1 + \Theta(\beta t_n - 1 - y)))$$

These can be collapsed by noting that

$$\begin{aligned} \beta t_n &= t_{n+1} \text{ if } \beta t_n - 1 < 0 \\ \beta t_n - 1 &= t_{n+1} \text{ if } \beta t_n - 1 > 0 \end{aligned}$$

and so

$$\begin{aligned} P &= \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} (\Theta(1 - \beta t_n) \Theta(t_{n+1} - y) + \Theta(\beta t_n - 1) (1 + \Theta(t_{n+1} - y))) \\ &= \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} (\Theta(t_{n+1} - y) [\Theta(1 - \beta t_n) + \Theta(\beta t_n - 1)] + \Theta(\beta t_n - 1)) \\ &= \frac{1}{F} \sum_{n=0}^{\infty} \frac{1}{\beta^n} (\Theta(t_{n+1} - y) + \Theta(\beta t_n - 1)) \\ &= \beta v(y) - \frac{\beta}{F} + \frac{1}{F} \sum_{n=0}^{\infty} \frac{\Theta(\beta t_n - 1)}{\beta^n} \\ &= \beta v(y) \end{aligned}$$

The last sum on the right is just the β -expansion for 1. That is, the β -expansion of x is

$$x = \sum_{n=0}^{\infty} \frac{\Theta(\beta t^n(x) - 1)}{\beta^{n+1}}$$

This is just eqn 6 written in a different way (making use of the equivalence 9). Thus $P = \beta v(y)$ and so $\mathcal{M}v = v$ as claimed.

2.2.2 Stretch–Cut–Stack Map

The measure can be geometrically constructed and intuitively understood as the result of the repeated application of a bakers-map-style stretch and squash operation. The idea is to iterate $v_{n+1} = \mathcal{M}_\beta v_n$, starting with $v_0 = 1$ and then take the limit $n \rightarrow \infty$; the result is $v_n \rightarrow v$ in the limit, given as above. The basic stretch-cut-stack operation provides intuition as to why v is invariant. The proof that it converges as desired is nearly identical to the (shorter) proof above.

Consider a first approximation that v_0 is constant on the interval $[0, 1]$, so that $v_0(y) = 1$. The operation of \mathcal{M} acting on this is to stretch it out to the interval $[0, \beta]$, chop off the $[1, \beta]$ part, move it to $[0, \beta - 1]$, stack it on top, thus doubling the density in this region. The doubling, though, is partly counteracted by the stretching, which thins out the density to $1/\beta$ uniformly over the entire interval $[0, \beta]$. This operation preserves the grand-total measure on the unit interval. Writing $\beta - 1 = t_1 = t(t_0) = t(1)$ for the first iterate of the endpoint $t_0 = 1$, this stretch, cut and stack operation should result in

$$v_1(y) = \begin{cases} \frac{2}{\beta} & \text{for } y \in [0, t_1] \\ \frac{1}{\beta} & \text{for } y \in [t_1, t_0] \end{cases}$$

The stretch–cut–stack operation is the intuitive, geometrical explanation for obtaining the first iterate. The same result is obtained algebraically, by writing $v_1 = \mathcal{M}v_0$ and then plugging and chugging:

$$v_1(y) = \frac{1}{\beta} [1 + \Theta(t_1 - y)]$$

The derivation of a recursive formula for v_n follows along the line of the algebraic proof given in the previous section, this time, limiting the sums to finite order. The general form is

$$v_n = \rho_n - \sum_{k=0}^{n-1} c_{n-k} v_k$$

which is preserved under iteration as $v_{n+1} = \mathcal{M}v_n$. The two parts are ρ_n defined as

$$\rho_n = \sum_{k=0}^n \frac{1}{\beta^k} \Theta(t_k - y)$$

This converges, up to overall normalization, to the invariant measure: $\rho_n \rightarrow Fv$ with F the normalization constant from before. The c_n are constants, defined as

$$c_n = 1 - \sum_{k=0}^{n-1} \frac{\Theta(\beta t_k - 1)}{\beta^{k+1}} \quad (15)$$

The sum in c_n is a partial β -expansion for 1. Each term can be recognized as a bit from the bit-expansion:

$$\Theta(\beta t_k - 1) = d_n \left(\frac{1}{2} \right) = k_n \left(\frac{\beta}{2} \right) = \varepsilon_n \left(\frac{1}{\beta} \right)$$

with k_n as defined in eqn 5, ε_n defined as in eqn 11 and d_n as in eqn 19 (apologies for the variety of notations; each is “natural” in a specific context.) The β -expansion of a real number x is as given in eqn 6. In the present context, this is the expansion for $x = \beta/2$, or, after rescaling, the expansion for $x = 1$:

$$1 = \sum_{k=0}^{\infty} \frac{\Theta(\beta t_k - 1)}{\beta^{k+1}}$$

Thus, $|c_n| < \beta^{-n+1}$ and $c_n \rightarrow 0$ as $n \rightarrow \infty$. The sum in c_n is a partial β -expansion for 1. Thus, $|c_n| < \beta^{-n+1}$ and $c_n \rightarrow 0$ as $n \rightarrow \infty$.

Each iteration preserves the measure; this is built into the construction. That is, $\int_0^1 v_n(y) dy = 1$ for all n . Plugging through gives a curious identity:

$$1 = \int_0^1 v_n(y) dy = \sum_{k=0}^n \frac{t_k}{\beta^k} - \sum_{k=1}^n c_k$$

which holds for all n . This is readily verified by further plugging through; the underlying identity is that

$$t_n = \beta^n c_n$$

This follows by noticing that each $\Theta(\beta t_k - 1)$ records a decision to decrement, or not, the product βt_k occurring in the iteration $t_{k+1} = \beta t_k \pmod 1$.

The Gelfond–Parry normalization was the $n \rightarrow \infty$ limit of the first sum: $F = \sum_{k=0}^{\infty} \beta^{-k} t_k$.

2.2.3 Stacking Generic Functions

A slightly different calculation results if one works with an arbitrary function $v_0(y)$. The iteration to obtain $v_n = \mathcal{M}^n v_0$ proceeds much as before, with only minor modifications needed to obtain the general form.

The trick is to track two distinct travellers: one that travels with $\Theta(t_k - y)$ and another that travels with the bitsequence $b_k = \Theta(\beta t_k - 1)$. The general solution has the form

$$\begin{aligned} v_n(y) &= a_n(y) + \sum_{k=1}^n h_{nk}(y) \Theta(t_k - y) \\ a_n(y) &= \sum_{k=0}^{n-1} b_k e_{nk}(y) \end{aligned} \tag{16}$$

The functions h_{nk} and e_{nk} can be solved for recursively; the recursive relations are simple enough that they can be rolled up as series summations. These are given by

$$h_{nk}(y) = \frac{1}{\beta^k} a_{n-k} \left(t_1 - \frac{\beta t_k - 1}{\beta^k} + \frac{y}{\beta^k} \right) \quad \text{for } n \geq k$$

which express h_{nk} in terms of the function a_n . This is given as a recursive series, obtained by iterating

$$e_{nk}(y) = \frac{1}{\beta^n} \left[v_0 \left(1 - \frac{t_k}{\beta^k} + \frac{y}{\beta^n} \right) + \Theta(k) \sum_{m=1}^{n-k-1} \beta^m a_m \left(t_1 - \frac{\beta t_k - 1}{\beta^k} + \frac{\beta^m y}{\beta^n} \right) \right]$$

The starting point for iteration is $a_0 = v_0$.

The function a_n is polynomial if v_0 is, and of the same degree; a_n is analytic, if v_0 is, and so on. The function h_{nk} is likewise, on the domain $0 \leq y \leq t_k$; the discontinuities in v_n are entirely due to the $\Theta(t_k - y)$ term.

The proper calculation of the $n \rightarrow \infty$ limit of a_n remains a mystery. This is the primary obstacle to constructing general eigenfunctions from this series.

Verification When $v_0 = 1$, all three functions a_n, e_{nk} and h_{nk} become constants. In this case, $a_n \rightarrow 1/F$ as $n \rightarrow \infty$, with F the normalization constant as before. Defining $\alpha_n = \beta^n a_n$, the recursion relation takes the curious form

$$\alpha_n = 1 + \sum_{k=1}^{n-1} b_k \sum_{m=0}^{n-k-1} \alpha_m$$

This is an integer sequence: each α_n is an integer, starting with $\alpha_0 = \alpha_1 = 1$. This has the form of a generalized Fibonacci sequence. Denote the partial sum as $s_i = \sum_{m=0}^i \alpha_m$. This implies $\alpha_i = s_i - s_{i-1}$ and so

$$\begin{aligned} \alpha_n &= 1 + b_1 s_{n-2} + b_2 s_{n-3} + \cdots \\ s_n &= 1 + s_{n-1} + b_1 s_{n-2} + b_2 s_{n-3} + \cdots \end{aligned}$$

If the orbit is of finite length p , so that $b_k = 0$ for $k > p$, then this can be recast directly into generalized Fibonacci form, by defining $f_n = s_n + \sum_{k=1}^p b_k$. The recursion relation is then

$$f_n = f_{n-1} + b_1 f_{n-2} + \cdots + b_p f_{n-p}$$

For example, the orbit generated by $\beta = \varphi = 1.618 \cdots$ the golden mean has $p = 1, b_1 = 1$ and the sequence is $\alpha_m = 1, 1, 2, 3, 5, 8, 13, \cdots$. Finite orbits and generalized Fibonacci sequences will be treated at length in the next chapter. One of the interesting properties is that $\beta = \lim_{n \rightarrow \infty} f_n / f_{n-1}$ holds in the general case, and not just for the golden mean. The bitsequences are self-describing; this is byproduct from the identity $\beta = \sum_{k=0}^{\infty} b_k \beta^{-k}$.

2.2.4 Scaled iteration

It is convenient to rescale the iterate \mathcal{M}^n by a constant factor, so that instead, one is examining a sequence of functions $\mathcal{M}\mu_n = \lambda\mu_{n+1}$. The iterated operator then has the form

$$\begin{aligned}\mu_n(y) &= c'_n(y) + \sum_{k=1}^n \frac{\Theta(t_k - y)}{(\lambda\beta)^k} c'_{n-k} \left(t_1 - \frac{\beta t_k - 1}{\beta^k} + \frac{y}{\beta^k} \right) \\ c'_n(y) &= v_0(y) + \frac{1}{(\lambda\beta)^n} \sum_{k=1}^{n-1} b_k \left[v_0 \left(1 - \frac{t_k}{\beta^k} + \frac{y}{\beta^n} \right) + \sum_{m=1}^{n-k-1} (\lambda\beta)^m c'_m \left(t_1 - \frac{\beta t_k - 1}{\beta^k} + \frac{\beta^m y}{\beta^n} \right) \right]\end{aligned}$$

This is identical to the earlier forms when $\lambda = 1$. The starting point for iteration is $c'_0 = v_0$. For $v_0 = 1$ and $\lambda = 1$, one regains the Gelfond–Parry distribution in the limit of $n \rightarrow \infty$. In this limit, $c'_n \rightarrow 1/F$.

2.3 Analytic Gelfond–Parry function

The technique above can be repeated verbatim for a “rotated” or “coherent” function

$$v_{\beta;z}(y) = \sum_{n=0}^{\infty} z^n \frac{\Theta(t_n - y)}{\beta^n} \quad (17)$$

for a given complex-valued z . No changes are required, and the result can be read off directly:

$$\begin{aligned}[\mathcal{M}v_{\beta;z}](y) &= \frac{v_{\beta;z}(y)}{z} - \frac{1}{z} + \sum_{n=0}^{\infty} z^n \frac{\Theta(\beta t_n - 1)}{\beta^{n+1}} \\ &= \frac{v_{\beta;z}(y)}{z} + C(\beta; z)\end{aligned}$$

with $C(\beta; z)$ being a constant independent of y . If there are values of β and/or z at which $C(\beta; z) = 0$, then this becomes the eigenequation for \mathcal{M} .

The eigenfunction for \mathcal{L} is the same, up to rescaling of $y \mapsto \beta x/2$. Recycling notation slightly, write

$$v_{\beta;z}(x) = \sum_{n=0}^{\infty} \frac{d_n(x)}{\beta^n} z^n \quad (18)$$

where the $d_n(x)$ are exactly the same digits as defined by Parry, just rescaled for the beta-shift. That is,

$$d_n(x) = \varepsilon_n \left(\frac{2x}{\beta} \right) = \Theta \left(\frac{\beta}{2} t_n - x \right) = \Theta \left(T^n \left(\frac{\beta}{2} \right) - x \right) = \begin{cases} 1 & \text{if } x < T^n \left(\frac{\beta}{2} \right) \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

where T the beta-shift map of eqn 4 and eqn 9 is used. The iterated end-point becomes the iterated midpoint:

$$t^n(1) = \frac{2}{\beta} T^n \left(\frac{\beta}{2} \right)$$

Holding both β and x fixed, the summation is clearly convergent (and holomorphic in z) for complex numbers z within the disk $|z| < \beta$. The eigenequation has the same form:

$$[\mathcal{L}_\beta v_{\beta;z}](x) = \frac{1}{z} v_{\beta;z}(x) + C(\beta; z)$$

where $C(\beta; z)$ is a constant independent of x . Numeric verification reveals we were a bit glib: $C(\beta; z)$ is a constant for $x \leq \beta/2$ and is zero otherwise! (This is normal; \mathcal{L} was defined in such a way that it is always exactly zero for $x > \beta/2$.)

The interesting limit is where $|z| \rightarrow \beta$ and so it's convenient to re-express C in terms of $\zeta = z/\beta$, so that everything is mapped to the unit disk. With some rearrangements, one obtains

$$E(\beta; \zeta) \equiv \zeta \beta C(\beta; \zeta \beta) = -1 + \zeta \sum_{n=0}^{\infty} \zeta^n d_n \left(\frac{1}{2} \right) \quad (20)$$

Given that $d_n(1/2) = \Theta(\beta t_n - 1)$, the above can be recognized as the rotated/coherent form of eqn 15 in the $n \rightarrow \infty$ limit. The primary task is to characterize the zeros of $E(\beta; \zeta)$. This is a straight-forward sum to examine numerically; results will be presented in the section after the next.

2.4 Analytic ergodics

This section proposes that the entire β -subshift can be tied together with a single holomorphic equation. The holomorphic equation effectively provides a continuum (*i.e.* uncountable number) of distinct relationships between different parts of the subshift. This can be interpreted either as a form of interaction across the subshift, or as a kind of mixing. Given the nature of the relationship, the moniker of “fundamental theorem of analytic ergodics” is an amusing name to assign to the result.

The constant term can be independently derived through a different set of manipulations. Explicitly plugging in eqn 18 into the transfer operator yields

$$\begin{aligned} C(\beta; z) &= [\mathcal{L}_\beta v_{\beta;z}](y) - \frac{v_{\beta;z}(y)}{z} \\ &= \sum_{n=0}^{\infty} \frac{z^n}{\beta^n} \left[\frac{1}{\beta} \left[d_n \left(\frac{y}{\beta} \right) + d_n \left(\frac{y}{\beta} + \frac{1}{2} \right) \right] \Theta \left(\frac{\beta}{2} - y \right) - \frac{1}{z} d_n(y) \right] \end{aligned}$$

Replacing z by $\zeta = z/\beta$ gives

$$\begin{aligned} E(\beta; \zeta) &= \zeta \beta C(\beta; \zeta \beta) \\ &= \sum_{n=0}^{\infty} \zeta^n \left[\zeta \left[d_n \left(\frac{y}{\beta} \right) + d_n \left(\frac{y}{\beta} + \frac{1}{2} \right) \right] \Theta \left(\frac{\beta}{2} - y \right) - d_n(y) \right] \end{aligned}$$

This is holomorphic on the unit disk $|\zeta| < 1$, as each individual d_n is either zero or one; there won't be any poles inside the unit disk. Note that $d_n(y) = 0$ for all $y > \beta/2$, and

so one may pull out the step function to write

$$E(\beta; \zeta) = \sum_{n=0}^{\infty} \zeta^n \left[\zeta d_n \left(\frac{y}{\beta} \right) + \zeta d_n \left(\frac{y}{\beta} + \frac{1}{2} \right) - d_n(y) \right] \Theta \left(\frac{\beta}{2} - y \right)$$

confirming the earlier observation that $E(\beta; \zeta)$ vanishes for all $y > \beta/2$.

The bottom equation holds without assuming that $E(\beta; \zeta)$ is independent of y . However, we've already proven that it is; and so a simplified expression can be given simply by picking a specific y . Setting $y = 0$, noting that $d_n(0) = 1$ and canceling terms, one obtains eqn. 20 again.

Staring at the right-hand side of the sum above, it is hardly obvious that it should be independent of y . In a certain sense, this is not “one equation”, this holds for a continuum of y , for all $0 \leq y \leq 1$. It is an analytic equation tying together the entire subshift. For each distinct y , it singles out three completely different bit-sequences out of the subshift, and ties them together. It is a form of mixing. Alternately, a form of interaction: the bit-sequences are not independent of one-another; they interact. This section attempts to make these notions more precise.

The tying-together of seemingly unrelated sequences seems somehow terribly important. It is amusing to suggest that this is a kind of “fundamental theorem of analytic ergodics”.

For such a claim, it is worth discussing the meaning at length, taking the effort to be exceptionally precise and verbose, perhaps a bit repetitive. Equation 4 defined a map, the β -shift. Equation 5 defined a bit-sequence, the β -expansion of a real number $0 \leq x \leq 1$, where equation 6 is the definition of the β -expansion. The set of all such bit-sequences defines the shift. To emphasize this point, its best to compare side-by-side. Copying equation 5, one bitsequence records the orbit of x relative to the midpoint:

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

while a different bitsequence records the orbit of the midpoint, relative to x :

$$d_n(x) = \begin{cases} 1 & \text{if } x < T^n \left(\frac{\beta}{2} \right) = T^{n+1} \left(\frac{1}{2} \right) \\ 0 & \text{otherwise} \end{cases}$$

The iterations are running in opposite directions; this is as appropriate, since the the transfer operator was a pushforward.

It is useful to return to the language of sigma algebras and cylinder sets, as opposed to point dynamics. Recall that the Borel algebra \mathcal{B} was defined as the sigma algebra, the collection of all cylinder sets in the product topology of $\{0, 1\}^\omega$. A subshift is a subset $\mathcal{S} \subset \mathcal{B}$ together with a map $T : \mathcal{S} \rightarrow \mathcal{S}$ that lops off the leading symbol of a given cylinder set but otherwise preserves the subshift: $T\mathcal{S} = \mathcal{S}$. The inverted map T^{-1} is a pushforward, in that it defines the transfer operator, a linear operator $\mathcal{L}_T : \mathcal{F} \rightarrow \mathcal{F}$ on the space \mathcal{F} of all functions $f : \mathcal{S} \rightarrow \mathbb{R}$; explicitly, it is given by $\mathcal{L}_T : f \mapsto f \circ T^{-1}$. Insofar as the d_n arose in the exploration of the transfer operator, it is not surprising that the shift is acting “backwards”.

The problem with the language of point dynamics is that one cannot meaningfully write $T^{-n}(x)$ for a real number, a point x , at least, not without severe contortions that lead back to the Borel algebra. Not for lack of trying; the $T^{-n}(x)$ is called the “Julia set” (to order n) of x : it is the preimage, the set of all points that, when iterated, converge onto x .

Can the analytic relation be restated in terms of cylinder sets? Yes, and it follows in a fairly natural way. The first step is to extend d_n to a map $d_n : \mathcal{S} \rightarrow [0, 1]$ into the unit interval, as opposed to being a single bit. Let $\mu : \mathcal{B} \rightarrow [0, 1]$ be the canonical Bernoulli measure. Using the Bernoulli mapping 3, the interval $[0, T^n(\beta/2)]$ maps to some cylinder; call it Δ_n . Then, given some cylinder $A \in \mathcal{S}$, define

$$d_n(A) = \mu(A \cap \Delta_n)$$

The rotated (pre-)measure is extended likewise:

$$v(A) = \sum_{n=0}^{\infty} \zeta^n d_n(A)$$

with $\zeta = z/\beta$ as before, recovering the Parry measure by setting $z = 1$.

The Parry measure should be invariant under the action of $T^{-1} : \mathcal{S} \rightarrow \mathcal{S}$, and otherwise yield eqn 20. Let’s check. The proof will mirror the one of the previous section. Again, it is convenient to use the β -transform t_β instead of the β -shift T_β . This is primarily a conceptual convenience; the subshift is more easily visualized in terms of the mod 1 map. Otherwise, the same notation is used, but rescaled, so that Δ_n is the cylinder corresponding to the interval $[0, t^n(1)]$.

Recall that for every $A \in \mathcal{S}$ and every $y \in A$, one will find that $y/\beta \in t^{-1}(A)$ and, whenever $y \leq \beta - 1$ that also $(y+1)/\beta \in t^{-1}(A)$. Thus, $t^{-1}(A)$ naturally splits into two parts: the cylinder that maps to $[0, 1/\beta]$, call it D , and the complement \bar{D} .

The pushforward action is then

$$\begin{aligned} v(t^{-1}(A)) &= \sum_{n=0}^{\infty} \zeta^n \mu(\Delta_n \cap t^{-1}(A)) \\ &= \sum_{n=0}^{\infty} \zeta^n [\mu(\Delta_n \cap D \cap t^{-1}(A)) + \mu(\Delta_n \cap \bar{D} \cap t^{-1}(A))] \end{aligned}$$

Two distinct cases emerge. When $t^n(1) < 1/\beta$ then one has that $\Delta_n \cap \bar{D} = \emptyset$. Thus, the second term can be written as

$$\begin{aligned} \mu(\Delta_n \cap \bar{D} \cap t^{-1}(A)) &= \Theta \left(t_n - \frac{1}{\beta} \right) \mu(\Delta_n \cap \bar{D} \cap t^{-1}(A)) \\ &= \Theta \left(t_n - \frac{1}{\beta} \right) \frac{1}{\beta} \mu(\Delta_{n+1} \cap A) \end{aligned}$$

where the second line follows from the first by linearity, and that \bar{D} selected out one of the two branches of $t^{-1}(A)$. Meanwhile, when $t^n(1) > 1/\beta$, then $D \subset \Delta_n$ so that

$D \cap \Delta_n = D$. Thus, the first term splits into two:

$$\begin{aligned}\Theta\left(t_n - \frac{1}{\beta}\right) \mu(\Delta_n \cap D \cap t^{-1}(A)) &= \Theta\left(t_n - \frac{1}{\beta}\right) \mu(D \cap t^{-1}(A)) \\ &= \Theta\left(t_n - \frac{1}{\beta}\right) \frac{1}{\beta} \mu(A)\end{aligned}$$

while

$$\Theta\left(\frac{1}{\beta} - t_n\right) \mu(\Delta_n \cap D \cap t^{-1}(A)) = \Theta\left(\frac{1}{\beta} - t_n\right) \frac{1}{\beta} \mu(\Delta_{n+1} \cap A)$$

Reassembling these pieces and making use of $\Delta_0 \cap A = A$ one gets

$$\begin{aligned}v(t^{-1}(A)) &= \sum_{n=0}^{\infty} \frac{\zeta^n}{\beta} \left[\mu(A) \Theta\left(t_n - \frac{1}{\beta}\right) + \mu(\Delta_{n+1} \cap A) \right] \\ &= \frac{1}{z} v(A) - \frac{\mu(A)}{z} + \mu(A) \sum_{n=0}^{\infty} \frac{\zeta^n}{\beta} \Theta\left(t_n - \frac{1}{\beta}\right) \\ &= \frac{1}{z} v(A) + \frac{\mu(A)}{z} E(\beta; z)\end{aligned}$$

with the constant term as before, in eqn 20:

$$\begin{aligned}E(\beta; z) &= -1 + \zeta \sum_{n=0}^{\infty} \zeta^n \Theta\left(t_n - \frac{1}{\beta}\right) \\ &= -1 + \zeta \sum_{n=0}^{\infty} \zeta^n d_n\left(\frac{1}{2}\right)\end{aligned}$$

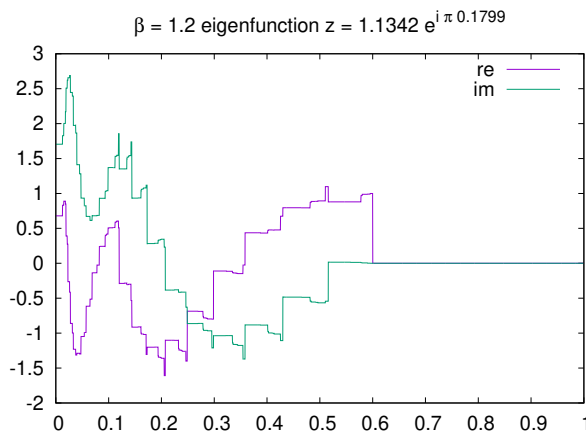
As before, one has for $z = 1$ that $E(\beta; 1) = 0$ and so $v \circ t^{-1} = v$ is indeed the measure invariant under t^{-1} . Other eigenvalues can be found for those values of z for which $E(\beta; z) = 0$. The task at hand is then to characterize $E(\beta; z)$.

2.5 Exploring $E(\beta; z)$

The function is easily explored numerically. It is clearly convergent in the unit disk $|\zeta| < 1$ and has no poles in the disk. For almost all β , there seem to be a countable number of zeros within the disk, accumulating uniformly on the boundary as $|\zeta| \rightarrow 1$. An example is shown in figure 10. The notion of ‘‘uniformly’’ will be made slightly more precise in the next section, where it is observed that, for certain special values of β , the bit-sequence $d_n\left(\frac{1}{2}\right)$ is periodic, and thus is a polynomial. When it is polynomial, there are a finite number of zeros (obviously; the degree of the polynomial), which are distributed approximately uniformly near the circle $|\zeta| = 1$. As the degree of the polynomial increases, so do the number of zeros; but they remain distributed approximately evenly. In this sense, the limit of infinite degree seems to continue to hold.

A handful of selected zeros are listed in the table below. The numbers are accurate to about the last decimal place.

Figure 9: Typical Eigenfunction



A typical eigenfunction $v_{\beta;z}(x)$ solving $\mathcal{L}_\beta v = \lambda v$ for $\beta = 1.2$. This is the eigenfunction as defined in eqn 18. This eigenfunction corresponds to the zero $z = 1.1342 \exp i\pi 0.1799$ of $E(\beta; z)$, alternately of $C(\beta; z)$, as defined in eqn 20. The eigenvalue is $\lambda = 1/z$. Since the eigenvalue is complex, so is the eigenfunction. The real and imaginary parts are paired in a way that vaguely resembles sine and cosine; such phased offsets are generic.

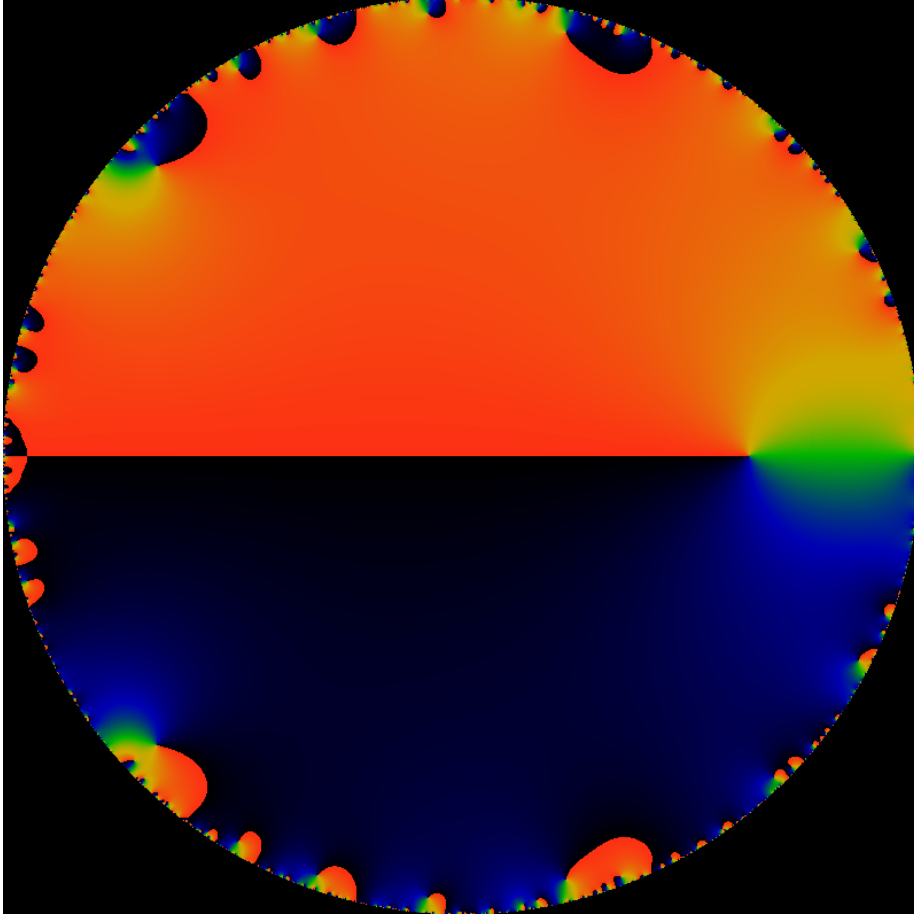
β	z	$ z $	$1/z$
1.8	-1.591567859	1.59156785	-0.6283112558
1.8	-1.1962384 + i 1.21602223	1.70578321	-0.41112138 - i 0.41792066
1.8	0.99191473 + i 1.44609298	1.75359053	0.32256553 - i 0.47026194
1.6	-1.06365138 + i 1.00895989	1.46606764	-0.49487018 - i 0.46942464
1.4	0.55083643 + i 1.17817108	1.30057982	0.32564816 - i 0.69652119
1.2	0.95788456 + i 0.60733011	1.13419253	0.74462841 - i -0.47211874

These are not particularly meaningful numbers; they just give a flavor for some locations of eigenvalues. Given a zero, the corresponding eigenfunction is also very easily computed. A typical eigenfunction is shown in figure 9; this is for the zero listed in the last row of the table above. Although it is unlike the figure 1, in that it is not strictly decreasing, it does have the same general plateau-like regions. Note that all such eigenfunctions are bounded and generally, differentiable-nowhere.

As the zeros accumulate onto the circle $|\zeta| \rightarrow 1$, there appears to be no way to holomorphically continue the function $E(\beta; z)$ outside of the unit circle. This indicates that there is a lower bound on the possible eigenvalues that can be reached via the eigenfunctions of eqn 18. The lower bound is $|\lambda| = 1/\beta$; the eigenfunctions with smaller eigenvalue do not arise from eqn 18.

The apparent reason for this is that the coherent state was constructed as a deformed version of the repeated iteration of $v_0 = 1$ a constant, as described in earlier sections.

Figure 10: Disk of $E(\beta; z)$



The above shows a visualization of $E(\beta; z)$ for $\beta = 1.6$ in the complex $\zeta = z/\beta$ plane. The disk consists of all ζ values with $|\zeta| \leq 1$. The plot is a phase plot, showing the argument $\arg E(\beta; z) \in [-\pi, \pi]$. The color-coding of blue-black puts the phase just above $-\pi$, green near 0, and red just below $+\pi$. Locations where the phase wraps around counter-clockwise (right-handed) are zeros of $E(\beta; z)$; this follows from Cauchy's principle. The most prominent zero at the right-hand side of the image corresponds to $z = 1$, located at $\zeta = 1/\beta$ so well inside the circle. This corresponds to the Gelfond–Parry invariant measure. Other zeros are seen at the end of black-red whiskers, wrapping through green. These accumulate at the $\zeta = 1$ boundary; in general, there are a countable number of such zeros. The corresponding eigenvalues are located at $\lambda = 1/z = 1/\beta\zeta$, so the accumulation ring of zeros occurs at $|\lambda| = 1/\beta$.

Iterating on $v_0(y)$ a polynomial appears to generate eigenfunctions with eigenvalue bounded by $1/\beta^p$ for a polynomial of degree p . One explicit example, generated by a parabola, is given in the next section. A simple, direct construction remains elusive.

A different derivation of this same holomorphic form will be given in the next chapter. There, the beta shift is shown to be conjugate to an infinite shift matrix, in explicit upper-Hessenberg form. Thus, the shift is explicitly a shift on the monomials ζ^n as the basis elements.

2.5.1 Convergence

The analytic structure of the ergodic disk

$$E(\beta; z) = -1 + \zeta \sum_{n=0}^{\infty} b_n \zeta^n$$

is daunting. Clearly, when the bitsequence b_n is not finite, then there is a pole at $\zeta = 1$. There is also a pole at $\zeta = -1$ unless the number of even and odd bits asymptotically cancel out, and then the series is conditionally convergent.

More generally, the series will be conditionally convergent for $\zeta = e^{i\phi}$ only when the bitsequence b_n is ultimately periodic. In this case, $\zeta = e^{i\phi}$ that are roots of unity can be arranged so that they precisely cancel for each period.

For ergodic bitsequences b_n , it is not clear how to formulate a proper conjecture for the sum.

2.5.2 Jentzsch's theorem

Relevant to the present case is Jentzsch's theorem (1914), which states that if

$$g(z) = 1 + a_1 z + \dots + a_n z^n + \dots$$

has a radius of convergence of one, then every point on the unit circle is a cluster-point of zeros of partial sums

$$s_n(z) = 1 + a_1 z + \dots + a_n z^n$$

A proof is given in [28], which generalizes to polynomial sequences that converge on arbitrary regions of the complex plane (having positive capacitance).

Perhaps a simpler statement is that this follows from **Rouché's theorem**, in that $1 - z^n$ will have n roots, located at the roots of unity, and all the other terms only perturb the location of the roots.

2.6 Other eigenfunctions

Eigenfunctions with eigenvalues λ less than $1/\beta$ clearly exist. Below is an example with $\lambda = 1/\beta^2$ for $\beta = \varphi = 1.618\dots$ the golden mean. A general solution in terms of piece-wise polynomial eigenfunctions will be presented several sections hence, after the theory of finite orbits is developed.

2.6.1 Piecewise quadratic eigenfunfuntions

To portray the problem, a piece-wise parabolic presentation is pursued. For $\beta = \varphi$, the function

$$\rho(y) = \begin{cases} \varphi y^2 - y + \frac{1}{8} & \text{for } 0 \leq y < \frac{1}{2} \\ y^2 - y + \frac{\varphi}{8} & \text{for } \frac{1}{2} \leq y < \frac{\varphi}{2} \\ 0 & \text{for } \frac{\varphi}{2} \leq y \leq 1 \end{cases}$$

satisfies $\mathcal{L}_\varphi \rho = \lambda \rho$ for $\lambda = \varphi^{-2}$.

2.6.2 Quasi-resonances

Returning to the expression for $\mathcal{M} \mu_n = \mu_{n+1}$ given in section 2.2.4, one might search for resonances by hypothesizing μ_0 to be some polynomial, and then look for sequences $\lambda^n \mu_n \rightarrow \mu$ that converge uniformly in the $n \rightarrow \infty$ limit. Performing such a search reveals surprises. The following was observed.

Fix $\beta = 1.6$. Then:

- If $\mu_0(x) = x - 1/2$ then μ_n converges to $\approx -0.0869229 \nu$ with ν the Gelfond–Parry invariant measure.
- If $\mu_0(x) = x - 1/2 + 0.0869229$ (so that the convergent above is subtracted), then the μ_n appear to bounce around ergodically, maintaining a bounded norm. Three behaviors are apparent: $\beta^n \int \mu_n = \text{const} \approx 0.2767$. The L_1 and L_2 norms bounce around but remain bounded: $0.3 < \beta^n \int |\mu_n| < 0.8$ and $0.1 < \beta^n \int |\mu_n|^2 < 1.0$. Each iterate is approximately orthogonal to the prior one: $\beta^{2n+1} \int \mu_n \mu_{n+1} \approx 0$. Here, $\int f = \int_0^1 f(x) dx$ is just short-hand notation.

So there are two questions: what is this magic constant 0.0869229? What is this ergodic bouncing?

The answer to the second question is that μ_n has converged onto a superposition of eigenvectors, each having an eigenvalue of modulus $1/\beta$ but differing in phase. Further iteration shows unitary evolution, up to a scale factor. The mixture over-all resembles ergodic behavior. More precisely, assume there are multiple eigenfunctions $\mathcal{M}(\psi_n + i\chi_n) = \frac{1}{\beta} e^{i\phi_n} (\psi_n + i\chi_n)$. Then if $\mu_n \rightarrow \sum_k a_k \psi_k$ further iteration will resemble an ergodic walk:

$$\mathcal{M} \mu_n = \mu_{n+1} = \frac{1}{\beta} \sum_k a_k (\psi_k \cos \phi_k - \chi_k \sin \phi_k)$$

and so the numerics for μ_n will look like a bouncing mess, as the various phase angles ϕ_k make the factors come and go.

At any rate, iteration from a simple starting point cannot cross the barrier-wall of eigenvalues located in a circle of radius $1/\beta$. One becomes trapped at this location. To get past it, one must design an initial μ_0 that is orthogonal to all of these, much as part of the recipe was to design an initial μ_0 that was orthogonal to ν .

If one considers instead a polynomial $\mu_0(x) = x^n + c_{n-1}x^{n-1} + \dots + c_0$, one can find an analogous behavior: $\mu_n \rightarrow Kv$. The leading divergence can be subtracted, so that iterating on $\mu_0(x) = x^n + c_{n-1}x^{n-1} + \dots + c_0 - K$ gives a sequence of μ_n whose length diminishes uniformly, in that $\beta^n \int \mu_n = \text{const}$, but otherwise bounce around ergodically. The relationship between K and the polynomial is opaque. Of course, it can be computed exactly; the polynomial can be inserted into eqn 16, binomials can be expanded with binomial coefficients, and one obtains K . However, these formulas are very tangled; an intuitive description of the manifold or variety is absent.

It has been shown that the beta transform is weak mixing. Is the above a symptom of this? Is this how mixing manifests itself? If so, then it seems that there should be a descriptive framework that makes this explicit.

2.7 Conclusion

It appears that a complete description of the transfer operator remains elusive. For larger eigenvalues, the analytic approach given above appears to be sufficient, at least for the case where the transfer operator acts on a sufficiently tame space of functions. A cohesive description of the spectrum for $|\lambda| < 1/\beta$ remains absent. Worse, there appear to be a variety of almost-resonances that are relatively easy to discover and explore numerically, but defy coherent explanation. Particularly annoying is the lack of a theoretical framework that articulates how mixing manifests itself in the structure of the transfer operator.

3 Finite Orbits

The iteration of the midpoint $m_0 = \beta/2$, that is, the iterated series $m_n = T_\beta^n(\beta/2)$ is ergodic in the unit interval, for almost all values of β . However, for certain values of β , the midpoint iterates will hit the point $x = 1/2$ where the β -shift map has a discontinuity. Here, iteration stops: at the next step, this point is defined to iterate to zero, in eqn 4. Zero is a fixed point, and so there is nowhere further to go. This section explores these special values of β .

Aside from the definition in eqn 4, one can consider the modified map, where the less-than sign has been altered to a less-than-or-equals:

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

In this map, the point $x = 1/2$ iterates to $\beta/2$, which is just the initial midpoint itself. In this case, the halted orbits become periodic orbits.

The β values at which the midpoint has a periodic or terminating orbit will be called “trouble spots”, for lack of a better term. They can be imagined to be prototypes of a bifurcation point, “depending delicately on initial conditions”: where two choices are possible, depending on infinitesimally small perturbations of m_0 , or, alternately, of β .

Before the end of the section, it will be shown that these “trouble spots” are dense in the interval $1 \leq \beta \leq 2$, that they can be placed in a one-to-one bijection with the dyadic rationals, and that this bijection can be extended to the rationals (describing ultimately-periodic orbits) and then the reals. The map is monotonically increasing, continuous but not differentiable; it vaguely resembles the blancmange curve. Of course, it has a peculiar fractal self-similarity, as all maps to the infinite binary tree do.

3.1 The β -generalized Golden Ratio

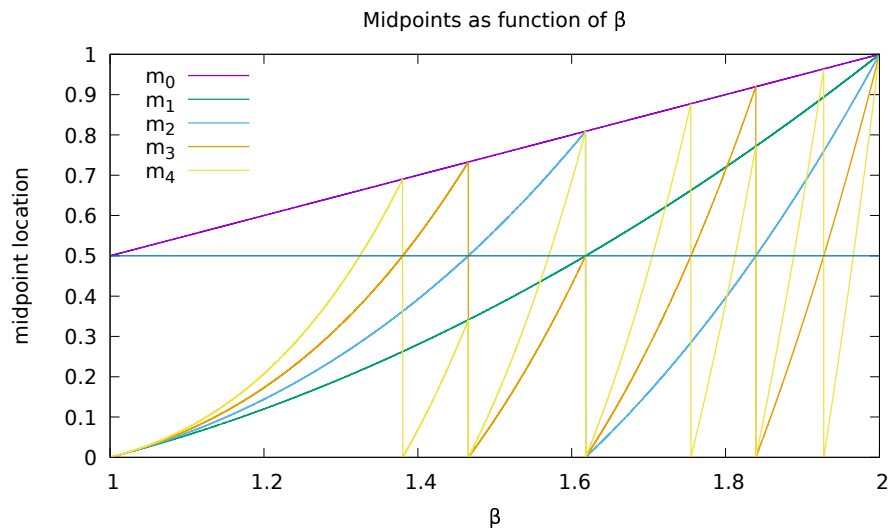
Trouble spots occur whenever the p 'th iterate $m_p = T_\beta^p(m_0)$ lands at the discontinuity, so that one may take either $m_p = 0$ or $m_p = m_0$. The iteration immediately before corresponds to $m_{p-1} = 1/2$. The length of the orbit is p .

The first finite orbit can be found when $\beta = \varphi = (1 + \sqrt{5})/2$ the Golden Ratio. In this situation, one has that $m_0 = \varphi/2$ and $m_1 = 1/2$. The length of the orbit is $v = 2$. For $v = 3$, there are two such 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 $\beta = 1.465571231876768\dots$ and $\beta = 1.839286755214161\dots$.

Where else are such 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 11. As made clear in the image, three new degenerate points appear. These are located at $\beta = 1.380327757\dots$ and $\beta = 1.754877666\dots$ and $\beta = 1.927561975\dots$, 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 suggestion by Dajani[4], the β numbers corresponding to the trouble spots may be called “generalized golden means”. Unfortunately, the term “generalized

Figure 11: Location of Midpoints



This chart illustrates the location of the first five midpoints, m_0, m_1, \dots, m_4 as a function of β . When $m_k = 0.5$, further iteration is ambiguous, as this is the location of the discontinuity in the shift map, and the next iteration leaves the midpoint bifurcated. These are the “trouble spots”. The first trouble spot is visible for $m_1 = 1/2$, corresponding to $\beta = \varphi$ and a length of $\nu = 2$. The midpoint m_2 crosses $1/2$ (in the ascending direction) at $\beta = 1.465\dots$ and $\beta = 1.839\dots$, corresponding to orbits of length $\nu = 3$. It also crosses discontinuously downwards at $\beta = \varphi$. This crossing point has already been assigned to a shorter orbit. The midpoint m_3 has three new crossings. It also rises to touch $1/2$ at $\beta = \varphi$; but this has already been assigned to shorter orbits. The midpoint m_4 has six new crossings. Crossings will generally fall to the left and right of earlier crossings, and so are related in a bracketing relationship. The bracketing is not the full binary tree; it is pruned, as shorter orbit assignments can knock out longer one. For example, m_4 falls back down at the m_2 crossing at $\beta = 1.465\dots$, and so, here, m_4 never even gets close to $1/2$; it won’t bracket until later. This is a formalized version of figure 3, which shows midpoints of all orders.

golden mean” is in common use, and is applied to a variety of different systems. Not all are relevant to the present situation; one that is, is given by Hare *et al.*[29] who provide series expansions for the real roots of $\beta^p - \sum_{k=0}^{n-1} \beta^k = 0$; these are known as the n -bonacci constants (Fibonacci, tribonacci, tetranacci, *etc.*). Stakhov[30] considers $\beta^{p+1} - \beta^p - 1 = 0$ in general settings. Some, but not all of these numbers are known to be Pisot numbers or Salem numbers[14]. In what follows, these will be referred to as the “beta golden means”, since all of the ones that appear here have explicit origins with the beta shift.

3.2 Counting Orbits

How many trouble spots are there? The table below shows the count M_ν of the number of “new” trouble spots, as a function of the orbit length ν .

ν	2	3	4	5	6	7	8	9	10	11	12
M_ν	1	2	3	6	9	18	30	56	99	186	335

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 values are given by Moreau’s necklace-counting function. The trouble spots are the positive real roots of polynomials of the form

$$p_{\{b_k\}}(\beta) = \beta^\nu - \beta^{\nu-1} - b_1\beta^{\nu-2} - b_2\beta^{\nu-3} - \dots - 1 = 0$$

with the $\{b_k\}$ being certain binary bit sequences. There is just one such (positive, real) root for each such polynomial. These polynomials are relatively prime, in the sense that a bit-sequence b_k is disallowed if it has the same root as some lower-order polynomial. For example, $\beta^4 - \beta^3 - \beta - 1$ is disallowed; it has the same root as $\beta^2 - \beta - 1$. Equivalently, the quadratic is a factor of the quartic; the quartic is not relatively prime with respect to the quadratic.

The reason for the appearance of the necklace-counting function is straightforward: it is counting the number of distinct orbits of a given length. An orbit of length ν is, by definition, a point x such that $x = T_\beta^\nu(x)$. Such an orbit generates a binary string, of length ν corresponding to whether $T_\beta^j(x) < 1/2$ is true or not. A cyclic permutation of such a string still corresponds to the same orbit; a reversed permutation does not: thus, it is a necklace without reversal. The necklace-counting function gives the number of distinct, unique orbits of a given length that cannot be factored into shorter orbits. The beta transformation has the property that every possible orbit does occur; none are prohibited.

Yet, the bit-string defining the polynomial is not a necklace; it cannot be rotated. Each bit-string corresponds to a unique polynomial, having roots that differ from those of other polynomials. The polynomials also have a canonical order, fixed as the integer that generates the bit-string; they cannot be reordered. Ideas such as Lyndon words apply to the orbits, but not to the defining polynomials. The ordering of the polynomials is *not* the lexicographic ordering of the Lyndon words, and cannot be brought into this order.

The values of M_n are given explicitly by Moreau’s necklace-counting function

$$M_n = \frac{1}{n} \sum_{d|n} 2^d \mu\left(\frac{n}{d}\right)$$

where the sum runs over all integers d that divide n and μ is the Möbius function. The generating function is

$$\frac{t}{\frac{1}{2} - t} = \sum_{n=1}^{\infty} n M_n \frac{t^n}{1 - t^n}$$

which has a radius of convergence of $|t| < 1/2$. For large n , the asymptotic behavior can be trivially deduced from the defining sum:

$$M_n = \frac{2^n}{n} - \mathcal{O}\left(\frac{2^{n/2}}{n}\right)$$

The above counting function is for necklaces with only two colors. In general, one can have necklaces with 3 or more colors; can that happen here? Yes, of course: if one considers the general β -transform for $2 < \beta$, then, in general, it can be taken as a “kneading transform” with $\lceil \beta \rceil$ branches or folds in it. The analogous trouble-spots again appear, and they can appear after an arbitrary finite-length orbit. Insofar as they correspond to periodic orbits, they are necessarily counted by the necklace-counting function. That is, one must consider all possible strings of $\lceil \beta \rceil$ letters, modulo a cyclic permutation: this is the very definition of a necklace (or “circular word”). The number of such necklaces is given by the necklace-counting function. Each such orbit is necessarily represented by a Lyndon word, which is a representative of the conjugacy class of the orbit.

The isomorphism of different systems described by necklace polynomials is a subject that gets some fair amount of attention. Golomb gives an isomorphism between the irreducible polynomials over \mathbb{F}_p , for p prime and necklaces built from Lyndon words.[31, 32] A number of other results exist, including [33, 34]. At any rate, a closer study of the beta-polynomials seems to be called for.

3.3 β -Golden Polynomials

The “trouble spots” occur whenever the k ’th iterate $m_k = T_\beta^k(m_0)$ of the midpoint $m_0 = \beta/2$ lands on the starting midpoint $m_k = m_0$; alternately, when $m_{k-1} = 1/2$. Because of the piece-wise linear form of T_β , the k ’th iterate will be a piece-wise collection of polynomials, each of order k , each of the form $p_{\{b_k\}}(\beta)$. These must be arranged such that $p_{\{b_k\}}(\beta) = 0$ at each discontinuity, as illustrated in figure 11. This constrains the polynomials that can appear; it constrains the possible coefficients $\{b_k\}$; not all bit-sequences appear. The sequences that do appear encode the orbit of the mid-point; see below.

The table below explicitly shows the polynomials for the first few orders. A polynomial is included in the table if it is an iterate of a previous polynomial, and if it’s

real root is bracketed by the roots of the earlier iterates. Adopting ordinal numbering, $p_n(\beta)$ must have the form

$$p_n(\beta) = \begin{cases} \beta(p_{n/2}(\beta) + 1) - 1 & \text{for } n \text{ even} \\ \beta p_{(n-1)/2}(\beta) - 1 & \text{for } n \text{ odd} \end{cases} \quad (21)$$

This recursion terminates at $p_0(\beta) = \beta - 1$.

The positive real root r_n satisfying $p_n(r_n) = 0$ is unique; the other $n - 1$ roots are complex; they are arranged in a roughly evenly-spaced ring on the complex plane, not far from the unit circle, reminiscent of roots of unity. There is always a positive real root, which satisfies $1 \leq r_n < 2$; the real roots and the polynomials are in one-to-one correspondence. The roots must be bracketed (to the left and right) by the roots of polynomials occurring earlier in the sequence; if the root is not bracketed, then the corresponding polynomial does not appear in the list.

The bracketing constraint can be represented by a recursive function $\theta_n(\rho)$ returning a boolean true/false value, as to whether a given polynomial is acceptable. It is

$$\theta_n(\rho) = \begin{cases} \Theta(r_{n/2} - \rho) \cdot \theta_{n/2}(\rho) & \text{for } n \text{ even} \\ \theta_{\lfloor n/2 \rfloor}(\rho) & \text{for } n \text{ odd} \end{cases} \quad (22)$$

The Heaviside $\Theta(x)$ used here is one for strictly positive $x > 0$, and is zero otherwise. This is important, as using $x \geq 0$ will not work. In numerical work, the test should be bounded away from zero. The recursion compares the candidate ρ to some root at each lower order. The recursion terminates at $\theta_0(\rho) = 1$. The index n corresponds to a valid polynomial, and thus a valid root, if and only if $\theta_n(r_n) = 1$. Effectively, this states that roots of higher-order polynomials must be less than a certain sequence of lower-order roots. This is visible in the location of the discontinuities in figure 11: new discontinuities at higher orders must occur to the left of earlier ones.

For example, the polynomial $\beta^3 - \beta - 1$ is excluded from the list simply because it is not an iterate of an earlier polynomial, even though it has the interesting real root 1.324717957244746..., the ‘‘silver constant’’. The numbering scheme does not even have a way of numbering this particular polynomial. Despite this, the silver constant does appear, but a bit later, as the root of $p_8 = \beta^5 - \beta^4 - 1$, which is an allowed polynomial.

The polynomial $p_5 = \beta^4 - \beta^3 - \beta - 1$ is excluded because it has $\varphi = 1.618\dots$ as a root, which was previously observed by p_1 . The polynomial $p_9 = \beta^5 - \beta^4 - \beta - 1$ is excluded because its root, $r_9 = 1.497094048762796\dots$ is greater than its predecessor r_2 ; the recursive algorithm does not compare it to r_4 . Note that p_9 is relatively prime to the earlier polynomials, so irreducibility is not a sufficient criterion; the root must also be less.

The indexing has the property that, whenever $\theta_n(r_n) = 1$, the integer $2n + 1$, expressed as a binary bitstring, encodes both the coefficients of the polynomial, and also the orbit of the midpoint. This can be taken as an alternate, non-recursive definition of θ_n : it is one if and only if the orbit of r_n encodes the bitsequence of $2n + 1$.

The degree v of the polynomial is identical to the length v of the orbit; it is $v = \lceil \log_2(2n + 1) \rceil$. The bits b_i of the bitstring $2n + 1 = b_0 b_1 b_2 \dots b_v$ correspond to the

orbit as

$$b_i = \Theta \left(T_\beta^i \left(\frac{\beta}{2} \right) - \frac{1}{2} \right) = d_i \left(\frac{1}{2} \right) = k_i \left(\frac{\beta}{2} \right)$$

where the d_i are as given before, in eqn 19, and the k_i as in eqn 5. Note that $b_0 = 1$ always corresponds to $1/2 < \beta/2$, always. By convention, the last digit is always 1, also.

order ν	$p_n(\beta)$	n	binary	root r_n
0	1			
1	β		0	0
	$\beta - 1$	0	1	1
2	$\beta^2 - \beta - 1$	1	11	$\varphi = \frac{1+\sqrt{5}}{2} = 1.618\dots$
3	$\beta^3 - \beta^2 - 1$	2	101	1.465571231876768...
	$\beta^3 - \beta^2 - \beta - 1$	3	111	1.839286755214161...
4	$\beta^4 - \beta^3 - 1$	4	1001	1.380277569097613...
	$\beta^4 - \beta^3 - \beta^2 - 1$	6	1101	1.754877666246692...
	$\beta^4 - \beta^3 - \beta^2 - \beta - 1$	7	1111	1.927561975482925...
5	$\beta^5 - \beta^4 - 1$	8	10001	1.324717957244746...
	$\beta^5 - \beta^4 - \beta^2 - 1$	10	10101	1.570147312196054...
	$\beta^5 - \beta^4 - \beta^3 - 1$	12	11001	1.704902776041646...
	$\beta^5 - \beta^4 - \beta^3 - \beta - 1$	13	11011	1.812403619268042...
	$\beta^5 - \beta^4 - \beta^3 - \beta^2 - 1$	14	11101	1.888518845484414...
	$\beta^5 - \beta^4 - \beta^3 - \beta^2 - \beta - 1$	15	11111	1.965948236645485...

The next table lists the acceptable polynomial indexes for order 5, 6 and 7. Again, the coefficients appearing in the polynomial are encoded by the binary value of $2n + 1$ in the sequence. This sequence has recently been published as [OEIS A368747](#).

order ν	valid indexes
5	8,10,12,13,14,15
6	16,20,24,25,26,28,29,30,31
7	32,36,40,42,48,49,50,52,53,54,56,57,58,59,60,61,62,63

The properties of this sequence are briefly reviewed in the next section.

3.4 Properties of the theta sequence

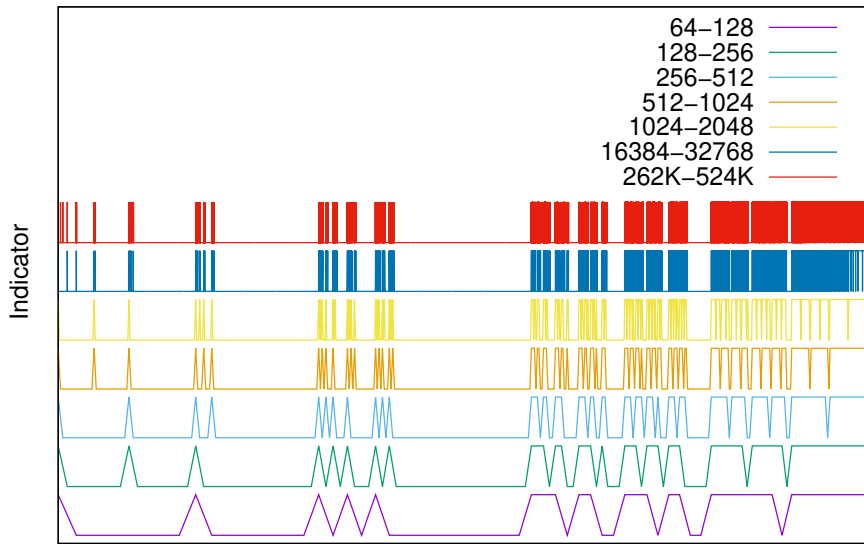
Define the “validity set” of valid or acceptable indexes as

$$\Psi = \{n \in \mathbb{N} : \theta_n(r_n) = 1, p_n(r_n) = 0\} \quad (23)$$

This is just the list of indexes from the previous two tables; it is $\Psi = \{0, 1, 2, 3, 4, 6, 7, 8, 10, 12, 13, 14, 15, 16, 20, 24, 25, 26, 28, 29, 30, 31, \dots\}$. The θ_n is the acceptance function from eqn 22. By abuse of notation, write $\theta_n = \mathbb{1}_\Psi(n) = \theta(n)$ for the membership indicator function for this set. This function is visualized in figure 12.

Figure 12: Indicator function θ

Valid Index Indicator Function



This figure shows the indicator function θ_n , as defined in eqn 23, over the range of $2^6 \leq n \leq 2^{19}$. The function is approximately periodic as a function of the order $v = \lfloor \log_2 n \rfloor + 2$ of the corresponding polynomial; the order is just the length of a finite orbit. This figure just stacks seven such ranges, for orders $v=6,7,8,9,10,14,18$. One can view each row as a comb. The teeth of the comb are the members of Ψ_v . The number of teeth at each rank is given by Moreau's necklace-counting function M_v . The width of the teeth go as 2^{-v} , but the number of teeth goes as $M_v \sim 2^v/v$. In the limit $v \rightarrow \infty$, the teeth converge onto a set of measure zero. This can be understood as a representation of the set Ψ , taken with a real-valued index, rather than an integer index.

The elements of the validity set Ψ are ordered; this is the validity sequence $\psi_m = \psi(m)$. It is convenient to start the sequence at $\psi_0 = 0$. This corresponds to $\beta = 2$ at the right; while, at the far left, for $\beta = 1$, write $\psi_{-1} = \infty$. The function ψ_m encodes the locations of one-bits in the bitmask θ , and so $\theta(\psi_m) = 1$.

The summatory function of the indicator is $S(k) = \sum_{n=1}^k \theta_n$. It counts the total number of one-bits below the location k . The validity sequence is the pullback of the summatory function. Each ψ_m is the smallest integer k for which $m = S(k)$ holds true, and so one has $m = S(\psi_m)$. The pullback can be expressed as $\psi(m) = \psi(S(\psi(m)))$.

A few additional properties may be noted:

- For all m , $\theta(2^m) = \theta(2^m - 1) = 1$.
- If $\theta(m) = 1$ then $\theta(2m) = 1$. By recursion, $\theta(2^n m) = 1$ for all n .
- If m is odd, and if $\theta(m) = 1$, then $\theta((m-1)/2) = 1$. This is reminiscent of the Collatz conjecture.
- If $\theta(m) = 0$ then $\theta(2^n(2m+1)) = 0$ for all n .

Each of these properties is visible in figure 12. The second bullet accounts for the stability of the comb-teeth, once they appear, while the last bullet accounts for the large spaces that open up, and stay open, never filling in.

It is convenient to partition Ψ into ranks v that correspond to the length of the orbits, or equivalently, the order of the defining polynomial. Examining the earlier tables, the rank of n is $v(n) = \lfloor \log_2 n \rfloor + 2$. The partition is then $\Psi = \bigcup_{v=1}^{\infty} \Psi_v$ with $\Psi_v = \{n \in \Psi : v(n) = v\}$. As before, it is convenient to extend the partition so that it can deal with the endpoints $\beta = 1$ and $\beta = 2$; this is a kind of (two-point) compactification of these and other various sequences and sets. The compactification here is to write $v(0) = 1$ and $v(-1) = \infty$, which allows components $\Psi_1 = \{0\}$ and $\Psi_{\infty} = \{-1\}$. The size of each component is $|\Psi_v| = M_v$ given by Moreau's necklace counting function M_v .

The representation of θ as a real number is $\theta = \sum_{n=1}^{\infty} \theta_n 2^{-n} = 0.93258880035365 \dots$. At this time, this does not appear in OEIS.

3.4.1 Leaders

An important subsequence consists of the leaders of the doubling sequences. These can be defined in several equivalent ways. One property of the bitmask is that if $\theta(m) = 1$, then $\theta(2^n m) = 1$ for all n . A leader λ is the smallest such m at the front of such a doubling sequence: it is either an odd number $\lambda = (2k+1)$ satisfying $\theta(\lambda) = 1$ or it is an even number of the form $\lambda = 2^h(2k+1)$, $h > 0$ such that $\theta(2^h(2k+1)) = 1$ but $\theta(2^{h-1}(2k+1)) = 0$. The minimal power h defining the leader will be called the "height of the leader".

Equivalently, the leaders can be defined as a subsequence of ψ_m . By definition, one has that $\theta(\psi_m) = 1$. Thus, if ψ_m is odd, then ψ_m is a leader. If ψ_m is even, then it is a leader if and only if $\theta(\psi_m/2) = 0$.

Given a valid index ψ_m , it is useful to define a function Λ that provides the map

$$\Lambda(\psi_m) = 2^h(2\psi_m + 1) \tag{24}$$

This map is not monotonic: $\Lambda(1) = 3$ and $\Lambda(2) = 10$ but $\Lambda(3) = 7$. It is, however, one-to-one, as leaders are always a product of an odd number times some power of two.

Sorting the leaders into ascending order, the start of the sequence is 1, 3, 7, 10, 13, 15, 25, 29, 31, ... This sequence is not currently known to OEIS. It is handy to count from one, and to assign $\lambda_0 = 0$, so that $\lambda_1 = 1$ and $\lambda_2 = 3$ and so on.

The number of leaders of non-zero height ($h \neq 0$) at each order v , starting at $v = 2$, is 0, 1, 1, 3, 3, 7, 10, 18, 28, 52, 83, 151, 256, 457, 798, 1439, 2549, 4627, 8328, 15190, 27659, 50800, 93270, 172330, 318717, 591996, ... This sequence is not currently known to OEIS.

The number of leaders of a given order v can be defined as the size of the set $\Omega(v) = \{\lambda : v = \lfloor \log_2 \lambda \rfloor, \lambda = \Lambda(\psi_m)\}$. This can be understood as the necklace function, minus the elements of non-zero height, plus those promoted up from a lower grade. This grading is 1, 1, 3, 3, 9, 12, 26, 43, 87, 149, 295, 531, 1021, 1898, 3630, 6822, 13062, 24783, 47481, 90699, 174165, 334148, 643306, 1238619, 2390213, ... This sequence is not currently known to OEIS.

The number of leaders of a given order v with a non-zero height is the size of the set $\{\lambda : v = \lfloor \log_2 \lambda \rfloor, \lambda = \Lambda(\psi_m), 0 < h\}$. This sequence is 0, 0, 1, 0, 3, 1, 6, 5, 16, 15, 43, 52, 116, 173, 348, 551, 1079, 1816, 3432, 6031, 11267, 20226, 37706, 68773, 128135, ... This sequence is not currently known to OEIS.

3.5 Location of β -Golden Roots

The location of the roots can be visualized by using the normalization of the Parry–Gelfond measure. The function in eqn 10 or more generally 17 can be integrated in a straightforward manner. One has

$$I(\beta; z) = \int_0^1 v_{\beta; z}(x) dx = \sum_{n=0}^{\infty} \frac{z^n}{\beta^n} \int_0^1 d_n(x) dx = \sum_{n=0}^{\infty} \frac{z^n}{\beta^n} T^n \left(\frac{\beta}{2} \right)$$

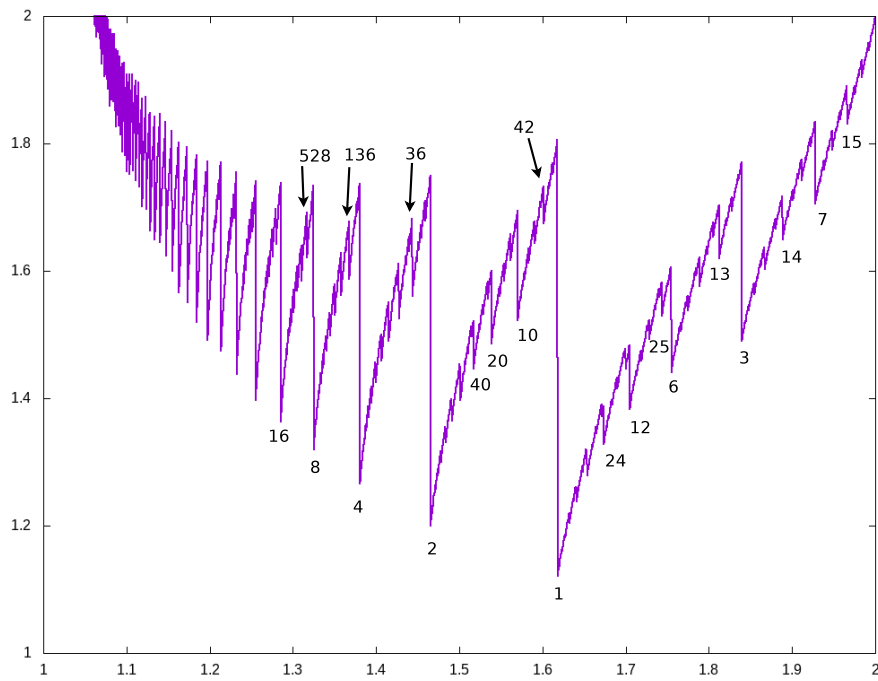
The result is a sawtooth, shown in figure 13. Each discontinuity corresponds to the real root of one of the polynomials. The first few are labeled by the integer labels from the previous table. The doubling sequences and their leaders are easy to identify.

3.5.1 Bracketing intervals

That this figure is a self-similar fractal is presumably self-evident. Thus, for example, the graph to the right of $r_1 = \varphi = 1.618\dots$ repeats again between r_2 and r_1 and again between r_4 and r_2 . Each such bracketed interval contains a unique largest discontinuity; it can be seen as being at the front of a doubling sequence. Each discontinuity is in one-to-one correspondence with a bracketing interval; the bracketing intervals are all self-similar to one-another.

A distinct notation for bracketed intervals is useful. Write $\ell \Rightarrow f \Leftarrow \rho$ for the discontinuity f bracketed on the left and right by ℓ and ρ . By “left” and “right”, it is literally meant that the three roots are in order, with $r_\ell < r_f < r_\rho$ with the inequalities being strict. Not all ascending sequences of three roots form a valid bracket; valid

Figure 13: Normalization Integral



This figure shows the integral $I(\beta) = \sum_{n=0}^{\infty} \beta^{-n} T^n \left(\frac{\beta}{2} \right)$ with $1 < \beta \leq 2$ running along the horizontal axis. Each discontinuity corresponds to the location of a real root of one of the β -Golden polynomials. Some of these are manually labeled by integers, corresponding to the polynomial labels from the previous polynomial table.

brackets are obtained by recursive subdivision; this is given in the next section. But first, some examples.

Taking the liberty to write $r_0 = 2$, the interval to the right of r_1 is then $1 \Rightarrow 3 \Leftarrow 0$. The most prominent self-similar intervals are then $2 \Rightarrow 10 \Leftarrow 1$ and $4 \Rightarrow 36 \Leftarrow 2$ and $8 \Rightarrow 136 \Leftarrow 4$. In each of these examples, the front f was also a leader, with leadership as defined in the previous section. This is not always the case: the brackets $1 \Rightarrow 6 \Leftarrow 3$ and $1 \Rightarrow 12 \Leftarrow 6$ and $1 \Rightarrow 24 \Leftarrow 12$ are clearly visible; they are a part of an index-doubling sequence.

The extreme left side can be assigned the index of ∞ so that $r_\infty = 1$. Thus, the entire interval $1 \leq \beta \leq 2$ corresponds to the bracket $\infty \Rightarrow 1 \Leftarrow 0$.

3.5.2 The bracket tree

The brackets can be arranged into a binary tree, recursively defined. Any valid interval $\ell \Rightarrow f \Leftarrow \rho$ can be split into two: the left side and the right side. These two pieces are $\ell \Rightarrow 2f \Leftarrow f$ on the left, and $f \Rightarrow \Lambda(f) \Leftarrow \rho$ on the right, where Λ is the leader function given by eqn 24 in the previous section. Recursion starts with $\infty \Rightarrow 1 \Leftarrow 0$. A bracketing relationship is valid if and only if it appears in this recursive binary tree.

The left and right moves L, \mathfrak{R} on the binary tree can then be written as

$$\begin{aligned} L : (\ell \Rightarrow f \Leftarrow \rho) &\mapsto (\ell \Rightarrow 2f \Leftarrow f) \\ \mathfrak{R} : (\ell \Rightarrow f \Leftarrow \rho) &\mapsto (f \Rightarrow \Lambda(f) \Leftarrow \rho) \end{aligned} \quad (25)$$

The right-move is denoted with a fraktur \mathfrak{R} instead of a roman R in order to distinguish between the action of the leader function, and the conventional dyadic map

$$\begin{aligned} L : m &\mapsto 2m \\ R : m &\mapsto 2m + 1 \end{aligned}$$

The issue is that not all dyadic R moves result in a valid index: having $\theta(m) = 1$ does not generally imply that $\theta(2m + 1)$ is one. However, the leader function does provide a successor that is always valid: $\theta(\Lambda(m)) = 1$ is guaranteed by construction.

Any sequence of L, \mathfrak{R} moves is guaranteed to produce a valid interval; every location in the binary tree is mapped to a valid index by the bracket recursion moves. A general location on the bracket tree is

$$G = L^{a_1} \mathfrak{R}^{a_2} L^{a_3} \dots \mathfrak{R}^{a_k}$$

and this has the property that if $\theta(m) = 1$ then $\theta(G(m)) = 1$. Equivalently, if $m \in \Psi$ then $G(m) \in \Psi$. This implies that the bracket recursion relations generate all of Ψ . Starting at $m = 1$, all left-right moves produce elements of Ψ ; conversely, every element of Ψ can be expressed as some sequence of left-right moves applied to $m = 1$. It is convenient to take G as the function that produces only the “good” indexes. Thus

$$G : L^{a_1} R^{a_2} L^{a_3} \dots R^{a_k} \mapsto L^{a_1} \mathfrak{R}^{a_2} L^{a_3} \dots \mathfrak{R}^{a_k}$$

Labeling the nodes of the binary tree with the natural numbers \mathbb{N} provides a bijection $G : \mathbb{N} \rightarrow \Psi$.

The bracket recursion relations split the interval $r_\ell < r_f < r_\rho$ into left and right pieces, as well. That is, each node of the binary tree is labeled by some (unique) r_f , with the root r_ℓ and r_ρ appearing as predecessors in the tree. Of course, the binary tree can also be mapped to the dyadic rationals; thus the bracketing recursion relations give a bijection between the dyadic rationals and the finite-length orbits of the β -map. The bracket tree and the dyadic tree are in one-to-one correspondence. The correspondence is shown in figure 14.

Perhaps the most interesting aspect to the figure is that it appears to be continuous. That is, the roots r_n appear to be dense in the interval $1 \leq \beta \leq 2$. Although the midpoints r_f in an interval $r_\ell < r_f < r_\rho$ are not evenly spaced, the midpoint always appears to be sufficiently far away from either endpoint so that the convergents are dense in the reals, which implies in turn that the map is continuous. It would appear that the map to infinite-length sequences can be taken, without pathologies.

3.5.3 The finite comb

The figure 12 shows each indicator rank Ψ_ν as a comb. A better understanding is gained by mapping this to the infinite binary tree. Each rank ν corresponds to a single horizontal row in the tree. The numbering that has been adopted is that $\nu = 2$ is at the root of the tree. The dyadic left and right moves on the comb are

$$\begin{aligned} L : \theta_m &\mapsto \theta_{2m} \\ R : \theta_m &\mapsto \theta_{2m+1} \end{aligned}$$

The indicator function marks each node in the tree with a zero or a one.

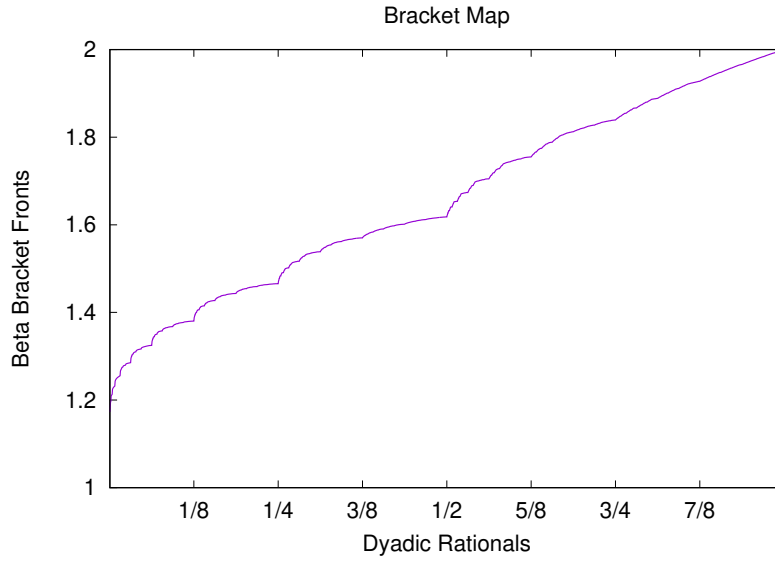
One property of this marking is that when a node is marked with a zero, all nodes in the right subtree are marked with a zero, as well: the presence of a zero removes the entire right-hand branch under that point. This follows from the property noted earlier, that if $\theta(m) = 0$, then $\theta(2^n(2m+1)) = 0$ for all n . Restating in terms of moves, if $\theta_m = 0$ then, taking $n = 0$, one has $R\theta_m = \theta_{2m+1} = 0$. Of course it, follows that $RR\theta_m = 0$. Taking $n = 1$, one has that $LR\theta_m = 0$, and so both left and right sub-branches are gone.

A converse marking is given by the property that, if $\theta(m) = 1$ then $\theta(2m) = 1$. On the tree, this means that left-branches of a node marked with a one are never trimmed: if $\theta_m = 1$, then $L\theta_m = 1$.

Write $2^{<\omega}$ for the collection of all finite strings in the letters L, R and 2^ω as the set of all infinite strings. The indicator function is then a map $\theta : 2^{<\omega} \rightarrow \{0, 1\}$ that indicates when a given finite, unbounded walk down the tree might have more branches, or definitely does not. This definitely-maybe marking defines a filter, and dually, an ideal. The filter/ideal can be taken on the integers, or equivalently, on the reals. The second possibility is what the figure 12 is hinting at.

The filter has the form that if $n \in \Psi_\nu$ then $2n \in \Psi_{\nu+1}$. The converse is not true: $10 \in \Psi_5$ but $5 \notin \Psi_4$. Written as a filter, this is $L\Psi_\nu \subset \Psi_{\nu+1}$ and the subset relation is strict. Written as an ideal, if $n \notin \Psi_\nu$ then $(2n+1) \notin \Psi_{\nu+1}$. This is more clearly stated with set complements at a given rank. Define the unit interval $I_\nu = \{n : 2^{\nu-2} \leq n < 2^{\nu-1}\}$ and the set complement $\bar{\Psi}_\nu = I_\nu \setminus \Psi_\nu$. The ideal is then $R\bar{\Psi}_\nu \subset \bar{\Psi}_{\nu+1}$ together with $(LR\bar{\Psi}_\nu \cup RR\bar{\Psi}_\nu) \subset \bar{\Psi}_{\nu+2}$.

Figure 14: Beta Bracket Map



This figure shows the mapping from the dyadic rationals, interpreted as strings of left-right moves, and the corresponding $\beta = r_f$ root for the front (center) of the corresponding bracket. Thus, $1/2$ maps to $r_1 = \varphi = 1.61803\dots$ and $1/4$ maps to $r_2 = 1.46557\dots$ while $3/4$ maps to $r_3 = 1.83929\dots$ It would appear that the curve is continuous. The curve does drop to $\beta = 1$ at the left, but it does so very slowly. The brackets $\ell \Rightarrow f \Leftarrow \rho$ remain quite wide, as the limit approaches $\beta = 1$. It does get there, though. An explicit derivation of this limit is given in a later section. The sharp takeoff at the left is reminiscent of the Minkowski question-mark mapping. The left generator for the bracket map is quite close to left generator $x/(1+x)$ of the Minkowski fractal. The Minkowski fractal has $1/(2-x)$ as the right generator. Replacing this with $x + 1/2$ generates a curve resembling the bracket curve above, having Minkowski-like behavior to the left, an linear-like behavior to the right. The de Rham curve construction may be used to generate such curves, given arbitrary L,R maps. The L,R maps for the bracket curve are both fractal themselves, so finding affine generators would be a surprise.

3.5.4 The comb bijection

The good-index bijection $G : \mathbb{N} \rightarrow \Psi$ provides a mechanism to map the full binary tree into the trimmed tree, so that every node in the trimmed tree is in one-to-one correspondence with the full binary tree.

The leadership function provides the desired mapping. If $n \in \Psi_v$, then $\Lambda(n) \in \Psi_{v+1+h}$ where h was the height of the leader. Any valid index n gets “kicked upstairs” by the leadership function; one can write $\Lambda\Psi_v \subset \bigcup_{h=0}^{\infty} \Psi_{v+1+h}$. This just corresponds to the bracket move $\mathfrak{R} : \theta_m \rightarrow \theta_{\Lambda(m)}$ on the trimmed tree: given any location $\theta(m) = 1$ in the trimmed tree, the bracket right move returns the next valid right branch in the trimmed tree: $\mathfrak{R}(m) = \Lambda(m)$, since, by construction, $\theta(\Lambda(m)) = 1$ whenever $\theta(m) = 1$.

3.5.5 The infinite comb

The comb is mapped to the reals by defining open subsets bounded by the dyadic rationals. Let

$$I(m, v) = \{x \in \mathbb{R} : m < x2^{v-2} < (m+1); 0 \leq m < 2^{v-2}\} \quad (26)$$

so that $I(0, 2) = \{x \in \mathbb{R} : 0 < x < 1\}$. The left and right moves are the obvious ones: $LI(m, v) = I(2m, v+1)$ and $RI(m, v) = I(2m+1, v+1)$. These are interpreted as the left and right halves of the (fat) Cantor set, on the reals; the fat Cantor set being taken as the reals with the dyadic rationals removed.

The filters on Ψ_v become filters on $I(m, v)$ in the obvious way, by means of a commuting diagram. Write $I(\Psi_v) = \bigcup_{m \in \Psi_v} I(m - 2^{v-2}, v)$ as the union of intervals that cover the nonempty runs in the binary tree. This is effectively what is being graphed in the figure 12. The limit set is then

$$\bar{\theta} = \bigcap_{v=2}^{\infty} I(\Psi_v) \quad (27)$$

where the notation $\bar{\theta}$ is happily abused to denote the limit of the indicator function θ_n as a subset of the reals. This is possible precisely because θ_n can be mapped to the binary tree, which can then be partitioned as filters and ideals.

Since $2^{v-2} \in \Psi_v$ for all v , i.e. the leftmost branch is never trimmed, one easily concludes that $0 \in \bar{\theta}$. Since $(2^{v-1} - 1) \in \Psi_v$ for all v , the rightmost branch is never trimmed, and so one may conclude that $1 \in \bar{\theta}$.

It should be clear that the set $\bar{\theta}$ is isomorphic to the Cantor set. The comb bijection shows how to map points in $\bar{\theta}$ back into the untrimmed infinite tree; the full binary tree is isomorphic to the Cantor set. From this, one concludes that the set $\bar{\theta}$ is uncountable, and can be placed in bijection with the reals. This is explored in the next section. A measure can be assigned to $\bar{\theta}$. This is given in the section after next. A more formal examination of $\bar{\theta}$, tightening up some of the loose language above, will be given in a later section on eventually-periodic orbits.

3.5.6 Self-describing orbits

What is the meaning of the finite and infinite combs? The finite comb is a mapping of the valid-index set Ψ to the dyadics. The infinite comb is the closure of the finite comb in the reals.

A polynomial index $n \in \Psi$ if and only if n encodes a self-describing finite orbit. That is, $n \in \Psi$ if and only if the real root r_n of $p_n(r_n) = 0$ iterates under the β -map such that the iterate reproduces the bit-sequence of n . That is, $n \in \Psi$ if and only if the bits b_i of the bitstring $2n+1 = b_0b_1b_2 \cdots b_v$ are given by the (finite) orbit as

$$b_i = \Theta \left(T_\beta^i \left(\frac{r_n}{2} \right) - \frac{1}{2} \right) = k_i \left(\frac{r_n}{2} \right)$$

If n does not have this property, then $n \notin \Psi$. At each rank v , the elements of Ψ_v correspond to finite orbits of length v . In the limit of $v \rightarrow \infty$, one gets self-describing orbits of unbounded (infinite) length.

By construction, if $n \in \Psi_v$ then $Ln \in \Psi_{v+1}$ and $\Lambda n \in \Psi_{v+1+h}$. That is, if n is a self-describing orbit, then it appears as the prefix of longer self-describing orbits.

Polynomials of infinite order are holomorphic functions. Given an infinite bit-sequence $\{b\} = b_0b_1 \cdots$, define a holomorphic function

$$q^{\{b\}}(\zeta) = 1 - \sum_{j=0}^{\infty} b_j \zeta^{j+1} \quad (28)$$

If the bit-sequence is finite, in that all $b_j = 0$ when $j > k$, then this is related to the polynomials as

$$\zeta^{k+1} p_n \left(\frac{1}{\zeta} \right) = 1 - b_0 \zeta - b_1 \zeta^2 - \cdots - b_k \zeta^{k+1}$$

Given any arbitrary sequence $\{b\}$, the holomorphic function $q^{\{b\}}(\zeta)$ will have a single, unique real, positive root. To make contact with the polynomials, write this as the reciprocal, so that the root r satisfies $q^{\{b\}}(1/r) = 0$. This root will have some orbit, given by

$$a_i = \Theta \left(T_\beta^i \left(\frac{r}{2} \right) - \frac{1}{2} \right) = k_i \left(\frac{r}{2} \right)$$

Such an orbit is self-describing if and only if $\{a\} = \{b\}$.

The claim being pursued here is that the infinite comb contains all self-describing sequences, and conversely, every element of the comb corresponds to a self-describing bit-sequence. That is, if

$$x = \sum_{j=0}^{\infty} \frac{b_j}{2^j} \in \bar{\theta}$$

then

$$x = 2 \left(1 - q^{\{b\}} \left(\frac{1}{2} \right) \right)$$

Every real number $x \in \overline{\theta}$ in the comb corresponds to such a self-describing orbit. A proof of these claims will be given in a later section, after the development of some formal definitions.

In the meanwhile, it can be noted that every rational number corresponds to a bitsequence that is ultimately periodic. After an initial unstable finite sequence, the bitstring settles down to a cyclic orbit. One task ahead is to examine the set $\mathbb{Q} \cap \overline{\theta}$: this is the set of self-describing eventually-periodic infinite-length orbits. It turns out these can be readily described as root of a finite polynomial. The holomorphic function $q^{\{b\}}(\zeta)$ to be factored into two finite polynomials, one describing the initial aperiodic segment, and a second describing the cyclic segment. Such orbits are examined in a later section.

3.6 Formal definitions

A sufficient number of distinct concepts have been introduced, that some basic house-keeping is in order. The definitions that follow are straight-forward and conventional. The goal is to provide a workable vocabulary for further discussion.

Let \mathbb{B} denote the finite but unbounded binary tree, and $\overline{\mathbb{B}}$ it's closure as the infinite tree. The infinite tree is, of course, isomorphic to the Cantor space 2^ω ; but this mechanism is not currently needed. A more careful definition of the finite tree is needed. Thus, let \mathbb{B} be the graph of vertices and connecting edges $\mathbb{B} = \{v_i, e_{ij} : i \in \mathbb{N}, j \in \{2i, 2i+1\}\}$. Let $\eta : \mathbb{N} \rightarrow \mathbb{B}$ denote the canonical labeling of the binary tree by the positive integers, so that the root of the tree is given the label 1, the first row is 2,3 and the next row is 4,5,6,7. This is a bijection: every finite walk down the tree can be labeled with a positive integer. The walks are left and right moves, in the canonical sense: $L : \mathbb{N} \rightarrow \mathbb{N}$ with $L : m \mapsto 2m$ and likewise $R : \mathbb{N} \rightarrow \mathbb{N}$ with $R : m \mapsto 2m+1$. The pushforward of L, R provide the canonical walks on the tree, as a commuting diagram, so that $L \circ \eta = \eta \circ L$ and likewise $R \circ \eta = \eta \circ R$. It is useful to adjoin the pre-root elements $\{0, \infty\}$ so that $R : 0 \rightarrow 1$ and $L : \infty \rightarrow 1$.

Let \mathbb{D} be the dyadic rationals, with the canonical bijection to the natural numbers $\delta : \mathbb{D} \rightarrow \mathbb{N}$ given by $\delta : (2n+1)/2^m \mapsto 2^{m-1} + n$. This labels the root of the tree with 1/2, and the first row under it as 1/4 and 3/4. The left and right moves are $L : (2n+1)/2^m \mapsto (4n+1)/2^{m+1}$ and $R : (2n+1)/2^m \mapsto (4n+3)/2^{m+1}$. This was set up so that $L \circ \delta = \delta \circ L$ and likewise $R \circ \delta = \delta \circ R$. The map $\delta^{-1} : \mathbb{N} \rightarrow \mathbb{D}$ is equally familiar: it is $n = b_0 b_1 \dots b_k \mapsto \sum_{n=0}^k b_n 2^{-n-1}$ that interprets $n \in \mathbb{N}$ as a sequence of L, R moves in the binary tree, returning the fraction found at that location.

This allows the bracket tree and the good-index bijection to be specified more precisely. The validity set $\Psi \subset \mathbb{N}$ is defined in eqn 23 as the collection of natural number indexes corresponding to polynomials with self-describing roots. This set came with two functions $L, \mathfrak{R} : \Psi \rightarrow \Psi$ given by index-doubling $L : m \mapsto 2m$ and leadership (eqn24) $\mathfrak{R} : m \mapsto \Lambda(m)$. The good-index bijection $G : \mathbb{N} \rightarrow \Psi$ is then defined as the pullback $L \circ G = G \circ L$ and $\mathfrak{R} \circ G = G \circ R$.

The trimmed tree $\mathfrak{B} \subset \mathbb{B}$ is the image of Ψ under the mapping η , so that $\mathfrak{B} = \eta\Psi$. This consists of those nodes and edges in the finite binary tree that are labeled by integers from the validity set Ψ . The good-index function G places the trimmed tree and the finite tree in a bijection, so that $\mathfrak{B} = \eta G \eta^{-1} \mathbb{B}$.

The root map $r : \Psi \rightarrow [1, 2]$ takes valid integer indexes, and maps them to the roots of the corresponding (finite-orbit) polynomials, so that $p_n(r_n) = 0$. The bracket map, depicted in figure 14, can now be written more precisely as the function $\rho = r \circ G \circ \delta^{-1} : \mathbb{D} \rightarrow [1, 2]$. It maps the dyadics to the β values that have finite orbits.

The notation for the finite comb θ is abused in several ways. In eqn 22 it is used to indicate whether a given polynomial root has a self-describing orbit. It is then defined as the indicator function for set membership $\theta = \mathbb{1}_\Psi : \Psi \rightarrow \{T, F\}$ and finally as the finite comb $\theta \subset \mathbb{D}$. Formally, the finite comb can now be written as $\theta = \delta^{-1}\Psi$.

The good-index bijection $G : \mathbb{N} \rightarrow \Psi$ can be commuted with the dyadic bijection to obtain the “good dyadic map” $G = \delta \circ G \circ \delta^{-1} : \mathbb{D} \rightarrow \theta$. Given any dyadic fraction, this map returns another dyadic that lies within the finite comb. As all the other maps discussed so far, it is a bijection. It is depicted in figure 15.

The infinite comb $\bar{\theta}$ was the closure of the finite comb, so that $\theta \subset \bar{\theta}$. The closure, defined in eqn 27, is constructed as the infinite intersection of open sets. Corresponding to this is the closure $\bar{\mathfrak{B}} \subset \mathbb{B}$ of infinitely long paths in the trimmed tree. It is taken as the limit of the finite but unbounded-length paths in \mathfrak{B} . The dyadics can be closed in several ways; one way is to the rationals \mathbb{Q} , and then further to the reals \mathbb{R} . In the present case, we restrict attention to the unit interval $I = [0, 1] \subset \mathbb{R}$ and to $\mathbb{Q}_I = \mathbb{Q} \cap [0, 1]$. Proper diligence requires distinguishing $\bar{\mathfrak{B}}_{\mathbb{Q}}$ from $\bar{\mathfrak{B}}_{\mathbb{R}}$ and also $\theta_{\mathbb{Q}}$ from $\theta_{\mathbb{R}}$.

The rationals correspond to infinite length orbits that are ultimately periodic; these will be examined in a later section. A closure to the root map to the rationals appears naturally; it can be written as $\bar{r} : \bar{\theta} \rightarrow [1, 2]$. Combined with the closure $G : \mathbb{Q}_I \rightarrow \theta$ this gives a closure of the bracket map $\bar{\rho} = \bar{r} \circ G : \mathbb{Q}_I \rightarrow [1, 2]$. Closing to the reals gives $G : [0, 1] \rightarrow \theta$ and $\bar{\rho} = \bar{r} \circ G : [0, 1] \rightarrow [1, 2]$. This is the primary achievement of this section: the definition of the bracket map as a continuous monotonic ascending bijection between the reals in the unit interval, and β values understood as self-describing orbits. Proving continuity will require some theoretical machinery and lemmas; these will be developed in a later section. However, the net result can already be seen in figure 14.

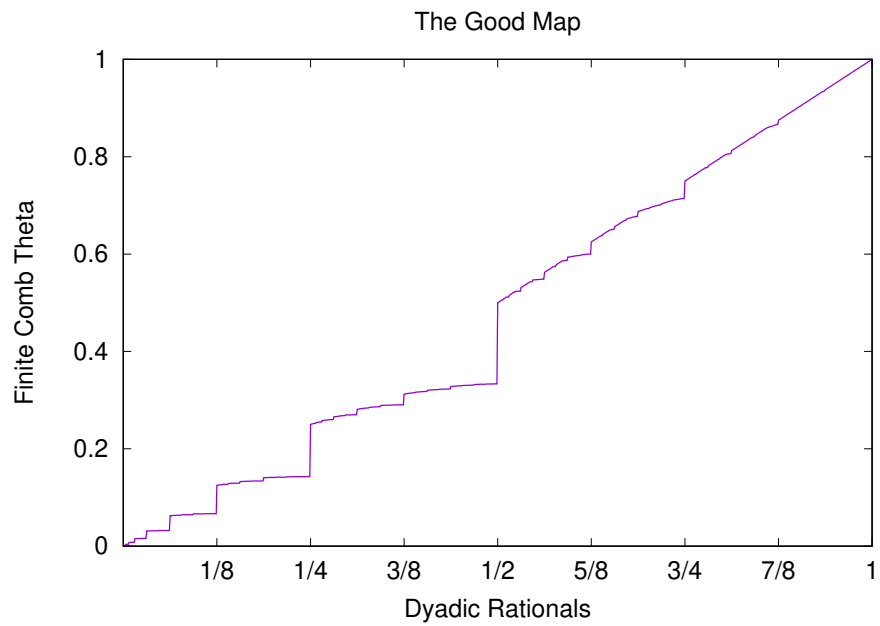
3.7 The Good Map and Measures

The formal definitions allow the description of the bracket map to be completed. This is a matter of reviewing the remaining congruences. The fundamental one, moving forward, is the “good map” $G : \mathbb{D} \rightarrow \theta$ that is a correspondence between the dyadic rationals to the comb. The comb, in turn, can be understood as inducing a measure. The measure, combined with the good map, induces the “beta measure”, which is just the bracket map. This closes the circle of commuting diagrams.

3.7.1 The good map

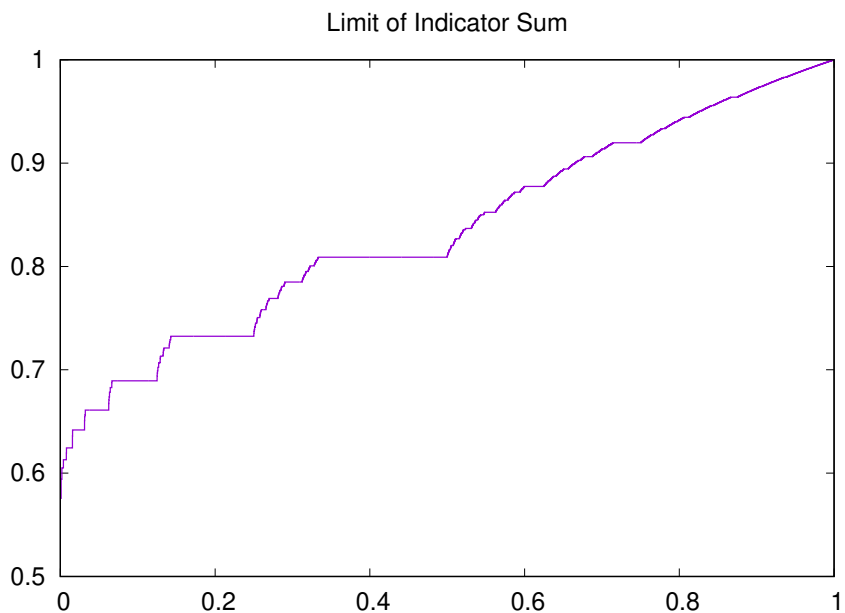
The validity map $G : \mathbb{N} \rightarrow \Psi$ can be commuted with the canonical mapping $\delta : \mathbb{N} \rightarrow \mathbb{D}$ between the natural numbers and the dyadic rationals. This defines a map $G = \delta \circ G \circ \delta^{-1}$ that is a bijection between the dyadic rationals and the finite comb: $G : \mathbb{D} \rightarrow \theta$. It is shown in figure 15.

Figure 15: The Good Map



A graph of the “good map” $G : \mathbb{D} \rightarrow \theta$. It is defined as the commutator $G = \delta \circ G \circ \delta^{-1}$ of the valid-index map $G : \mathbb{N} \rightarrow \Psi$ by the canonical mapping $\delta : \mathbb{N} \rightarrow \mathbb{D}$ between the natural numbers and the dyadic rationals. The finite comb is just a mapping of the good indexes Ψ to the dyadics: $\theta = \delta^{-1}\Psi$.

Figure 16: Limit of Indicator Sum



This figure shows the limit of the indicator sum $A(x) = \lim_{v \rightarrow \infty} A_v(x)$ of eqn 29. More precisely, it shows $A_v(x)$ for $v = 20$. By this point, convergence is sufficient that any differences from the limit are not visible to the naked eye.

3.7.2 The comb measure

The infinite comb was constructed as a closure or limit of the finite comb. An interesting trick is to interpret it as a measure, and so to integrate over it. This can be obtained as a limit over sums of ranks in the finite comb. The measure is depicted in figure 16.

The sum over the indicator function $S(k) = \sum_{n=1}^k \theta_n$ shows power-of-two periodicity, same as each rank in the finite comb. Each rank v can be separated out as

$$S_v(k) = S(k) - S(2^{v-2} - 1)$$

The intent is to isolate the range $2^{v-2} \leq k < 2^{v-1}$, and to split the sum $S(k)$ into a collection of ranks S_v .

At the end of the range, the sum S_v achieves Moreau's necklace-counting function: $S_v(2^{v-1} - 1) = M_v$. Dividing by M_v gives each rank the same vertical scale: zero to one. It is also useful to rescale the horizontal range, to run zero-to-one as well. This gives a normalized version

$$F_v(x) = \frac{S_v(\lfloor 2^{v-2}(1+x) \rfloor)}{M_v}$$

that runs from zero to one as x runs from zero to one. This function does not have an interesting limit as $v \rightarrow \infty$, as it slowly drops to zero over the entire unit interval. However, it does so at a fixed rate, and can be held constant with a radical. The sequence of functions

$$A_v(x) = \exp(M_v 2^{-v} \log F_v(x)) \quad (29)$$

converge rapidly and more-or-less uniformly to a limit $A(x) = \lim_{v \rightarrow \infty} A_v(x)$. It appears to be well-defined on the unit real interval $0 \leq x \leq 1$. This is the limit shown in figure 16. It is perhaps useful to keep in mind the asymptotic limit of Moreau's function, $M_v = 2^v/v - \mathcal{O}(2^{v/2}/v)$ and so the radical scales as $M_v 2^{-v} = 1/v - \mathcal{O}(2^{-v/2}/v)$.

The convergence to the limit appears to be uniform and rapid, except at $x = 0$, which proceeds slowly. This is easily demonstrated. The $x = 0$ limit is

$$A(0) = \lim_{v \rightarrow \infty} (F_v(0))^{M_v 2^{-v}} = \lim_{v \rightarrow \infty} \left(\frac{1}{M_v} \right)^{M_v 2^{-v}} = \lim_{v \rightarrow \infty} \left(\frac{v}{2^v} \right)^{1/v} = \frac{1}{2} \lim_{v \rightarrow \infty} v^{1/v} = \frac{1}{2}$$

The slow convergence is entirely due to the last limit, above.

The offsets to the sums and limit above were defined above, so as to avoid having to debate the meaning of $\lim_{v \rightarrow \infty} \sqrt[v]{0}$. Yet clearly, the intent is that $A(x)$ should provide a measure for the infinite comb. Should the infinite comb be thought of as having a large point-weight at $x = 0$? Perhaps not; thus, perhaps a more suitable measure is $\mu_\theta(x) = 2A(x) - 1$, which runs from zero to one over the unit interval.

To summarize: The function $\mu_\theta(x)$ provides a measure for the infinite comb given in eqn 27. The non-flat sections are where the infinite-length self-describing sequences are accumulating.

3.7.3 The Beta measure

The significance of the comb measure is revealed by superimposing its graph 16 on the bracket map 14, so that both appear side by side. This is shown in 17. The stair-treads line up with the blancmange dips in the bracket map. The good map can be used to eliminate the stair treads. The identity is

$$2A \circ G = \rho = r \circ G \circ \delta^{-1} = r \circ \delta^{-1} \circ G$$

Since $G = \delta \circ G \circ \delta^{-1}$ is a bijection, it can be peeled off, to give

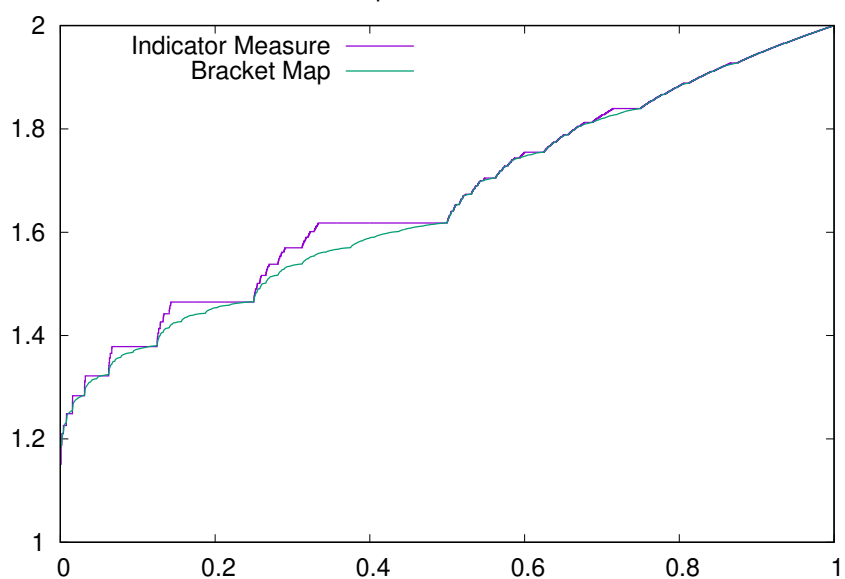
$$2A = r \circ \delta^{-1}$$

This is surprising. The map r is purely local: it takes integers n to the enumerated roots $p_n(r_n) = 0$. It is just specifying locations of roots. The map A is global: it is counting how many roots there are. It is the limit of a sum, a kind of peculiar integral, that captures information about all other β values, and how they behaved.

An alternative interpretation is that this provides a way of estimating the number of orbits of length v , satisfying $\beta < \alpha$ for some fixed $1 \leq \alpha \leq 2$. The total number of orbits of length v is given by Moreau's M_v . The counting function $S_v(k)$ returns the total number of orbits of length v that occurred at some β with $\beta \leq r_k$. The normalized version provides this same number slightly more elegantly: the total number of orbits

Figure 17: The Beta Measure

Bracket Map and Indicator Measure



This figure superimposes the two figures 14 and 16 into one. The indicator measure has been rescaled, so that the y-axis aligns with the interval $1 \leq \beta \leq 2$, as that is the nominal topic of discussion. The comb is also a map through β values, but taken sideways, as it were. This figure indicates visually what that correspondence is. When the stair-treads are removed with the “good map”, the two curves are identical. This is surprising, as they have entirely different origins: the indicator measure is counting periodic orbits, while the bracket map is providing the locations.

of length v with $\beta < \rho(x)$ is given by $M_v F_v(x)$. The indirection with ρ is annoying; define $f_v(\alpha) = F_v(\rho^{-1}(\alpha))$, so that $f_v(\alpha)$ counts the fraction of all orbits of length v occurring at some (any) $\beta \leq \alpha$. This fraction is approximated as $f_v(\alpha) \approx (\alpha/2)^v$, which holds exactly in the limit $v \rightarrow \infty$.

This last again illustrates the local-global tie between these two: $f_v(\alpha)$ is a counting function, an integral of sorts, while α is just a number. It's quite rare to find such specific analytic results in this project. Very rare: this is the first.

3.8 Complex Roots

What are the complex roots? Numerical work clearly indicates that they seem to be approximately cyclotomic in some sense or another. They seem to be more-or-less uniformly distributed in an approximate circle, always. The modulus of most of the complex roots appear to be less than one. This is violated for the complex roots of $p_{2^k}(\beta) = \beta^{k+2} - \beta^{k+1} - 1$, where some of the roots in the right-hand quadrant have a modulus larger than one. By contrast, the complex roots of $p_{2^k-1}(\beta) = \beta^{k+1} - \sum_{j=0}^k \beta^j$ seem to always have a modulus less than one. These two seem to be the extreme cases: in general, the polynomials appear to be "approximately cyclotomic". Its not clear how to make this statement more precise.

These numerical results can be argued heuristically: just divide the polynomial by its leading order. That is, a general polynomial of this form is

$$p_n(z) = z^{k+1} - \sum_{j=0}^k b_j z^{k-j}$$

with the convention that $b_0 = b_k = 1$, and the bit-sequence $2n+1 = b_0 b_1 b_2 \dots b_p$ corresponding to a terminating orbit. Dividing by z^{k+1} gives a series

$$1 - \frac{1}{z} - \frac{b_1}{z^2} - \frac{b_2}{z^3} - \dots$$

Clearly, this can have a zero only when $|z| < 2$ as otherwise, the terms get too small too quickly.

3.9 β -Golden β -Fibonacci Sequences

It is well known that the golden ratio occurs as limit of the ratio of adjacent Fibonacci numbers:

$$\varphi = \lim_{m \rightarrow \infty} \frac{F_m}{F_{m-1}}$$

where $F_m = F_{m-1} + F_{m-2}$. There is a generalization of this, which also has received attention: the tribonacci, quadronacci, *etc.* sequences, whose limits are

$$\alpha_n = \lim_{m \rightarrow \infty} \frac{F_m^{(n)}}{F_{m-1}^{(n)}}$$

where $F_m^{(n)} = F_{m-1}^{(n)} + F_{m-2}^{(n)} + \dots + F_{m-n}^{(n)}$.

Is it possible that the real roots of the polynomials $p_n(\beta)$ are also the roots of such sequences? But of course they are! Given a finite string of binary digits $\{b\} = b_0, b_1, \dots, b_k$, write the beta-Fibonacci sequence as

$$F_m^{\{b\}} = b_0 F_{m-1}^{\{b\}} + b_1 F_{m-2}^{\{b\}} + \dots + b_k F_{m-k}^{\{b\}}$$

The name “beta-Fibonacci” is needed because the term “generalized Fibonacci sequence” is already in wide circulation for the special case of all bits being one. The ratio of successive terms is

$$\alpha^{\{b\}} = \lim_{m \rightarrow \infty} \frac{F_m^{\{b\}}}{F_{m-1}^{\{b\}}}$$

and is given as the (positive) real root of the polynomial

$$p_n(\beta) = \beta^{k+1} - b_0 \beta^k - b_1 \beta^{k-1} - \dots - b_k = 0$$

These polynomials and their roots were already enumerated and tabulated in the previous section.

The beta-Fibonacci sequences do not appear by accident: these sequences have an ordinary generating function (OGF) given by the polynomial! That is,

$$\sum_{m=0}^{\infty} z^m F_m^{\{b\}} = \frac{z^k}{1 - b_0 z - b_1 z^2 - \dots - b_k z^{k+1}} = \frac{1}{z p_n\left(\frac{1}{z}\right)}$$

The factor of z^k in the numerator serves only to initiate the sequence so that $F_0^{\{b\}} = \dots = F_{k-1}^{\{b\}} = 0$ and $F_k^{\{b\}} = 1$.

These sequences are generic: they indicate how many different ways one can partition the integer m into elements of the set $\{b_0, 2b_1, 3b_2, \dots, (k+1)b_k\}$. So, for example, the entry for $n = 12$ in the table below corresponds to [OEIS A079971](#), which describes the number of ways an integer m can be partitioned into 1, 2 and 5 (or that $5m$ can be partitioned into nickels, dimes and quarters). This corresponds to the bit sequence $\{b\} = 11001$; that is, $\{b_0, 2b_1, 3b_2, \dots, (k+1)b_k\} = \{1 \cdot 1, 2 \cdot 1, 3 \cdot 0, 4 \cdot 0, 5 \cdot 1\} = \{1, 2, 5\}$. From such partitions, it appears that one can build partitions of the positive integers that are disjoint, and whose union is the positive integers. This suggests a question: can these partitions be expressed as Beatty sequences?

The previous table is partly repeated below, this time annotated with the OEIS sequence references.

n	binary	root	root identity	sequence
0	1	1		
1	11	$\varphi = \frac{1+\sqrt{5}}{2} = 1.618$	golden ratio	Fibonacci A000045
2	101	1.465571231876...	OEIS A092526	Narayana A000930
3	111	1.839286755214...	tribonacci A058265	tribonacci A000073
4	1001	1.380277569097...	2nd Pisot A086106	A003269 , A017898
6	1101	1.754877666246...	A109134	A060945
7	1111	1.927561975482...	tetranacci A086088	tetranacci A000078
8	10001	1.324717957244...	silver A060006	A003520 , A017899
10	10101	1.570147312196...	Pisot A293506	A060961
12	11001	1.704902776041...		A079971
13	11011	1.812403619268...		A079976
14	11101	1.888518845484...		A079975
15	11111	1.965948236645...	pentanacci A103814	A001591
16	100001	1.285199033245...		A005708 , A017900

All of these integer sequences and roots participate in a number of curious relations having a regular form; this is, of course, the whole point of listing them in the OEIS. This suggests a question: do the known relationships generalize to the beta-shift setting?

For example, the polynomials of the form

$$\beta^{k+1} - \beta - 1 = 0$$

are the Lamé polynomials, they arise as solutions to Lamé's equation, a kind of ellipsoidal harmonic differential equation. In the present notation, these correspond to polynomials $p_n(\beta) = 0$ for $n = 2^k$.

Another example is the Fibonacci-tribonacci-tetranacci sequence of "generalized golden means". These are the roots of the series for which all $b_k = 1$, that is, the roots of

$$\beta^{k+1} - \beta^k - \beta^{k-1} - \dots - 1 = 0$$

In the present notation, these would be the polynomials $p_n(\beta) = 0$ for $n = 2^k - 1$. Such roots can be rapidly computed by a series provided by Hare, Prodinger and Shallit[29]:

$$\frac{1}{\alpha_k} = \frac{1}{2} + \frac{1}{2} \sum_{j=1}^{\infty} \frac{1}{j} \binom{j(k+1)}{j-1} \frac{1}{2^{j(k+1)}}$$

This series is obtained by making good use of the Lagrange inversion formula. Here, α_k is the k 'th generalized golden mean, i.e. the solution $p_{2^k-1}(\alpha_k) = 0$. Can the Hare series be extended to provide the roots r_n of $p_n(r_n) = 0$ for general n ?

Another set of observations seem to invoke the theory of complex multiplication on elliptic curves, and pose additional questions. So:

The tribonacci root r_3 is given by

$$r_3 = \frac{1}{3} \left(1 + \sqrt[3]{19 + 3\sqrt{33}} + \sqrt[3]{19 - 3\sqrt{33}} \right) \simeq 1.839\dots$$

The silver number (plastic number) r_8 is given by

$$r_8 = \frac{1}{6} \left(\sqrt[3]{108 + 12\sqrt{69}} + \sqrt[3]{108 - 12\sqrt{69}} \right) \simeq 1.324\dots$$

The Narayana's cows number (supergolden ratio) r_2 is given by

$$r_2 = \frac{1}{6} \sqrt[3]{116 + 12\sqrt{93}} + \frac{2}{3 \sqrt[3]{116 + 12\sqrt{93}}} + \frac{1}{3} \simeq 1.465\dots$$

The root r_6 is related to the silver number r_8 as $r_8 = r_6(r_6 - 1)$ and is given by

$$r_6 = \frac{1}{6} \sqrt[3]{108 + 12\sqrt{69}} + \frac{2}{\left(\sqrt[3]{108 + 12\sqrt{69}} \right)^2} \simeq 1.754\dots$$

Do the other roots have comparable expressions? To obtain them, is it sufficient to articulate the theory of “complex multiplication” on elliptic curves? Are these just disguised solutions to Lamé’s ellipsoidal harmonic differential equation? The appearance of only the cube and square roots is certainly suggestive of an underlying process of points on elliptic curves.

3.10 β -Fibonacci sequences as shifts

The nature of the β -Fibonacci sequences as shift sequences can be emphasized by noting that they arise from the iteration of the **companion matrix** for the polynomial $p_n(x)$. This is a $(k+1) \times (k+1)$ matrix in **lower-Hessenberg form**:

$$B = \begin{bmatrix} b_0 & 1 & 0 & 0 & \dots & 0 \\ b_1 & 0 & 1 & 0 & \dots & 0 \\ b_2 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ b_{k-1} & 0 & 0 & 0 & \dots & 1 \\ b_k & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (30)$$

Iteration produces a **linear recursive sequence** this is the β -Fibonacci sequence. The m 'th element of the sequence is obtained from the m 'th iterate B^m .

Define the **exchange matrix** as

$$J = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}$$

This can be used to write the above in the more conventional companion-matrix form:

$$C = [JB]J^T = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ b_k & b_{k-1} & b_{k-2} & b_{k-3} & \cdots & b_0 \end{bmatrix}$$

Some explicit examples are in order. For the golden ratio, one has

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

and the iterates are

$$B^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, B^3 = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}, B^4 = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, B^n = \begin{bmatrix} F_n & F_{n-1} \\ F_{n-1} & F_{n-2} \end{bmatrix}$$

with F_n being the n 'th Fibonacci number, as usual. For the general case, after $m \geq k - 1$ iterations, one gets the **Hankel matrix**

$$B^m = \begin{bmatrix} F_m^{(b)} & F_{m-1}^{(b)} & F_{m-2}^{(b)} & \cdots & F_{m-k+1}^{(b)} & F_{m-k}^{(b)} \\ F_{m-1}^{(b)} & F_{m-2}^{(b)} & F_{m-3}^{(b)} & \cdots & F_{m-k}^{(b)} & F_{m-k-1}^{(b)} \\ F_{m-2}^{(b)} & F_{m-3}^{(b)} & & \cdots & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ F_{m-k+1}^{(b)} & & & \cdots & & \\ F_{m-k}^{(b)} & & & \cdots & & F_{m-2k}^{(b)} \end{bmatrix}$$

so that the top row consists of the latest sequence values. When multiplied by the bits, this just generates the next iterate in the sequence. The upper-diagonal 1's just serve to shift columns over by one, with each iteration: that is why it's a shift!

The product $B^m J$ is a **Toeplitz matrix**.

The characteristic polynomial of this matrix is, of course, the polynomial p_n :

$$\det[B - xI] = (-1)^k p_n(x)$$

Thus, we can trivially conclude that the eigenvalues of B are given by the roots of $p_n(x)$. This matrix is in lower-Hessenberg form; this makes it obvious that it's a shift; a finite shift, in this case.

For every Hankel matrix, there are corresponding **moment problems** that ask if there is a measure that generates the corresponding matrix entries (as moments). The moment problem on the unit interval is the **Hausdorff moment problem**. In the present case, it appears that the corresponding measure is precisely the Parry–Gelfond measure. The only point of confusion is that the measure is placed in the denominator, not the numerator, as the measure is for the pushforward, rather than being a pullback. There do appear to be various interrelationships between the moments, these remain to be articulated.

3.11 Infinite sequences and shifts

Both of the last two sections extend in a natural way for infinite sequences. The β -Fibonacci sequences are just as before,

$$F_m^{\{b\}} = \sum_{j=1}^m b_{j-1} F_{m-j}^{\{b\}}$$

starting with $F_0^{\{b\}} = 1$. The sum is always finite, and all that one needs is the first m bits of the (now infinite) bit-sequence $\{b\}$. The integer sequence still converges as before,

$$\beta = \lim_{m \rightarrow \infty} \frac{F_m^{\{b\}}}{F_{m-1}^{\{b\}}}$$

with β value is the one associated to $\{b\}$. The real number β and the bit sequence $\{b\}$ label exactly the same orbit.

The shift of eqn 30 is replaced by an infinite-dimensional matrix $\mathcal{H}_\beta = B^T / \beta$. The transpose puts the matrix into upper-Hessenberg form; the subdiagonal provides the shift. It is rescaled so that the largest eigenvalue is 1. To be explicit: given the bit-sequence $\{b\}$, the operator \mathcal{H}_β has the matrix elements

$$\begin{aligned} \langle 0 | \mathcal{H}_\beta | j \rangle &= \frac{b_j}{\beta} \\ \langle j+1 | \mathcal{H}_\beta | j \rangle &= \frac{1}{\beta} \end{aligned}$$

with all other entries being zero. It is perhaps useful to visualize this operator; in pictorial form, it is

$$\mathcal{H}_\beta = \frac{1}{\beta} \begin{bmatrix} b_0 & b_1 & b_2 & b_3 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & & \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix} \quad \text{with} \quad \psi = \begin{bmatrix} 1 \\ (\lambda\beta)^{-1} \\ (\lambda\beta)^{-2} \\ (\lambda\beta)^{-3} \\ \vdots \end{bmatrix}$$

where w is the shift state, satisfying $\mathcal{H}_\beta \psi = \lambda \psi$ whenever the eigenvalue λ can satisfy the top row:

$$\lambda = \sum_{j=0}^{\infty} \langle 0 | \mathcal{H}_\beta | j \rangle \psi_j = \frac{1}{\beta} \sum_{j=0}^{\infty} \frac{b_j}{(\lambda\beta)^j} = \lambda (E+1) = \lambda \left(1 - q^{\{b\}} \left(\frac{1}{\lambda\beta} \right) \right) = \lambda$$

with $q^{\{b\}}(\zeta)$ the same holomorphic function as defined earlier, in eqn 28 and with E as defined in eqn 20, where we'd previously identified solutions $E = 0$ as eigenvalues.

Explicit eigenfunctions on the unit interval can be recovered directly from this shift state. The eigenvector ψ can be expanded in terms of the canonical basis vectors e_j

so that $\psi = \sum_j \psi_j e_j$ and e_j is the vector with a one in the j 'th position, and zero everywhere else. Then make the identification

$$e_j \mapsto \varepsilon_j(y) = \Theta(t_j - y)$$

where $\varepsilon_j(y)$ is the step function previously given in 11, with $t_n = t_\beta^n(1)$ being the iterated endpoint from eqn 8. Alternately, after rescaling with $x = y\beta/2$ so as to work with the rescaled shift,

$$e_j \mapsto d_j(x) = \Theta(m_j - x)$$

with $m_j = T_\beta^j(\beta/2)$ being the iterated midpoint from eqn 9, and $d_j(x)$ the rescaled step function from eqn 19. With this identification, one has

$$\psi(x) = \sum_{j=0}^{\infty} d_j(x) \psi_j = \sum_{j=0}^{\infty} \frac{1}{(\beta\lambda)^j} \Theta(m_j - x)$$

which is just eqn 18 in a slightly different form. It satisfies $\mathcal{L}_\beta \psi = \lambda \psi$ and, for $\lambda = 1$ recovers the original Gelfond–Parry invariant measure from eqn 10.

The mapping $S: e_j \mapsto d_j(x)$ can be thought of as a similarity transform, so that

$$\mathcal{L}_\beta S = S \mathcal{H}_\beta$$

with \mathcal{L}_β the transfer operator from eqn 12. This mapping is not invertable; it is a projection from some larger space of functions on the unit interval, to those functions that can be represented with piecewise constant segments. Subsection 2.6 demonstrated that there are also piece-wise quadratic solutions, and a general form for piece-wise polynomial solutions will be given in section 6.4.

Clouding the issue is that when a bitsequence $\{b\}$ is not of finite length, the holomorphic function $q^{\{b\}}(\zeta)$ has a countable number of zeros, accumulating on a circular ring at $|\zeta| = 1$. These appear to be dense on the circle, and so Jentzsch's theorem blocks conventional analytic continuation to the region $|\zeta| > 1$. But clearly, \mathcal{L}_β has eigenvalues in this region; the piecewise quadratic eigenfunction exhibited in section 2.6 has the eigenvalue $\lambda = 1/\beta^2$ which corresponds to $\zeta = \beta > 1$. Classical analysis doesn't allow us into that region. This suggests some incomplete description or paradox.

The paradox is resolved in 6.4. The resolution maps the e_j to piece-wise polynomial functions. When the polynomials are of degree m , then \mathcal{L}_β has a block upper-triangular form, with the diagonal blocks containing $\beta^{-k} \mathcal{H}_\beta$, for $0 \leq k \leq m$. Off-diagonal blocks couple terms. This form reproduces the spectrum: when a bitsequence $\{b\}$ is not of finite length, the spectrum still accumulates on a ring at $|\zeta| = 1$, but this time, that ring corresponds to β^{-m} , so that there is an accumulation of eigenvalues at $|\lambda| \geq \beta^{-m}$. For piecewise-polynomial solutions, there are accumulation rings at each m .

These are the piecewise-polynomial solutions. In the $m \rightarrow \infty$ limit, the piecewise real-analytic solutions lie in the kernel $\lambda = 0$. In addition to these, there are also the piecewise-fractal solutions, termed “coherent states”, as they are (again) constructed as shifts. These belong to the continuous spectrum, and can be constructed for any $|\lambda| < 1$. This construction is given in section 6.6.

3.12 Equivalent labels for orbits

There are many equivalent ways of labeling the various expressions and properties under consideration. These are recapped here.

3.12.1 Orbits

For every given $1 < \beta < 2$ there is a unique orbit of midpoints $\{m_p\}$ given by $m_p = T_\beta(m_{p-1}) = T_\beta^p(m_0)$ and $m_0 = \beta/2$. The orbits are in one-to-one correspondence with β . The midpoints are the same as the Parry sequence; namely $T_\beta^p(\beta/2) = (\beta/2)t_\beta^p(1)$, recalling here the notation of eqn 8 and 10. Some orbits are of finite length; the rest are either eventually periodic, or are ergodic.

3.12.2 Orbit encoding

The midpoint generates a unique sequence of bits $\{b\} = \{b_0, b_1, \dots, b_k, \dots\}$ given by the left-right moves of the mid-point, as it is iterated. That is, $b_k = \Theta(m_k - 1/2)$ so that b_k is one if the midpoint is greater than half, else b_k is zero. Each bit-sequence is in one-to-one correspondence with β . Finite orbits have finite-length sequences.

3.12.3 Monotonicity

The compressor function $w(\beta) = \sum_k b_k 2^{-k}$ is a monotonically increasing function of β , so that values of $w(\beta)$ are in one-to-one correspondence with β .

3.12.4 Polynomial numbering

If the orbit is finite, then there exists a polynomial $p_n(z) = z^{k+1} - b_0 z^k - b_1 z^{k-1} - \dots - b_{k-1} z - 1$ with $k = 1 + \lceil \log_2(2n+1) \rceil$ being the length of the orbit. The positive real root r_n of $p_n(r_n) = 0$ is $\beta = r_n$. That is, the iteration of r_n will generate the finite-length bit-sequence $\{b\} = \{b_0, b_1, \dots, b_k\}$. The integer n is in one-to-one correspondence with the bit sequence, and with the value of β . The integer is explicitly given by $2n+1 = \sum_{j=0}^k 2^j b_j$.

If the orbit is not finite, there is a function $q^{\{b\}}(\zeta) = 1 - \sum_{j=0}^{\infty} b_j \zeta^{j+1}$ holomorphic on the unit disk, having one unique positive real zero $q^{\{b\}}(r_{\{b\}}) = 0$ where this $r_{\{b\}} = 1/\beta$ is the same β that generated the bit-sequence $\{b\}$. Iterating $r_{\{b\}}$ generates $\{b\}$. If $\{b\}$ is finite, then $q^{\{b\}}(\zeta) = \zeta^{k+1} p_n(1/\zeta)$, so these functions agree on finite-length sequences.

3.12.5 Brackets

If the orbit is finite, then there exists a unique bracketing relationship $\ell \Leftrightarrow n \Leftarrow \rho$ for which n is the polynomial index. The left and right bounds ℓ, ρ are strictly smaller indexes: $\ell < n$ and $\rho < n$, and even more strongly, $2\ell \leq n$ and $2\rho \leq n$ that have the property of bounding the positive real roots of the corresponding polynomials: $r_\ell < r_n < r_\rho$, with $p_\ell(r_\ell) = p_n(r_n) = p_\rho(r_\rho) = 0$.

3.12.6 Binary tree

Each bracket is in one-to-one correspondence with a node in the full, unbounded binary tree. Sub-brackets define left and right subintervals that are disjoint, and whose union makes up the whole interval. Every node in the full binary tree can be labeled with a unique sequence of left-right moves to get to that node. This places the brackets (and thus, the polynomials and the roots and the mid-point orbits) in unique, one-to-one correspondences with finite-length strings of L,R moves. Such strings are, in turn, in one-to-one correspondence with the dyadic rationals. The L,R strings are in one-to-one correspondence with the orbits $\{b\}$ but they are *not* numerically the same! This are distinct sequences! In particular, all possible L,R moves are allowed. Only a limited number of orbits $\{b\}$ are possible, as limited by necklace-counting considerations.

3.12.7 Baire sequences

If the orbit is finite, then there exists a unique integer sequence $[m_1, m_2, \dots, m_k]$ such that the index is given by $\eta[m_1, m_2, \dots, m_k]$. This is a bijection between all valid indexes and all possible finite-length sequences. Due to the bounding property of the brackets, limits of $k \rightarrow \infty$ can be taken, and these limits are unique. Thus, all β values, including those with non-finite orbits, can be placed in a one-to-one bijection with infinite-length sequences $[m_1, m_2, \dots] \in \mathbb{N}^\omega$.

3.12.8 Beta-Fibonacci sequences

If the orbit is finite, then there exists a sequence of integers $F^{\{b\}}$, the beta-Fibonacci sequence, that is in one-to-one correspondence with the finite bit sequence $\{b\} = b_0, b_1, \dots, b_k$, and with the value of β . There are also sequences for each infinite-length orbit $\{b\}$. These are briefly touched on, below.

3.12.9 Shift matrix

If the orbit is finite, then the finite bit sequence $\{b\} = b_0, b_1, \dots, b_k$ defines a lower-Hessenberg “golden shift” matrix B , as shown in eqn 30. The limit of $k \rightarrow \infty$ can be taken in a relatively straight-forward manner, given below.

3.12.10 Summary

To summarize: any one of these: the integer n , the polynomial $p_n(x)$, the bracket location in the binary tree, a dyadic rational, a point in Baire space, the integer sequence $F_m^{\{b\}}$, the orbit of midpoints $m_p = T^p(\beta/2)$, the orbit encoding $\{b\}$, the shift matrix B , the value of the compressor function $w(\beta)$ and, of course, β itself can each be used as a stand-in for the others, as they are all in one-to-one correspondence. Specifying one determines the others; all uniquely map to one-another. The formulas that provide maps between each of these can all be given in closed form, except for the handful of recursively-defined formulas. The recursive formulas are all invertible; thus they are computable (decidable). They are all equivalent labels. Fashionably abusing notation, $n \equiv p_n(x) \equiv r_n \equiv \{b\} \equiv F_m^{\{b\}} \equiv m_p \equiv w(\beta) \equiv \beta \equiv B$.

An explicit expression relating the orbit encoding and the orbit can be read off directly from eqn 7. Plugging in,

$$m_p = T_\beta^{p+1} \left(\frac{\beta}{2} \right) = \frac{\beta}{2} \left[\beta^{p+1} - \sum_{j=0}^p b_j \beta^{p-j} \right] \quad (31)$$

for $p < k$ the length of the bit sequence, and $m_k = T_\beta^{k+1} (\beta/2) = \beta p_n(\beta) / 2 = 0$ terminating, since β is the positive root of $p_n(x)$.

Four of the correspondences given above ask for finite orbits. Three of these can be extended to non-finite orbits in an unambiguous and uncontroversial way. The extensions are covered in the next two sections. The fourth is the numbering n of the finite orbits. These are countable; there is no way to extend the counting number n to the non-finite orbits. Indeed, there are too many: the non-finite orbits are uncountable.

4 Eventually-periodic orbits

The previous section characterized the set of β values which have midpoint orbits of finite length. Another interesting class is the set of eventually-periodic orbits: orbits of infinite length, settling down to a stable, periodic cycle after an initial bout of chaotic motion. These β values occur as the roots of a slightly different set of self-describing polynomials, as a sum of two parts: one for the initial chaotic motion, and a second polynomial for the cyclic motion. These can be (monotonically) paired with the rational numbers, with infinite paths through the binary tree, and with locations on the comb function. They naturally extend the bracket mapping from the dyadic rationals to all rational numbers.

4.1 Bitsequence polynomials

Consider the set of all eventually-periodic bit-sequences. These consist of a leading chaotic prefix of length L followed by a periodic orbit of length N . Such sequences can be placed in one-to-one correspondence with the rationals, in the conventional fashion. Select such a sequence $\{b_k\}$. The cyclic condition has that $b_k = b_{k+N}$ for all $k \geq L$. Split this into two parts: a finite length- L digit sequence $d_k = b_k$ for $k < L$ and a finite length- N cyclic bit sequence $c_k = b_{k+L}$ for $k < N$. Associated to this is a rational number $x = \sum_{k=0}^{\infty} b_k 2^{-k-1}$. To get it's value, write

$$\begin{aligned} S(\beta) &= \sum_{k=0}^{\infty} b_k \beta^{-k} \\ &= \sum_{k=0}^{L-1} b_k \beta^{-k} + \sum_{k=L}^{L+N-1} b_k \beta^{-k} + \sum_{k=L+N}^{L+2N-1} b_k \beta^{-k} + \dots \\ &= \sum_{k=0}^{L-1} b_k \beta^{-k} + \frac{\beta^N}{\beta^N - 1} \sum_{k=L}^{L+N-1} b_k \beta^{-k} \\ &= \frac{1}{\beta^{L-1}} \sum_{k=0}^{L-1} d_k \beta^{L-k-1} + \frac{1}{\beta^{L-1} (\beta^N - 1)} \sum_{k=0}^{N-1} c_k \beta^{N-k-1} \end{aligned}$$

Plugging in $\beta = 2$ gives ratios of whole numbers: a rational that corresponds to this bit-sequence.

Any given $1 \leq \beta \leq 2$ generates a mid-point orbit bit-sequence. Sadly, we've introduced too many different but equivalent notations for this. Starting with the mid-point $x = m_0 = \beta/2$, the characteristic bit-sequence is

$$b_n = \Theta \left(m_n - \frac{1}{2} \right) = \Theta \left(T_{\beta}^n \left(\frac{\beta}{2} \right) - \frac{1}{2} \right) = d_n \left(\frac{1}{2} \right) = k_n \left(\frac{\beta}{2} \right) = \varepsilon_n \left(\frac{1}{\beta} \right)$$

Such orbits have the self-describing property, that $\beta = S(\beta)$. This follows from eqn 6, as well as other identities. For the eventually-periodic orbits, $S(\beta)$ is a ratio of polynomials. It can clearly be placed in one-to-one correspondence with the rationals.

It is not hard to show that it has one unique real root $1 < \beta \leq 2$, and so these are in one-to-one correspondence as well.

This can be related to earlier notation. Writing $\beta - S(\beta) = 0$ and then multiplying through by the denominator,

$$(\beta^N - 1) \left(\beta^L - \sum_{k=0}^{L-1} d_k \beta^{L-1-k} \right) - \sum_{k=0}^{N-1} c_k \beta^{N-1-k} = 0$$

This touches the earlier notation for the polynomials describing finite orbits:

$$p^{\{d_0 \cdots d_{v-1}\}}(\beta) = \beta^v - \sum_{k=0}^{v-1} d_k \beta^{v-1-k}$$

and so the eventually-periodic orbits are the roots of

$$(\beta^N - 1) p^{\{d_0 \cdots d_{L-1}\}}(\beta) + p^{\{c_0 \cdots c_{N-1}\}}(\beta) - \beta^N = 0$$

Finite-length orbits have $N = 1$ and $c_0 = 0$, so that the cyclic term vanishes, and the earlier form is recovered, after ignoring an extra factor of $\beta - 1$.

Finite-length orbits can be converted to infinite-length periodic orbits simply by setting the last bit to zero, and then repeating the bit sequence cyclically. Recall, finite orbits always have the last bit equal to one, so this operation is always unambiguous. Restating this explicitly: if $b_0 b_1 \cdots b_{v-1}$ is a finite-length orbit, then set $L = 0$ and $N = v$ and $c_k = b_k$ for $k < v - 1$ and finally $c_{v-1} = 0$. Plugging through just yields the usual polynomial for finite orbits.

Collecting terms of the same order, write

$$0 = \beta^{N+L} - \sum_{k=0}^{N+L-1} a_k \beta^{N+L-1-k}$$

The coefficients are then

$$a_k = d_k + c_{k-L} + \delta_{k-N-1}^0 - d_{k-N}$$

with δ the Kronecker delta. The only coefficients that can appear are in the set $a_k \in \{-1, 0, 1, 2\}$, and at most one coefficient can be 2; it is specifically a_{N-1} , which can only ever be 1 or 2. This can be best understood visually, lining up columns. For $L > N$, get a_k by summing the columns:

$$\begin{array}{cccccccc} & d_0 & \cdots & & \cdots & d_{L-1} & c_0 & \cdots & c_{N-1} \\ + & 0 & \cdots & 0 & 1 & -d_0 & \cdots & & -d_{L-1} \\ \hline & a_0 & & \cdots & & a_N & & \cdots & a_{N+L-1} \end{array}$$

The top row shows the prefix string and the first run of the cyclic string. The second row shows minus the prefix string, shifted all the way to the right, establishing alignment at the right side. Just before it is a lone 1, and then padded on the left with zeros. The bottom row is just the sum of the two rows above it. Clearly, the coefficient a_{N-1} can only ever be 1 or 2. It is exactly the same for $L < N$, with the bottom row being shorter:

$$\begin{array}{cccccccccccc}
& d_0 & \cdots & d_{L-1} & c_0 & & \cdots & & c_{N-1} & & & & \\
+ & 0 & & \cdots & & 0 & 1 & -d_0 & \cdots & -d_{L-1} & & & \\
\hline
& a_0 & & & \cdots & & & a_N & \cdots & a_{N+L-1} & & &
\end{array}$$

4.2 Examples

Not all possible bit-sequences $\{b_k\}$ are allowed. They must be self-describing, so that the root of $\beta - S(\beta) = 0$, when iterated, generates $\{b_k\}$. For orientation, some examples are shown below.

$\frac{p}{q}$	$\frac{p+q}{2q}$	m, c	$b_0 b_1 \dots$	L	N	polynomial	β
0	1/2	1,0	1 $\bar{\cdot}$	1	1	$\beta - 1 = 0$	1
1	1	0,1	$\bar{1}$	0	1	$\beta - 2 = 0$	2
1/2	3/4	3,0	11 $\bar{\cdot}$	2	1	$\beta^2 - \beta - 1 = 0$	$\varphi = 1.61803\dots$
1/3	2/3	0,2	$\bar{10}$	0	2	above	above
2/3	5/6	1,2	1 $\bar{10}$	1	2	$\beta^3 - \beta^2 - 2\beta + 1 = 0$	$1.80\dots = 2\cos\frac{\pi}{7}$
5/6	11/12	3,2	11 $\bar{10}$	2	2	$\beta^4 - \beta^3 - 2\beta^2 + 1 = 0$	1.90516616775
11/12	23/24	7,2	111 $\bar{10}$	3	2	$\beta^5 - \beta^4 - 2\beta^3 + 1 = 0$	1.95468312004
						...	
1/4	5/8	5,0	101 $\bar{\cdot}$	3	1	$\beta^3 - \beta^2 - 1$	1.46557123187
1/7	4/7	0,4	$\bar{100}$	0	3	$\beta^3 - \beta^2 - 1$	above
4/7	11/14	1,4	1 $\bar{100}$	1	3	$\beta^4 - \beta^3 - \beta^2 - \beta + 1 = 0$	1.72208380573
11/14	25/28	3,4	11 $\bar{100}$	2	3	$\beta^5 - \beta^4 - \beta^3 - 2\beta^2 + \beta + 1 = 0$	1.87134931301
25/28	53/56	7,4	111 $\bar{100}$	3	3	$\beta^6 - \beta^5 - \beta^4 - 2\beta^3 + \beta + 1 = 0$	1.93992448793
2/7	9/14	1,2	1 $\bar{010}$	1	3	$\beta^4 - \beta^3 - 2\beta + 1 = 0$	1.55897987798
9/14	23/28	3,2	11 $\bar{010}$	2	3	$\beta^5 - \beta^4 - \beta^3 - \beta^2 + 1 = 0$	1.77847961614
4/5	9/10	1,12	111 $\bar{00}$	1	4	$\beta^5 - \beta^4 - \beta^3 - \beta^2 - \beta + 1 = 0$	1.88320350591

The rationals in the first column are from the set $\frac{0}{1} \leq \frac{p}{q} \leq \frac{1}{1}$. The β bit-sequences always have $b_0 = 1$ and thus, the second column shows $S(2) = (p+q)/2q \geq 1/2$. The third column m, c shows the prefix and the cyclic part as integers. The fourth column shows the actual bitsequence. An overline is drawn over the repeating digits. If there is one repeating digit, and it is zero, it is written as $\bar{\cdot}$ so as to keep things a bit more readable. The prefix must always start with a 1, and so the integer m is unique. The length of the prefix is in the L column. The cycle might have leading zeros, and so specifying c is not enough; a cycle length is required, given in the N column. The corresponding polynomial and its root are given in the last two columns.

Notes:

- The sequence $1\overline{10}$ is described in [OEIS A160389](#).
- The sequence $10\overline{10}$ is obviously not allowed, as it is the same as $\overline{10}$.
- The sequence $100\overline{10} = 100\overline{1}$ is not allowed. In the unreduced form, replacing $\overline{10}$ by $1\overline{10}$ gives $100\overline{10} \mapsto 10011 = 19$, which wasn't allowed as a finite orbit.
- The sequence $\overline{101\overline{10}}$ is not allowed; it reduces to a finite form $101\overline{10} \mapsto 10111 = 23$ that is a disallowed finite orbit.
- These last two observations are a coincidence, and do not hold in the general case. There are good periodic orbits that have disallowed finite versions, and vice-versa.
- The sequence $10\overline{100}$ is not allowed, because it is reducible: $10\overline{100} = \overline{1010}$.
- $4/7$ gives $1\overline{100}$ seems to be [OEIS A289917](#) which is $(1 + \sqrt{13} + \sqrt{2\sqrt{13}-2})/4$.
- $4/5$ gives $1\overline{1100}$ seems to be [OEIS A289915](#) which is $(1 + \sqrt{2} + \sqrt{2\sqrt{2}-1})/2$.
The polynomial factors as $\beta^5 - \beta^4 - \beta^3 - \beta^2 - \beta + 1 = (\beta + 1)(\beta^4 - 2\beta^3 + \beta^2 - 2\beta + 1)$.
- Rationals with prefix $m = 0$ correspond to the finite orbits.
- It seems the prefix $m = 2^{L-1}$ never occurs, when $L > 1$.
- The prefix is often but not always odd. Counterexamples include $7/12 = 1100\overline{1}$ and $13/20 = 110\overline{1001}$.
- The first time that $m = 5$ occurs is for $9/28 = 1010\overline{10}$.

Not all rationals give valid bit-sequences. The set of valid rationals are exactly the ones that are in the set $\mathbb{Q} \cap \overline{\theta}$ where $\overline{\theta}$ is the infinite comb, given in eqn 27, and depicted visually in figure 12. The comb is extremely fractal, and working directly with $\mathbb{Q} \cap \overline{\theta}$ would be a chore. Fortunately, the bracket map saves the day: it is a map of the entire unit interval onto $\overline{\theta}$. The bracket map, shown in figure 14, maps rationals to rationals; it provides a map $\mathbb{Q} \cap [0, 1] \rightarrow \mathbb{Q} \cap \overline{\theta}$. Thus, it provides a better, though more indirect way, of describing the ultimately-periodic orbits. Some examples are reviewed in the section after next.

4.3 Rational to periodic binary

As a practical matter, it is computationally useful to convert a given fraction p/q into the binary prefix m , its length L , and the cyclic part c , and its length N . This is not

hard, but also not entirely easy, and so is presented here. Write

$$\begin{aligned} \frac{p+q}{2q} &= \frac{m}{2^L} + \frac{1}{2^L} \left(\frac{c}{2^N} + \frac{c}{2^{2N}} + \frac{c}{2^{3N}} + \cdots \right) \\ &= \frac{1}{2^L} \left(m + \frac{c}{2^N} \left(1 + \frac{1}{2^N} + \frac{1}{2^{2N}} + \cdots \right) \right) \\ &= \frac{1}{2^L} \left(m + \frac{c}{2^N - 1} \right) \\ &= \frac{1}{2^L} \cdot \frac{m(2^N - 1) + c}{2^N - 1} \end{aligned}$$

In the last line, both numerator and denominator are integers. This allows the following algorithm:

1. Find $\gcd(p+q, 2q)$ and so reduce to lowest terms $a/b = (p+q)/2q$.
2. Factor $b = 2^L b'$ to obtain $L \geq 0$.
3. Solve $(2^N - 1) \bmod b' = 0$ for the smallest positive $N \geq 1$.
4. Define $r = (2^N - 1)a/b'$.
5. Solve $(r - c) \bmod (2^N - 1) = 0$ for the smallest c such that $0 \leq c < 2^N$.
6. Define $m = (r - c)/(2^N - 1)$.

This provides all four integers m, L, c, N that define the cyclic expansion.

4.4 The Bracket Map

The bracketing relationship provided a mapping to the dyadic rationals, via moves on the binary tree. What do ultimately-periodic moves correspond to, on that tree?

The left and right moves L, R on the valid-index binary tree were $L: (\ell \Rightarrow f \Leftarrow \rho) \mapsto (\ell \Rightarrow 2f \Leftarrow f)$ and $R: (\ell \Rightarrow f \Leftarrow \rho) \mapsto (f \Rightarrow \Lambda(f) \Leftarrow \rho)$. (The symbol R is used in place of \mathfrak{R} in this section, for ease of typography. They are isomorphic.) Consider starting at $\infty \Rightarrow 1 \Leftarrow 0$ and applying a sequence of alternating left and right moves. This is shown in the table below:

position	move	bracket	β
1/2	-	$\infty \Rightarrow 1 \Leftarrow 0$	1.61803398...
1/4	<i>L</i>	$\infty \Rightarrow 2 \Leftarrow 1$	1.46557123...
3/8	<i>LR</i>	$2 \Rightarrow 10 \Leftarrow 1$	1.57014731...
5/16	<i>LRL</i>	$2 \Rightarrow 20 \Leftarrow 10$	1.53849659...
11/32	<i>LRLR</i>	$20 \Rightarrow 82 \Leftarrow 10$	1.56175206...
21/64	<i>LRLRL</i>	$20 \Rightarrow 164 \Leftarrow 82$	1.55392112...
43/128	<i>LRLRLR</i>	$164 \Rightarrow 658 \Leftarrow 82$	1.55970265...
85/256	<i>LRLRLRL</i>	$164 \Rightarrow 1316 \Leftarrow 658$	1.55767530...
171/512	<i>LRLRLRLR</i>	$1316 \Rightarrow 5266 \Leftarrow 658$	1.55917021...
1/3	\overline{LR}	—	1.55897978...

The column labeled “position” is the location in the dyadic tree. By convention, 1/2 is at the top, and 1/4 lies to the left, and 3/4 lies to the right. The column labeled “move” consists of the sequence of left-right moves to get to a given tree position. For n moves, encoded as a binary integer m , the corresponding position in the dyadic tree is $(2m + 1)/2^n$. The bracket is likewise a position in the valid-index binary tree. The string of moves to get to that location are written in reverse applicative order, so that the first letter in the string is the first move. If the moves are treated as functions to be composed and applied, then the string needs to be reversed to get the applicative order. The beta is the beta value at the center of that bracket. Based on the beta value, we conclude that $G : 1/3 \mapsto 2/7$.

A second example helps cement the idea.

position	move	bracket	β
1/2	-	$\infty \Rightarrow 1 \Leftarrow 0$	1.61803398...
1/4	<i>L</i>	$\infty \Rightarrow 2 \Leftarrow 1$	1.46557123...
1/8	<i>LL</i>	$\infty \Rightarrow 4 \Leftarrow 2$	1.38027756...
3/16	<i>LLR</i>	$4 \Rightarrow 36 \Leftarrow 2$	1.44326879...
5/32	<i>LLRL</i>	$4 \Rightarrow 72 \Leftarrow 36$	1.42705896...
11/64	<i>LLRLR</i>	$72 \Rightarrow 580 \Leftarrow 36$	1.43949911...
21/128	<i>LLRLRL</i>	$72 \Rightarrow 1160 \Leftarrow 580$	1.43591015...
43/256	<i>LLRLRLR</i>	$1160 \Rightarrow 9284 \Leftarrow 580$	1.43866733...
85/512	<i>LLRLRLRL</i>	$1160 \Rightarrow 18568 \Leftarrow 9284$	1.43784133...
1/6	\overline{LLR}	—	1.43841656...

Based on the convergent, we conclude that $G : 1/6 \mapsto 2/15$, deduced below. In this case, $1/6 \notin \overline{\theta}$ and so the raw fraction 1/6 does not generate a self-describing orbit.

But $2/15$ is self-describing. Of course, that is the entire intent of the bracket map. Originally formulated as the good-index map $G : \mathbb{N} \rightarrow \Psi$ which maps natural numbers to brackets, the extension \mathbb{G} is the “good rational” map $\mathbb{G} : \mathbb{Q} \rightarrow \mathbb{Q} \cap \bar{\theta}$, which maps rationals to those that are self-describing. This is the same map as visualized in figure 14.

4.5 Bracketing Examples

A curated collection of examples of rationals passed through the bracketing function is given below. It serves mostly to give a sense of the patterns that develop, as well as counterexamples that defy simple patterns.

$\frac{a}{b}$	Ψ	moves	$\frac{p}{q} = \mathbb{G}\frac{a}{b}$	id	orbit	β
0/1	-1	ε	0/1	y	$1\bar{\cdot}$	1
$1/2^n - \varepsilon$		$\approx RL \cdots L\bar{R}$	$1/(2^{n+1} - 1) - \delta$	-	$\approx 1\bar{0} \cdots 0\bar{1}$	
$1/2^n$	2^{n-1}	$RL \cdots L\bar{R}$	$1/2^n$	y	$1\bar{0} \cdots 0\bar{1}$	
$1/8 - \varepsilon$		$\approx RLL\bar{R}$	$1/15 - \delta$	-	$\approx 1\bar{000}\bar{1}$	< 1.380277
1/8	4	RLL	1/8	y	1001	1.38027756
1/7		$RLL\bar{R}$	4/31	-	$1\bar{00100}$	1.42109608
1/6		$RLL\bar{R}$	2/15	-	$1\bar{0010}$	1.43841656
1/5		$RLL\bar{R}\bar{R}$	12/85	-	$1\bar{00100100}$	1.45394278
$1/4 - \varepsilon$		$\approx RLL\bar{R}$	$1/7 - \delta$	-	$\approx 1\bar{00}\bar{1}$	< 1.4655712
1/4	2	RL	1/4	y	$101\bar{\cdot}$	1.46557123
2/7		$R\bar{L}\bar{R}\bar{L}$	4/15	-	$1\bar{0100}$	1.52626195
1/3		$R\bar{L}\bar{R}$	2/7	-	$1\bar{010}$	1.55897987
2/5		$R\bar{L}\bar{R}\bar{L}$	20/63	-	$1\bar{010100}$	1.58925391
3/7		$R\bar{L}\bar{R}\bar{R}$	10/31	-	$1\bar{01010}$	1.60022189
$1/2 - \varepsilon$		$\approx RL\bar{R}$	$1/3 - \delta$	-	$\approx 1\bar{0}\bar{1}$	$< 1.618 \cdots$
1/2	1	R	1/2	y	$11\bar{\cdot}$	$\varphi = 1.618 \cdots$
6/11			8900/16383	-		1.69971346
5/9		$R\bar{R}L\bar{L}\bar{R}\bar{R}$	40/73	-		1.70348856
4/7		$R\bar{R}L\bar{L}$	4/7	y	$1\bar{100}$	1.72208380
3/5		$R\bar{R}L\bar{L}\bar{R}$	25/42	-	$1\bar{1001100}$	1.74720863
9/14		$R\bar{R}L\bar{R}\bar{L}$	9/14	y	$1\bar{1010}$	1.77847961
2/3		$R\bar{R}L$	2/3	y	$1\bar{10}$	1.80193773
7/10		$R\bar{R}L\bar{R}\bar{L}$	43/62	-		1.82000973

$\frac{a}{b}$	Ψ	moves	$\frac{p}{q} = G\frac{a}{b}$	id	orbit	β
5/7		$RRL\bar{R}$	7/10	-	110110	1.82651577
3/4	3	RR	3/4	y	111	1.83928675
7/9		$\overline{RRRLLL\bar{R}}$	227/292	-		1.86625406
11/14		$RR\bar{R}\bar{L}\bar{L}$	11/14	y	11100	1.87134931
4/5		$RR\bar{R}\bar{L}$	4/5	y	11100	1.88320350
5/6		$RR\bar{R}\bar{L}$	5/6	y	1110	1.90516616
6/7		$RR\bar{R}\bar{L}$	6/7	y	1110	1.92128960
8/9		$RRR\bar{R}\bar{L}\bar{L}$	8/9	y	1111000	1.93762945
25/28		$RRR\bar{R}\bar{L}\bar{L}$	25/28	y	111100	1.93992448
9/10		$RRR\bar{R}\bar{L}\bar{L}$	9/10	y	111100	1.94470383
10/11			10/11	y		1.94988340
11/12		$RRR\bar{R}\bar{L}$	11/12	y	11110	1.95468312
1/1	0	\bar{R}	1/1	y	$\bar{1}$	2

Table legend.

- The first column shows selected interesting rationals. Every possible rational is allowed here, and will generate an ultimately-periodic orbit.
- The second column shows the corresponding index Ψ ; only seven are shown, for the finite orbits that correspond to the dyadic rationals; all others have an infinite limit.
- The third column shows the bracket moves generated by that rational. These are obtained as the binary expansion of the fraction a/b , or, more properly, the expansion for $(a+b)/2b$. Thus, the first move is always R , to arrive from the far left to the center of the tree. The expansions for 6/11 and 10/11 were left blank; these are pointlessly long strings that would have cluttered the table.
- The fourth column shows the result of the “good rational” map $G : \mathbb{Q} \rightarrow \mathbb{Q} \cap \bar{\theta}$ that takes $G : \frac{a}{b} \mapsto \frac{p}{q}$. Thus, by definition, $p/q \in \bar{\theta}$. The fractions are those obtained by extending the “good index” map $G : \mathbb{N} \rightarrow \Psi$ to its limit points. A graph of the fourth column, relative to the first, is given in figure 15. Although that figure was prepared for the finite orbits, it is exactly the same for the periodic orbits.
- Some of the discontinuities visible in 15 are shown here, in the rows with $\delta, \varepsilon \ll 1$ in them. For example, the top of the tree is 1/2, which has a finite orbit $1\bar{0}$. It can also be written as an infinite cyclic orbit $1\bar{0}\bar{1} = 1/3$ and these two distinct

binary strings are mapped to distinct fractions; thus, the discontinuities. Such a discontinuity appears at every finite orbit.

- Notable rows are $6/11$, which has a horribly large denominator; and $7/9$ is the runner-up. The large denominator is due to leader heights that are greater than one, that are encountered during the expansion. Compare rows $7/9$ to $8/9$. Both have long move strings, but $G(7/9) = 7/9$. Similarly, $6/11$ and $10/11$ have intolerably long move strings, but $G(10/11) = 10/11$.
- The fifth column has a 'y' in it, whenever $G(a/b) = a/b$; when the first and fifth column are equal. Note that most 'y's occur at the end of the table, rather than at the beginning.
- The sixth column shows the bitsequence $b_0b_1 \dots$ for the binary expansion of p/q . Some rows are blank; these rows had horribly long expansions that provide no intuition.
- The seventh and rightmost column shows the corresponding β value for the orbit in the previous column.

Perhaps most notable is that the good map G maps rationals to rationals. Reassuringly, it extends from the dyadic rationals, to all rationals. Formal lemmas and proofs follow in a later section.

5 Formalities

Theorems, lemmas, proofs. A number of observations, claims and assumptions have been made. They seem “obviously true” from numerical work, but are lacking a formal proof. These are sketched below. As sketches, this chapter is still quite informal; the difference is that it uses a more sophisticated vocabulary to say things in a more abstract manner.

- Theorem: The finite orbits are dense in $1 \leq \beta \leq 2$. Proof: Provided by the bracketing relation, which states that between any two endpoints, there’s another orbit strictly in the middle. We even have more: the estimate $v^{1/v}$ from the limit diagram, showing how they don’t want to accumulate near $\beta = 1$.
- Lemma: The bracketed roots are always in strict ascending order, with $r_\ell < r_f < r_\rho$.
- Lemma: The valid-index map G is a bijection. Proof: Leadership function is monotonically increasing.
- Theorem: The good map G always maps rationals to rationals. More precisely, those rationals are always inside of $\bar{\theta}$ so that $G : \mathbb{Q}_I \rightarrow \bar{\theta} \cap \mathbb{Q}$ is a bijection, except at the dyadics, where we have a countable freedom of finite orbits and one infinite-periodic orbit to choose from.
- Theorem: $G : [0, 1] \rightarrow \bar{\theta}$ is a bijection for all reals $[0, 1] \subset \mathbb{R}$. That is, the closure to reals works as expected. This follows because the roots form a countable dense subset, the reals are separable, and the function is continuous.
- Proper diligence requires distinguishing $\bar{\mathfrak{B}}_{\mathbb{Q}}$ from $\bar{\mathfrak{B}}_{\mathbb{R}}$ and also $\bar{\theta}_{\mathbb{Q}}$ from $\bar{\theta}_{\mathbb{R}}$.

The remainder of this section consists of fragments of proofs of the above theorems and lemmas. If these were stitched together properly, whole proofs should emerge. As it is, it’s more an outline or sketch.

5.1 The Rational Comb

The dyadic comb, of figure 12, shows where valid finite orbits can occur. Its limit to infinite rank gives the infinite comb θ of eqn 27. The only valid rationals, that is, those giving self-describing orbits, are those that belong to $\mathbb{Q} \cap \theta$.

This can be understood as a limiting procedure. Every rational can be approximated by a dyadic rational; each set is dense in the other. The dyadic rationals are finite walks down the binary tree; the non-dyadic rationals are infinite walks. The infinite comb corresponds to the trimmed tree, described earlier. The trimming always maintains branches of unbounded length. These have a limit; in the limit, some of these will correspond to rationals. These are exactly the rationals that correspond to the ultimately-periodic orbits. The correspondence goes in both directions: given an infinite path down the tree, it is sufficient to truncate it to be of finite length. By construction, the truncated path corresponds to a valid dyadic rational. To recap: going

in one direction, there is a sequence of finite orbits that converge, in the limit, onto an eventually-periodic orbit. Conversely, given such an orbit, every truncated version thereof is valid.

5.2 Closures

The ultimately-periodic orbits live inside the closure $\overline{\mathfrak{B}} \subset \overline{\mathbb{B}}$. Write $\mathfrak{C} \subset \overline{\mathbb{B}}$ for the cyclic (ultimately periodic) orbits; these are infinite-length strings describing moves down the infinite binary tree. As already noted, $\mathfrak{C} \cap \mathbb{B} \subset \mathfrak{B}$; that is, every infinite-length cyclic orbit, when truncated to finite length, is a valid finite orbit. The closure $\overline{\mathfrak{B}}$ consists of all self-describing bit-strings obtained as solutions to $\beta = S(\beta)$ where $S(\beta) = \sum_{k=0}^{\infty} b_k \beta^{-k}$. The cyclic orbits are just a special case: $\mathfrak{C} \subset \mathfrak{B}$. There are presumably many more, uncountably many chaotic orbits in $\overline{\mathfrak{B}}$.

Just a little bit more machinery is needed. Let 2^ω be Cantor space, the space of infinitely-long binary strings. Let $\chi : 2^\omega \rightarrow [0, 1]$ be the canonical mapping of Cantor space to the unit interval: $\chi : (b_0 b_1 \dots) \mapsto \sum_k b_k 2^{-k-1}$. Note that this mapping manages to miss all the dyadic rationals, as these correspond to finite strings, of which there aren't any in the Cantor space. Let $2^{<\omega} = 2^*$ be the set of finite-length binary strings, and allow the symbol χ to perform double-duty, by writing $\chi : 2^* \rightarrow \mathbb{D}$ to map finite strings to dyadics. These binary strings can also be interpreted as strings in the symbols L, R , so that they are moves on the tree. Write $\iota : 2^* \rightarrow \mathbb{B}$ that maps the empty string to the root of the tree, and likewise $\iota : 2^\omega \rightarrow \overline{\mathbb{B}}$ that maps infinite strings to the leaves of the infinite tree.

The comb was a subset of the unit interval: $\overline{\theta} \subset [0, 1]$. The claim is that $\overline{\theta} = \chi \iota \overline{\mathfrak{B}}$. The mapping is onto, but not one-to-one: The finite orbits all have corresponding cyclic orbits, all of which are distinct elements in $\overline{\mathfrak{B}}$ but map to the (finite-orbit) polynomial, and thus the same root.

The good-index function provided a bijection $\mathfrak{B} = \eta G \eta^{-1} \mathbb{B}$ between the trimmed and untrimmed finite, unbounded trees. The claim is that this can be extended to a bijection $\overline{\mathfrak{B}} = \overline{G} \overline{\mathbb{B}}$. This requires a short detour into topology. The standard (weak) topology on Cantor space is the product topology. The base for the topology may be taken as the set of all finite strings, followed by an infinite number of don't-care markers. These are the open sets; the full topology is the finite intersection and infinite union of the open sets in this base.

Every set in the base of the topology is represented by a finite string. The function η^{-1} maps this to an integer. The function G is defined on all integers; it returns an integer, which is mapped by η to a finite string and thus an open set. To conclude, $\eta G \eta^{-1}$ maps open sets to open sets, and it is defined on *every* open set in the topology. Thus, it is safe to write $\overline{G} = \eta G \eta^{-1}$, as it is defined everywhere; all of $\overline{\mathbb{B}}$ is in its domain. The range of \overline{G} can be taken to be the definition of the closure $\overline{\mathfrak{B}}$. This has several benefits: it avoids having to make funny arguments that pass through the reals via the comb, and it also makes clear that \overline{G} is a bijection. Note \overline{G} is not continuous, in the context that $\overline{\mathfrak{B}} \subset \overline{\mathbb{B}}$, since \overline{G}^{-1} is defined almost nowhere on $\overline{\mathbb{B}}$.

The points r_n are dense in the interval $[1, 2]$. Within each rank v they are monotonically increasing. The bracketing relation guarantees that, for all r_n in the subtree, these

are all contained within the endpoints of the bracket, and that they are always totally ordered, by the natural tree-ordering. The notion of open sets is compatible. Given a basic open set in $\overline{\mathbb{B}}$, the function $\rho = r \circ G \circ \eta^{-1}$ maps it to an open set $U \subset [1, 2]$. Claim that the inverse map also maps open sets to open sets, and thus ρ is continuous. This is established as follows. Since the r_n are dense in the interval $[1, 2]$, any open set $U \subset [1, 2]$ can be written as a countable union of brackets. The function ρ^{-1} is defined on all brackets, and by bracketing, it is defined on all $U \subset [1, 2]$ and by bracketing $\rho^{-1}U$ is an open set in the product topology on $\overline{\mathbb{B}}$.

The goal is to extend this reasoning to $\rho : \mathbb{D} \rightarrow [1, 2]$, given by $\rho = \rho \circ \eta \circ \delta^{-1} = r \circ G \circ \delta^{-1}$ so as to obtain a continuous, monotonic function $\bar{\rho} : [0, 1] \rightarrow [1, 2]$ on the unit interval. And we are done, more or less.

The mapping χ maps the basic open sets of $\overline{\mathbb{B}}$ to open subintervals of the unit interval; more precisely, to subintervals with dyadics at each end. These are precisely the sets $I(m, v)$ defined in eqn 26. They allowed the infinite comb to be built up from unbounded-length but finite strings in \mathfrak{B} .

The base of the product topology was mapped to open intervals on the real-number line; and so $\bar{\rho}$ is continuous on the reals. This can also be seen in a different way. The standard measure μ on the reals says that $\mu I(m, v) = 2^{2-v}$. The function G is built from left and right moves. The left moves bump the rank by one, and so always map to sets that are exactly half the size on the real number line. The right moves are given by the leadership function: $\Lambda \Psi_v \subset \bigcup_{h=0}^{\infty} \Psi_{v+1+h}$ for each rank v . The resulting sets are (at a minimum) half the size, and possibly larger, as the union of additional small intervals. This allows a conventional delta-epsilon proof of continuity to go forward: for each epsilon, one can choose a rank v where the basis sets $I(m, v)$ are smaller than epsilon. The function G maps them to other open sets that are strictly smaller. Strictly, because the height of a leader is always a finite number; the union is written as a union over all possible heights, but the union is always finite.

5.3 Proof that roots are correctly bounded

Theorem: The bracket relationship gives roots with $r_\ell < r_f < r_\rho$ being strict inequalities.

Proof: Recursion starts with $\infty \Leftrightarrow 1 \Leftarrow 0$. The bracket moves are

$$\begin{aligned} L : (\ell \Leftrightarrow f \Leftarrow \rho) &\mapsto (\ell \Leftrightarrow 2f \Leftarrow f) \\ \mathfrak{R} : (\ell \Leftrightarrow f \Leftarrow \rho) &\mapsto (f \Leftrightarrow \Lambda(f) \Leftarrow \rho) \end{aligned}$$

So $p_1 = \beta^2 - \beta - 1$ and $p_\infty = p_{1/0} = \beta - 1$ and $p_{1/\infty} = \beta - 2$. Must show that bracket bounds are respected in all four cases. The proofs depend on having $p'_n(\beta) > 0$ for all $1 \leq \beta \leq 2$.

Case 1: Show that $r_{2f} < r_f$. Proof: The L move gives $p_{2f} = \beta(p_f + 1) - 1$. So $1 = r_{2f}(p_f(r_{2f}) + 1)$ but since $r_{2f} > 1$, must have $p_f(r_{2f}) < 0$ and since $p'_n > 0$, conclude that $r_{2f} < r_f$.

Case 2: Show that $r_\ell < r_{2f}$. Proof: For any $\ell \Leftrightarrow f \Leftarrow \rho$ it is always the case that ρ is in the tree rooted by ℓ . This can be seen by examining the two possibilities arising

from $\infty \Leftrightarrow 1 \Leftrightarrow 0$ and then keeping in mind all trees are isomorphic. The path to ρ from it's root ℓ is $\rho = L^n R \ell$ for some $n \geq 0$. Thus $p_f = \beta^{n+1} p_\ell - 1$ and so $p_{2f} = \beta^{n+2} p_\ell - 1$. But $p_\ell(r_\ell) = 0$ and so $p_{2f}(r_\ell) = -1 < 0$. Since $p'_n(\beta) > 0$, conclude that $r_\ell < r_{2f}$.

Case 3: Show that $r_f < r_{\Lambda(f)}$. Proof: Proceeds as a modified proof of case 1. For height zero, $\Lambda(f) = 2f + 1$ so $p_{\Lambda(f)} = \beta p_f - 1$. Thus, $1 = r_{\Lambda(f)} p_f(r_{\Lambda(f)})$ so $0 < p_f(r_{\Lambda(f)}) < 1$ and since $p'_n > 0$, conclude that $r_f < r_{\Lambda(f)}$. What about positive heights? For height h , the moves are $L^h \mathfrak{R}$ and so $p_{\Lambda(f)} = \beta^{h+1} p_f - 1$. Again, $1 = (r_{\Lambda(f)})^{h+1} p_f(r_{\Lambda(f)})$ and again $0 < p_f(r_{\Lambda(f)})$ and so again $r_f < r_{\Lambda(f)}$.

Case 4: Show that $r_{\Lambda(f)} < r_\rho$. Proof: This proceeds as a modified version of case 2, with the role of left and right reversed. In this case, ρ roots a tree, and the path to f is $f = R^n L \rho$ for some $n \geq 0$. For $n = 0$, one has $p_f = \beta(p_\rho + 1) - 1$ and so $p_{\Lambda(f)} = \beta^{h+1}(p_f) - 1$. Expanding, $p_{\Lambda(f)} = \beta^{h+2} p_\rho + \beta^{h+2}(\beta - 1) - 1$, and since $p_\rho(r_\rho) = 0$, one concludes that $p_{\Lambda(f)}(r_\rho) > 0$ and therefore $r_{\Lambda(f)} < r_\rho$. For $n > 0$, the proof proceeds the same way, with $p_{\Lambda(f)} = ap_\rho + b$ for some positive numbers a, b . Thus, $p_{\Lambda(f)}(r_\rho) = b > 0$ and again $r_{\Lambda(f)} < r_\rho$.

To conclude: the roots are ordered as $r_\ell < r_f < r_\rho$ with strict inequalities holding. QED.

6 Decaying Eigenfunctions

This section examines the general set of solutions to $\mathcal{L}_\beta \psi = \lambda \psi$ for $|\lambda| < 1$. Two general classes of solutions are presented. One consists of a discrete spectrum of (piecewise) polynomial eigenfunctions, occurring at the β values associated with the finite orbits. The other class consists of (piecewise) fractal eigenfunctions, which can be constructed for any complex λ with $|\lambda| < 1$. That is, the continuous spectrum.

The discrete spectrum is closely analogous to the discrete spectrum of the polynomial eigenfunctions of the Bernoulli operator. Recall that the Bernoulli operator is the beta transform, for $\beta = 2$. It has a discrete spectrum with $\lambda = 2^{-n}$ for all non-negative integers n ; the corresponding eigenfunctions are the Bernoulli polynomials. The Bernoulli polynomials are orthonormal to one-another. In the $n \rightarrow \infty$ limit, they converge to either $\sin 2\pi x$ or $\cos 2\pi x$, depending on whether n is even or odd. Both of these trig functions, and any function odd about $1/2$ lie in the kernel of the operator.

The Bernoulli operator also has a continuous spectrum that has a countable number of distinct eigenfunctions for each λ . [35] As these all share the same eigenvalue, one is free to choose the basis as desired. There are two generic possibilities. In one choice, the basis functions are square integrable, but are not differentiable at the dyadic rationals. These are built up as coherent states of underlying periodic waves. They resemble the blancmange curve; the periodic spectrum appears here, in the same way that the blancmange curve becomes a parabola for a specific parameter value.

A different choice of basis provides eigenfunctions that are smooth over the open interval $0 < x < 1$ but diverge at $x = 0, 1$ and are thus not square-integrable, due to these divergent endpoints. Equivalently, they are not smooth at $x = 0, 1$. The discrete spectrum emerges as those special values of λ for which the endpoints are finite, thus both smooth and square-integrable. Explicit constructions of both sets of basis functions, and the transformation between them, is given in [35].

This section presents explicit analogous solutions for the beta operator, for the β values corresponding to finite orbits. That is, both kinds of eigenfunctions are presented: piecewise-polynomial eigenfunctions, and fractal (coherent state) eigenfunctions. The development below is long and verbose, but never rises above the complexity level of basic algebra. There does not seem to be much need for abstract conceptualization here; it is mostly a matter of turning the crank and verifying identities.

6.1 General Framework

When an orbit is finite, of order v , the Parry–Gelfond invariant measure consists of v distinct plateaus. It is a piecewise constant function. This description can be extended to any piecewise continuous sections; the method is illustrated for piecewise polynomial intervals. For a polynomial of order j , the largest real eigenvalue is β^{-j} , with additional complex eigenvalues λ bounded in the range $\beta^{-j-1} < |\lambda| < \beta^{-j}$.

The method for obtaining these is straightforward. The first step is to obtain the segments on which the polynomials will live. This is easy: the endpoints of the intervals are determined by the midpoint orbit $m_k = T_\beta^{k+1}(1)$. A somewhat inelegant/difficult bit is that these need to be sorted into order, so that they define non-overlapping segments. Hypothesizing a distinct polynomial on each segment, these are plugged through \mathcal{L}_β to

obtain a matrix relating polynomial coefficients. This matrix is block-upper-triangular, and thus solvable, and that's it; one is done.

Besides the need to sort midpoints, there is one more inelegant or difficult aspect to this. The application of \mathcal{L}_β generates additional intervals; these are images of the original intervals under the pushforward \mathcal{L}_β . They are determined by the values m_k/β and $1/2 + m_k/\beta$ for each midpoint. These new intervals are entirely contained within the original set, but introduce a practical difficulty: the membership function is needed. Membership is easily computed with arithmetic; what is annoying is that there does not seem to be any obvious recursive relationship that provides this membership function. Instead, one proceeds by brute force: compute the m_k , sort them into order, compute the m_k/β and $1/2 + m_k/\beta$, and see where they land. It would be conceptually preferable if some relationship could be found, that gave the indicator function by means of some small, simple recurrence relation. If there is such a relation, it is not obvious.

6.2 Ruminations

The development of this section raises an overall bugaboo: what is the appropriate balance between “mathematics” and “algorithmics”? The general goal of the former is to provide understanding and insight by use of small, compact algebraic formulas establishing relationships between things. The latter is used to compute things that are too tedious to do by hand. But a thing is computable, only if there is a recursive formulation that terminates. The product of two real numbers is computable only if they themselves are individually computable. The “best” algorithm for multiplication requires a trade-off between speed, size and complexity; a choice must be made. There is a homotopic landscape of algorithms, but it is hardly clear what is homotopic to what.

The algorithmic difficulty of multiplying a pair of arbitrary numbers is avoided when those two numbers can be given a short, compact mathematical (arithmetical?) definition: as members of some ring or integral domain or whatever. Computing the m_k is far more annoying: the appropriate value of β must first be found by solving for the root of a polynomial $p_n(\beta) = 0$, when the orbit is finite, or by specifying β in some other way, when the orbit is not finite. But still, this can be brushed away as un concerning, since multiplication is an old and comfortable prosthesis, as is root-finding.

Thus, in one sense, it is quite fine to say “compute the m_k , sort them into order, compute the m_k/β and $1/2 + m_k/\beta$, and see where they land”; this is just fine as an algorithm. As a bit of mathematics, it is deeply unsatisfying: the relationship between the endpoints is opaque and unknown; the endpoints themselves can be explicitly graphed and charted and exhibited, showing hints of patterns present. But the form of those patterns is hidden. The mathematical structure is unresolved.

Is this acceptable? There is no algebraic equation for the real roots of a polynomial larger than a cubic; we've come to terms with this state of affairs and accept it. But this is only because there are existing, known proofs that explicitly demonstrate that such roots are fundamentally unobtainable by algebraic means. The state of affairs regarding the iterated β -transform midpoints is not settled: perhaps there are algebraic

expressions, and perhaps there are not. A tool-set to determine the answer to this question is absent. I don't know how to solve this, other than to blunder about.

6.3 Motivating example

The general solution can be placed into context by exhibiting an example of what it produces. This section demonstrates a piecewise quadratic solution to

$$\mathcal{L}_\beta f = \lambda f$$

The solution is developed for the $n = 2$ finite orbit.

For $n = 2$, the defining polynomial is the cubic $p_2(\beta) = \beta^3 - \beta^2 - 1$. It has a root $p_2(r_2) = 0$, numerically determined as $\beta = r_2 \approx 1.46557 \dots$. The midpoint orbit is

$$\begin{aligned} m_0 &= \beta/2 \approx 0.73278 \\ m_1 &= \beta(\beta - 1)/2 \approx 0.34116 \\ m_2 &= \beta^2(\beta - 1)/2 = 1/2 \end{aligned}$$

The length of the orbit is $\nu = 3$, which is necessarily the same as the degree of the cubic p_2 . The midpoints fall in the sequence $0 < m_1 < m_2 < m_0 < 1$, and so this fixes four intervals

$$I_3 = [0, m_1] < I_2 = [m_1, m_2] < I_1 = [m_2, m_0] < I_0 = [m_0, 1]$$

As the desired solution is a quadratic, hypothesize $f_k(y) = a_{k0} + a_{k1}y + a_{k2}y^2$ on each segment I_k with constants a_{kj} to be determined.

The pushforward is, as always

$$[\mathcal{L}_\beta f](y) = \frac{1}{\beta} \left[f\left(\frac{y}{\beta}\right) \Theta(m_0 - y) + f\left(\frac{y}{\beta} + \frac{1}{2}\right) \Theta(m_1 - y) \right]$$

and plugging through the f_k results in eigen-relations

$$\lambda f_k(y) = \frac{1}{\beta} \left[f_{Lk}\left(\frac{y}{\beta}\right) + f_{Rk}\left(\frac{y}{\beta} + \frac{1}{2}\right) \right]$$

The notation f_{Lk} and f_{Rk} is shorthand for the recursive interval maps $L : I_k \mapsto \frac{I_k}{\beta}$ and $R : I_k \mapsto \frac{I_k}{\beta} + \frac{1}{2}$. These two maps induce interval inclusion relations

$$\begin{aligned} LI_3 &\subset I_3 & \text{and} & & LI_2 &\subset I_3 & \text{and} & & LI_1 &= I_2 \\ RI_3 &= I_1 & \text{and} & & RI_2 &\subset I_0 & \text{and} & & RI_1 &\subset I_0 \end{aligned}$$

The shorthand notation is just a linear operator:

$$f_{Lk} = \sum_{m=0}^{\nu} \Theta(LI_k \subseteq I_m) f_m$$

where $\Theta(LI_k \subseteq I_m)$ is an indicator function that is true for only one k, m pair. It is generically the case that LI_k is wholly contained in only one interval I_m . Likewise for RI_k , and so the eigen-relations become

$$\lambda f_k(y) = \frac{1}{\beta} \sum_{m=1}^v \left[\Theta(LI_k \subseteq I_m) f_m\left(\frac{y}{\beta}\right) + \Theta(RI_k \subseteq I_m) f_m\left(\frac{y}{\beta} + \frac{1}{2}\right) \right]$$

That this is just a matrix equation can be manifested more directly by writing it as

$$\lambda \begin{bmatrix} f_1(y) \\ f_2(y) \\ f_3(y) \end{bmatrix} = \frac{1}{\beta} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1\left(\frac{y}{\beta}\right) \\ f_2\left(\frac{y}{\beta}\right) \\ f_3\left(\frac{y}{\beta}\right) \end{bmatrix} + \frac{1}{\beta} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_1\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ f_2\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ f_3\left(\frac{y}{\beta} + \frac{1}{2}\right) \end{bmatrix}$$

To make an overly-long story short, the hypothesized quadratics $f_k(y) = a_{k0} + a_{k1}y + a_{k2}y^2$ can be plugged through; the monomials in y^j gathered together, leaving a solvable matrix equation. The general form will be given shortly. The resulting solution is

$$f(y) = \begin{cases} f_3(y) = \frac{\beta-1}{\beta}y^2 - \frac{\beta-1}{\beta}y + \frac{\beta(\beta-1)}{4(\beta+1)} & \text{for } 0 \leq y < m_1 = \frac{\beta(\beta-1)}{2} \\ f_2(y) = (\beta-1)y^2 - \frac{\beta-1}{\beta}y + \frac{\beta-1}{4(\beta+1)} & \text{for } m_1 \leq y < m_2 = \frac{1}{2} \\ f_1(y) = y^2 - \frac{\beta-1}{\beta}y + \frac{\beta-1}{4\beta(\beta+1)} & \text{for } \frac{1}{2} \leq y < m_0 = \frac{\beta}{2} \\ 0 & \text{for } m_0 \leq y \leq 1 \end{cases}$$

During the solution, it becomes apparent that the eigenvalue is $\lambda = 1/\beta^2$. The solution consists of four intervals, three piece-wise quadratics. Keep in mind this holds only for $\beta = r_2 \approx 1.46557 \dots$ and so other finite orbits will have a different form.

6.4 General form

The general form follows the template given in the example above. It provides additional detail on how to obtain the solvable matrix. From the top:

- The goal is to solve $\mathcal{L}_\beta f = \lambda f$ to obtain the eigenvalue λ and eigenfunction f . This will be done for β values that have finite orbits. These values are given by the largest real root of the corresponding polynomial.
- Fix n , selecting a polynomial $p_n(\beta)$ of order v . Solve $p_n(\beta) = 0$ to obtain a root $r_n = \beta$.
- Iterate the midpoint $m_0 = \beta/2$ to get $m_1 = T_\beta(m_0) = \beta(\beta-1)/2, \dots, m_k = T_\beta^k(m_0), \dots, m_v = 1/2$.
- Use the midpoints to split the interval $[0, 1]$ into v distinct sub-intervals. Provide a label k for each interval, so that the intervals I_k can be referred to by their labels. It is generally easiest to sort these into ascending order, but this is not mandatory.

- Define the recursive interval maps $L : I_k \mapsto \frac{I_k}{\beta}$ and $R : I_k \mapsto \frac{I_k}{\beta} + \frac{1}{2}$.
- Starting with $I_0 = [m_0, 1]$, define an interval inclusion graph.
- The interval inclusion graph determines incidence matrices. These are square matrices B, C having matrix entries $B_{kj} = \Theta(I_j \subseteq LI_k)$ where

$$\Theta(I_j \subseteq LI_k) = \begin{cases} 1 & \text{if } I_j \subseteq LI_k \\ 0 & \text{if } I_j \not\subseteq LI_k \end{cases}$$

is the inclusion membership function. Likewise $C_{kj} = \Theta(I_j \subseteq RI_k)$.

- For each interval k , define the component $f_k(y) = f(y)|_{I_k}$ as the restriction of f to the interval I_k .
- On each interval, the pushforward \mathcal{L}_β takes the form of a recurrence relation

$$\lambda \beta f_k(y) = f_{Lk} \left(\frac{y}{\beta} \right) + f_{Rk} \left(\frac{y}{\beta} + \frac{1}{2} \right)$$

where $Lk = j$ when $I_j \subseteq LI_k$ and likewise for Rk .

- Explicitly inserting the membership function Θ into the recurrence relation gives the linear equation

$$\lambda f_k(y) = \frac{1}{\beta} \sum_{m=1}^v \left[\Theta(LI_k \subseteq I_m) f_m \left(\frac{y}{\beta} \right) + \Theta(RI_k \subseteq I_m) f_m \left(\frac{y}{\beta} + \frac{1}{2} \right) \right]$$

- This is more visually transparent when written in block form a

$$\lambda \begin{bmatrix} f_1(y) \\ \vdots \\ f_v(y) \end{bmatrix} = \frac{1}{\beta} B \begin{bmatrix} f_1 \left(\frac{y}{\beta} \right) \\ \vdots \\ f_v \left(\frac{y}{\beta} \right) \end{bmatrix} + \frac{1}{\beta} C \begin{bmatrix} f_1 \left(\frac{y}{\beta} + \frac{1}{2} \right) \\ \vdots \\ f_v \left(\frac{y}{\beta} + \frac{1}{2} \right) \end{bmatrix}$$

with the square matrices B, C as just defined.

- For each interval k , write $f_k(y) = \sum_j a_{kj} y^j$ for unknown constants a_{kj} to be solved for. Thus, the eigenfunction will be piece-wise polynomial. The above forms a linear system that is solvable.
- Insert the polynomial pieces, and gather terms into monomials y^j . This requires the use of binomial coefficients to write

$$\left(\frac{y}{\beta} + \frac{1}{2} \right)^j = \sum_{m=0}^j \binom{j}{m} y^m \frac{1}{\beta^m} \cdot \frac{1}{2^{j-m}}$$

so that

$$\begin{aligned} f_k \left(\frac{y}{\beta} + \frac{1}{2} \right) &= \sum_j a_{kj} \left(\frac{1}{2} + \frac{y}{\beta} \right)^j \\ &= \sum_{m=0}^{\infty} y^m \left(\frac{2}{\beta} \right)^m \sum_{j=m}^{\infty} a_{kj} \frac{1}{2^j} \binom{j}{m} \end{aligned}$$

- Grouping into monomials results in the block-triangular form

$$\lambda \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ \vdots \end{bmatrix} = \frac{1}{\beta} \begin{bmatrix} D & C_{01} & C_{02} & \cdots \\ 0 & \frac{1}{\beta} D & C_{12} & \cdots \\ 0 & 0 & \frac{1}{\beta^2} D & \\ \vdots & \vdots & & \ddots \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ \vdots \end{bmatrix}$$

with $D = B + C$ and with $C_{jm} = c_{jm} C$ with

$$c_{jm} = \left(\frac{2}{\beta} \right)^m \frac{1}{2^j} \binom{j}{m}$$

and block vectors

$$A_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{vj} \end{bmatrix}$$

- In other words,

$$\mathcal{L}_\beta = \frac{1}{\beta} \begin{bmatrix} D & C_{01} & C_{02} & \cdots \\ 0 & \frac{1}{\beta} D & C_{12} & \cdots \\ 0 & 0 & \frac{1}{\beta^2} D & \\ \vdots & \vdots & & \ddots \end{bmatrix}$$

is a matrix representation of \mathcal{L}_β in a monomial-basis form. It is upper-triangular. It is solvable. It can be truncated at any finite j to give a finite $v j \times v j$ matrix.

- For any fixed v , the eigenfunction consists of v pieces, each piece being a polynomial of the same degree j .
- For such a truncated, finite-size matrix, the spectrum of \mathcal{L}_β is determined entirely by the spectrum of the block D , as solvability implies that only the bottom-right block matters. Thus, one concludes that the spectrum of \mathcal{L}_β is just the spectrum of D times β^{-j} .
- The spectrum of D is given by the self-encoding polynomial; it is $\det[D - \omega I] = p_n(\omega)$. The roots of $p_n(\omega)$ completely determine the constructable eigenfunctions.

- Given this block-triangular structure, the consequence is that the largest real eigenvalue of \mathcal{L}_β will be β^{-j} and there will in general be additional complex eigenvalues λ satisfying $\beta^{-j-1} < |\lambda| < \beta^{-j}$. These are determined entirely by the spectrum of D .

This is a complete description for such piece-wise polynomial solutions. The solution decouples the behavior with respect to the monomials, which is captured in the D, C block matrix, from the behavior with respect to the intervals, which is captured in the individual blocks B, C . This allows limiting behavior can be explored in two distinct ways. One is the replacement of the monomial basis by some other basis, possibly not polynomial. Another is the $\nu \rightarrow \infty$ limit, where the number of plateaus becomes unbounded.

The factorization into block-triangular form indicates that the spectrum of \mathcal{L}_β is determined entirely by the spectrum of D , and thus it is this spectrum that should be the subject of focus; the rest is just decorative trimmings. The only irregular or “difficult” part to the above is the calculation of the interval inclusions, needed to obtain the left and right incidence matrices B and C .

That’s it. That’s the complete solution for the general case.

6.5 Interval Inclusion Tables

The most confusing aspect of the above solution is the form of the interval inclusions. This is explored below, by writing out explicit cases and organizing them into a table. Unfortunately, they shine no light on the situation.

Below is the interval inclusion table for the polynomials up to order $\nu = 5$.

ν	n	m_j	LI_5	LI_4	LI_3	LI_2	LI_1	RI_5	RI_4	RI_3	RI_2	RI_1
2	1	1,0				$\subset I_2$	$\subset I_2$				$= I_1$	$= I_0$
3	2	1,2,0			$\subset I_3$	$\subset I_3$	$= I_2$			$= I_1$	$\subset I_0$	$\subset I_0$
	3	2,1,0			$\subset I_3$	$\subset I_3$	$\subset I_3$			$= I_2$	$= I_1$	$= I_0$
4	4	1,2,3,0		$\subset I_4$	$\subset I_4$	$= I_3$	$= I_2$		$= I_1$	$\subset I_0$	$\subset I_0$	$\subset I_0$
	6	2,3,1,0		$\subset I_4$	$\subset I_4$	$\subset I_3$	$\subset I_3$		$= I_2$	$\subset I_1$	$\subset I_1$	$= I_0$
	7	3,2,1,0		$\subset I_4$	$\subset I_4$	$\subset I_4$	$\subset I_4$		$= I_3$	$= I_2$	$= I_1$	$= I_0$
5	8	1,2,3,4,0	$\subset I_5$	$\subset I_5$	$= I_4$	$= I_3$	$= I_2$	$= I_1$	$\subset I_0$	$\subset I_0$	$\subset I_0$	$\subset I_0$
	10	3,1,4,2,0	$\subset I_5$	$\subset I_5$	$\subset I_5$	$= I_4$	$= I_3$	$= I_2$	$= I_1$	$\subset I_0$	$\subset I_0$	$\subset I_0$
	12	2,3,4,1,0	$\subset I_5$	$\subset I_5$	$= I_4$	$\subset I_3$	$\subset I_3$	$= I_2$	$\subset I_1$	$\subset I_1$	$\subset I_1$	$= I_0$
	13	2,4,1,3,0	$\subset I_5$	$\subset I_5$	$\subset I_5$	$\subset I_5$	$= I_4$	$= I_3$	$= I_2$	$= I_1$	$\subset I_0$	$\subset I_0$
	14	3,4,2,1,0	$\subset I_5$	$\subset I_5$	$\subset I_4$	$\subset I_4$	$\subset I_4$	$= I_3$	$\subset I_2$	$\subset I_2$	$= I_1$	$= I_0$
15	4,3,2,1,0	$\subset I_5$	$\subset I_5$	$\subset I_5$	$\subset I_5$	$\subset I_5$	$= I_4$	$= I_3$	$= I_2$	$= I_1$	$= I_0$	

Table legend:

- The column labeled m_j gives the index order for j such that $0 < \dots < m_j < \dots < m_0 = \beta/2 < 1$.
- The columns labeled LI_k and RI_k indicate what happens when L, R are applied to I_k . The result is marked as equality or strict inclusion.

I don't have anything particularly notable to say about the above table. Although there is a pattern to the above table, it is difficult to express in words. More progress can be made by encoding the binary relationships as matrices, as below.

6.5.1 Interval inclusion incidence matrices

An alternate representation for the above tables is as matrices. These are the square matrices B, C , having matrix entries $B_{kj} = \Theta(I_j \subseteq LI_k)$ where

$$\Theta(I_j \subseteq LI_k) = \begin{cases} 1 & \text{if } I_j \subseteq LI_k \\ 0 & \text{if } I_j \not\subseteq LI_k \end{cases}$$

is the inclusion membership function. Likewise $C_{kj} = \Theta(I_j \subseteq RI_k)$. In tabular form:

v	n	m_j	B	C	seq	ind	P	D
2	1	1,0	$\begin{bmatrix} \cdot 1 \\ \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \\ 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} L \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \\ 1 \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \\ 11 \end{bmatrix}$
3	2	1,2,0	$\begin{bmatrix} \cdot 1 \cdot \\ \cdot \cdot 1 \\ \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \\ \cdot \cdot \cdot \\ 1 \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$	$\begin{bmatrix} L \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \\ \cdot \cdot 1 \\ 1 \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \\ \cdot \cdot 1 \\ 1 \cdot 1 \end{bmatrix}$
	3	2,1,0	$\begin{bmatrix} \cdot \cdot 1 \\ \cdot \cdot 1 \\ \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \\ 1 \cdot \cdot \\ \cdot 1 \cdot \end{bmatrix}$	$\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \\ 1 \cdot \cdot \\ \cdot 1 \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \\ 1 \cdot 1 \\ \cdot 11 \end{bmatrix}$
4	4	1,2,3,0	$\begin{bmatrix} \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix}$	$\begin{bmatrix} L \\ L \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot 1 \\ 1 \cdot \cdot 1 \end{bmatrix}$
	6	2,3,1,0	$\begin{bmatrix} \cdot \cdot 1 \cdot \\ \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \\ \cdot 1 \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} 3 \\ 1 \\ 4 \\ 2 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \cdot \\ 1 \cdot \cdot \cdot \\ \cdot \cdot \cdot 1 \\ \cdot 1 \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \cdot \\ 1 \cdot 1 \cdot \\ 1 \cdot \cdot 1 \\ \cdot 1 \cdot 1 \end{bmatrix}$
	7	3,2,1,0	$\begin{bmatrix} \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \\ \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \end{bmatrix}$	$\begin{bmatrix} 4 \\ 1 \\ 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ R \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot \\ \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot 1 \\ 1 \cdot \cdot 1 \\ \cdot 1 \cdot 1 \\ \cdot \cdot 11 \end{bmatrix}$
5	8	1,2,3,4,0	$\begin{bmatrix} \cdot 1 \cdot \cdot \cdot \\ \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 1 \end{bmatrix}$	$\begin{bmatrix} L \\ L \\ L \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \cdot \cdot \\ \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot 1 \cdot \cdot \cdot \\ \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot 1 \end{bmatrix}$
	10	3,1,4,2,0	$\begin{bmatrix} \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot \cdot 1 \\ \cdot \cdot \cdot \cdot 1 \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \\ \cdot \cdot \cdot \cdot \cdot \\ 1 \cdot \cdot \cdot \cdot \\ \cdot 1 \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} 3 \\ 4 \\ 5 \\ 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} L \\ L \\ L \\ R \\ R \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot \cdot \\ \cdot 1 \cdot \cdot \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot 1 \\ \cdot 1 \cdot \cdot 1 \end{bmatrix}$

v	n	m_j	B	C	seq	ind	P	D
	12	2,3,4,1,0	$\begin{bmatrix} \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} \dots \dots \\ 1 \dots \dots \\ 1 \dots \dots \\ 1 \dots \dots \\ \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} 3 \\ 1 \\ 4 \\ 5 \\ 2 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ L \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \dots 1 \dots \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots \dots 1 \\ \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} \dots 1 \dots \\ 1 \dots 1 \dots \\ 1 \dots 1 \dots \\ 1 \dots 1 \dots \\ \dots 1 \dots 1 \end{bmatrix}$
	13	2,4,1,3,0	$\begin{bmatrix} \dots 1 \dots \\ \dots \dots 1 \\ \dots \dots 1 \\ \dots \dots 1 \\ \dots \dots 1 \end{bmatrix}$	$\begin{bmatrix} \dots \dots \dots \\ \dots \dots \dots \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} 4 \\ 5 \\ 1 \\ 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} L \\ L \\ R \\ R \\ R \end{bmatrix}$	$\begin{bmatrix} \dots \dots 1 \dots \\ \dots \dots \dots 1 \\ 1 \dots \dots \dots \\ \dots 1 \dots \dots \\ \dots \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} \dots \dots 1 \dots \\ \dots \dots \dots 1 \\ 1 \dots \dots \dots \\ \dots 1 \dots \dots \\ \dots \dots 1 \dots \end{bmatrix}$
	14	3,4,2,1,0	$\begin{bmatrix} \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \\ \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} \dots \dots \dots \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots 1 \dots \\ \dots \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} 4 \\ 1 \\ 2 \\ 5 \\ 3 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ R \\ L \\ R \end{bmatrix}$	$\begin{bmatrix} \dots 1 \dots \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots \dots 1 \\ \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} \dots 1 \dots \\ 1 \dots 1 \dots \\ \dots 1 \dots 1 \\ \dots 1 \dots 1 \\ \dots 1 \dots 1 \end{bmatrix}$
	15	4,3,2,1,0	$\begin{bmatrix} \dots \dots 1 \\ \dots \dots 1 \\ \dots \dots 1 \\ \dots \dots 1 \\ \dots \dots 1 \end{bmatrix}$	$\begin{bmatrix} \dots \dots \dots \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots \dots 1 \dots \\ \dots \dots 1 \dots \end{bmatrix}$	$\begin{bmatrix} 5 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$	$\begin{bmatrix} L \\ R \\ R \\ R \\ R \end{bmatrix}$	$\begin{bmatrix} \dots \dots 1 \\ 1 \dots \dots \\ \dots 1 \dots \\ \dots \dots 1 \\ \dots \dots 1 \end{bmatrix}$	$\begin{bmatrix} \dots \dots 1 \\ 1 \dots \dots 1 \\ \dots 1 \dots 1 \\ \dots \dots 1 \dots 1 \\ \dots \dots 1 \dots 1 \end{bmatrix}$

The structure of the B and C matrices have no strong, easy-to-describe pattern that jumps out. It is all more subtle. First, some general comments. These all follow from the nature of the interval inclusions, and are fairly easily derived.

- The sum $D = B + C$ never has any entries that have 2 in it. This follows from the L, R maps.
- The first column of B is always zero. This follows because L always shifts down; the topmost interval can never be a subinterval of some other interval.
- The first column of C always has at least one entry that is one. This follows because R always shifts up, and some interval must always be above the one or more lower intervals.
- There are one or more non-zero entries in the last column of B .
- The last column of C is always all-zeros.
- Every row in B has exactly one non-zero entry. This follows, because every interval is a subinterval of one and only one other interval.
- The top row of C is always all-zeros.
- It is always the case that $B(v, v) = 1$ and $B(v - 1, v) = 1$.

A more opaque but important pattern is the shift structure across the two. For every row of B and C , the following holds: If $B(k, j) = 1$, then one and only one of the following possibilities hold:

- $B(k + 1, j + 1) = 1$ and $C(k + 1, 1) = 0$

- $B(k+1, j+1) = 1$ and there does not exist any i such that $C(k, i) = 1$ and $C(k+1, i+1) = 1$
- $B(k+1, j+1) = 0$ and $C(k+1, 1) = 1$
- $B(k+1, j+1) = 0$ and there exists i such that $C(k, i) = 1$ and $C(k+1, i+1) = 1$
- $j = v$ and $C(k+1, 1) = 1$
- $j = v$ and there exists i such that $C(k, i) = 1$ and $C(k+1, i+1) = 1$

Since for every row k , there is always some column j for for which $B(k, j) = 1$, the above holds for all rows. It is saying that there is always a shift, either in the B matrix, or in the C matrix, and never both at the same time, for every row.

The six bullet points above are expressed as logic statements written in English. To be useful for calculations, they need to be converted into formulas. This will be needed later, in the section on the continuous spectrum. The structure above can be captured by two distinct functions. One is a sequencing function: given interval k (row k), return the next shift $j+1$; the other function is the sequence indicator function: given interval k , was the shift obtained on the left or right? The table above shows these two functions, labeled as “seq” and “ind”.

The “seq” function is a permutation; the corresponding permutation matrix is shown in the P column. The “sporadic” matrix $S = B + C - P$ is what is left over, after removing P .

The indicator function “ind” shows how the permutation matrix is assembled from B and C . One has that

$$P(k, j) = \Theta(\text{ind}(k) = L)B(k, j) + \Theta(\text{ind}(k) = R)C(k, j)$$

$$P_{kj} = \delta_{\text{ind}(k), L}B_{kj} + \delta_{\text{ind}(k), R}C_{kj}$$

where $\Theta(x) = 1$ if x evaluates to true, else $\Theta(x) = 0$. The δ_{ab} is the Kronecker delta.

The sequence function “seq” is the permutation, expressed as a cycle:

$$P_{kj} = \delta_{\text{seq}(k), j} = \Theta(\text{seq}(k) = j)$$

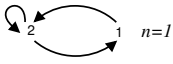
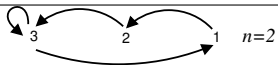
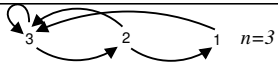
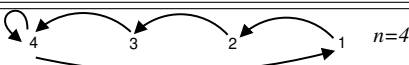
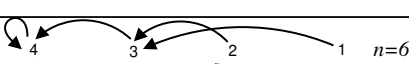
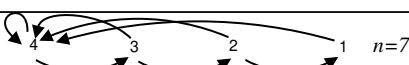
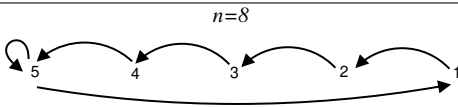
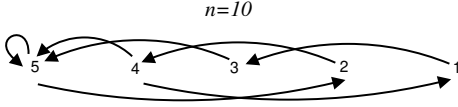
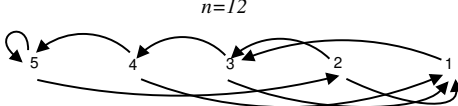
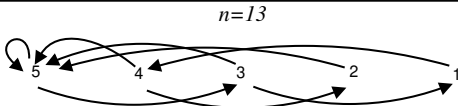
It is currently unclear how to obtain any of these matrices and functions, other than to brute-force iterate the midpoint, sort the midpoints into order, compute the intervals, and compute the overlaps. This is an algorithmic process that does not provide much insight.

The permutation matrices are orthogonal, in that $P^{-1} = P^T$. There are some additional sporadic identities. For $P = P_n$ for index n , note that $P_2^T = P_3$ and $P_7^T = P_4$ and $P_{15}^T = P_8$ and $P_{13}^T = P_{10}$. Unclear how these interchanges might be meaningful; there’s no corresponding relationship between P_{12} and P_{14} .

The product $P^T D$ is lower-triangular, and thus explicitly solvable.

The D matrix encodes the generating polynomial. That is, $\det[D - \lambda I] = p_n(\lambda)$ so that $\det[D - \beta I] = 0$.

Each matrix D can be taken to be the adjacency matrix for a graph. These are shown below; they show how intervals mix into one-another.

v	n	D	diagram
2	1	$\begin{bmatrix} \cdot 1 \\ 11 \end{bmatrix}$	
3	2	$\begin{bmatrix} \cdot 1 \cdot \\ \cdot \cdot 1 \\ 1 \cdot 1 \end{bmatrix}$	
	3	$\begin{bmatrix} \cdot \cdot 1 \\ 1 \cdot 1 \\ \cdot 11 \end{bmatrix}$	
4	4	$\begin{bmatrix} \cdot 1 \cdot \cdot \\ \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot 1 \\ 1 \cdot \cdot 1 \end{bmatrix}$	
	6	$\begin{bmatrix} \cdot \cdot 1 \cdot \\ 1 \cdot 1 \cdot \\ 1 \cdot \cdot 1 \\ \cdot 1 \cdot 1 \end{bmatrix}$	
	7	$\begin{bmatrix} \cdot \cdot \cdot 1 \\ 1 \cdot \cdot 1 \\ \cdot 1 \cdot 1 \\ \cdot \cdot 11 \end{bmatrix}$	
5	8	$\begin{bmatrix} \cdot 1 \cdot \cdot \cdot \\ \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot 1 \end{bmatrix}$	
	10	$\begin{bmatrix} \cdot \cdot 1 \cdot \cdot \\ \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot 1 \\ \cdot 1 \cdot \cdot 1 \end{bmatrix}$	
	12	$\begin{bmatrix} \cdot \cdot 1 \cdot \cdot \\ 1 \cdot 1 \cdot \cdot \\ 1 \cdot \cdot 1 \cdot \\ 1 \cdot \cdot \cdot 1 \\ \cdot 1 \cdot \cdot 1 \end{bmatrix}$	
	13	$\begin{bmatrix} \cdot \cdot \cdot 1 \cdot \\ \cdot \cdot \cdot \cdot 1 \\ 1 \cdot \cdot \cdot 1 \\ \cdot 1 \cdot \cdot 1 \\ \cdot \cdot 1 \cdot 1 \end{bmatrix}$	

v	n	D	diagram
	14	$\begin{bmatrix} \dots & 1 & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & 1 \end{bmatrix}$	<p style="text-align: center;">$n=14$</p>
	15	$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & 1 \end{bmatrix}$	<p style="text-align: center;">$n=15$</p>

Left-moving arrows indicate $LI_k \mapsto I_j = I_k/\beta$ while right-moving arrows indicate $RI_k \mapsto I_j = I_k/\beta + 1/2$. If only one arrow lands on an interval, then the map is an equality (one-to-one and onto); otherwise it is an inclusion. The arrow-heads are drawn in proper interval sequence; the L and R maps never alter the overall sequence. Due to the self-inclusion $LI_v \mapsto I_v = I_v/\beta$ for the left-most interval, for all orders v , this map is always an inclusion (recursively splitting the interval into ever-smaller pieces). These in turn are mapped back to the right, eventually creating a set of endpoint mappings that is dense in the unit interval. Note also that there is always a map $LI_{v-1} \mapsto I_v$ for all orders v .

Some open questions:

- Why these graphs, and not others?
- Are these somehow the prime graphs, in that all other (measure-preserving) graphs can be constructed from these? For example, all Markovian two-point graphs can be taken as linear combinations of the $n = 0$ and $n = 1$ graphs. In this example, $M_{ij} = a\delta_{i1}\delta_{j1} + b\delta_{i2}\delta_{j2} + c(\delta_{i1}\delta_{j1} + \delta_{i2}\delta_{j1} + \delta_{i1}\delta_{j2})$ represents all possible 2×2 graphs, where as much material flows into a vertex, as flows out of it. The a and b terms are just the $n = 0$ graph, while the c term is the $n = 1$ graph. (A Markov normalization additionally requires that $a + c = b + c = 1$.)

6.6 Coherent States

It is possible to construct coherent-state eigenfunctions satisfying $\mathcal{L}_\beta f = wf$ for any complex w with $|w| < 1$. By “coherent state” it is meant a fractal function

$$\psi_w(x) = \sum_{k=0}^{\infty} w^k s\left(T_\beta^k(x)\right)$$

where $s(x) = g(x \bmod \beta/2)$ is a sawtooth, and $g(x)$ is more-or-less any function on the interval $0 \leq x \leq 1/2$; it is constrained by recursion on $1/2 \leq x \leq \beta/2$. This construction uses the framework established above. It is relatively straightforward, although it has a few difficult turns that leave open it’s general structure.

The construction appears to yield a family of functions f that should be able to form a countable orthonormal basis, for any fixed w . This has some interesting ramifications. In the case of $\beta = 2$, the Bernoulli shift, this basis has been obtained long ago.[35] Since the collection of basis functions is countable, the space is quite large. It includes smooth functions, infinitely differentiable on the open interval $0 < x < 1$, but divergent at the endpoints. These also form a countable basis, which can be explicitly exhibited. Furthermore the (orthogonal) transform, taking the one basis into the other, can also be exhibited. Explicit expressions are given in [35]. No such explicit form is given below. It is hypothesized to exist, by analogy. However, the manipulations here are sufficiently daunting that an explicit demonstration remains an open problem. There are more than a few open problems that arise, here.

6.6.1 Construction

Fixing a self-describing polynomial $p_n(\beta) = 0$ for fixed n and order v , the construction follows the framework developed above. The solution f to $\mathcal{L}_\beta f = \lambda f$ consists of piece-wise components $f(y) = f_k(y)$, when y is restricted to the interval I_k . The interval inclusion maps give the twizzled form

$$\beta\lambda \begin{bmatrix} f_1(y) \\ \vdots \\ f_v(y) \end{bmatrix} = B \begin{bmatrix} f_1\left(\frac{y}{\beta}\right) \\ \vdots \\ f_v\left(\frac{y}{\beta}\right) \end{bmatrix} + C \begin{bmatrix} f_1\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ \vdots \\ f_v\left(\frac{y}{\beta} + \frac{1}{2}\right) \end{bmatrix}$$

Each segment is a superposition of coherent states, expressed with a rotation matrix R such that

$$\begin{bmatrix} f_1(y) \\ \vdots \\ f_v(y) \end{bmatrix} = R \begin{bmatrix} \psi_1(y) \\ \vdots \\ \psi_v(y) \end{bmatrix}$$

In index notation, this is $f_k(y) = R_{kj} a_j \psi_j(y)$ with $\psi_k(y) = \psi\left(T_\beta^{k-1}(y)\right)$. For the case of $n = 1$ and $n = 2$, it is possible to obtain a solution with R being the identity. For the case $n = 3$, this is no longer possible, and R is non-trivial, and depends on the eigenvalue. It is hypothesized that one can take $R = I$ whenever $\beta \lesssim 1.5$, while larger β values require a non-trivial R , and may also incur other constraints. The need for R arises from the form of the C matrix: if it has more than one non-zero entry, then it applies a constraint that must be resolved with a non-trivial R .

Several identities are needed. One is

$$\begin{aligned} \psi\left(T_\beta^k(y)\right) &= \sum_{n=0}^{\infty} w^n s\left(T_\beta^{n+k}(y)\right) \\ &= \frac{1}{w^k} \left[\psi(y) - \sum_{n=0}^{k-1} w^n s\left(T_\beta^n(y)\right) \right] \end{aligned}$$

Also, for $k > 0$,

$$\psi\left(T_\beta^k\left(\frac{y}{\beta}\right)\right) = \psi\left(T_\beta^k\left(\frac{y}{\beta} + \frac{1}{2}\right)\right) = \psi\left(T_\beta^{k-1}(y)\right)$$

while for $k = 0$

$$\begin{aligned}\psi\left(\frac{y}{\beta}\right) &= s\left(\frac{y}{\beta}\right) + w\psi(y) \\ \psi\left(\frac{y}{\beta} + \frac{1}{2}\right) &= s\left(\frac{y}{\beta} + \frac{1}{2}\right) + w\psi(y)\end{aligned}$$

Plugging through, this gives

$$\beta\lambda R \begin{bmatrix} a_1\psi_1(y) \\ \vdots \\ a_v\psi_v(y) \end{bmatrix} = BR \begin{bmatrix} a_1\psi_1\left(\frac{y}{\beta}\right) \\ a_2\psi_1(y) \\ \vdots \\ a_v\psi_{v-1}(y) \end{bmatrix} + CR \begin{bmatrix} a_1\psi_1\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ a_2\psi_1(y) \\ \vdots \\ a_v\psi_{v-1}(y) \end{bmatrix}$$

Setting $D = B + C$ (as before) and continuing,

$$\beta\lambda R \begin{bmatrix} a_1\psi_1(y) \\ \vdots \\ a_v\psi_v(y) \end{bmatrix} = DR \begin{bmatrix} a_1w\psi_1(y) \\ a_2\psi_1(y) \\ \vdots \\ a_v\psi_{v-1}(y) \end{bmatrix} + BR \begin{bmatrix} a_1s\left(\frac{y}{\beta}\right) \\ 0 \\ \vdots \\ 0 \end{bmatrix} + CR \begin{bmatrix} a_1s\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Shorten the notation for the recurrence relation:

$$\begin{aligned}\psi_{k+1}(y) &= \psi\left(T_\beta^k(y)\right) = \frac{1}{w^k} \left[\psi(y) - \sum_{n=0}^{k-1} w^n s\left(T_\beta^n(x)\right) \right] \\ &= \frac{1}{w^k} [\Psi(y) - S_{k-1}(y)]\end{aligned}$$

and then plug through and collect distinct terms. The ψ terms give

$$\beta\lambda R \begin{bmatrix} a_1 \\ a_2w^{-1} \\ \vdots \\ a_vw^{1-v} \end{bmatrix} = wDR \begin{bmatrix} a_1 \\ a_2w^{-1} \\ \vdots \\ a_vw^{1-v} \end{bmatrix}$$

which has the form of an eigenvalue equation. The solution is given by

$$\det[wD - \beta\lambda I] = 0$$

and since $\det[D - \beta I] = 0$ conclude that $\lambda = w$. Note that this is independent of the matrix R .

The generating terms, those having only s in them, give

$$\beta R \begin{bmatrix} 0 \\ a_2S_0(y) \\ \vdots \\ a_vw^{2-v}S_{v-2}(y) \end{bmatrix} = DR \begin{bmatrix} 0 \\ a_3w^{-1}S_0(y) \\ \vdots \\ a_vw^{2-v}S_{v-3}(y) \end{bmatrix} - BR \begin{bmatrix} a_1s\left(\frac{y}{\beta}\right) \\ 0 \\ \vdots \\ 0 \end{bmatrix} - CR \begin{bmatrix} a_1s\left(\frac{y}{\beta} + \frac{1}{2}\right) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The goal is to find a rotation matrix that allows the above to be solved. The rotation matrix has $v(v-1)/2$ free parameters, whereas the above provides only v constraints. This would appear to allow for a fair amount of freedom.

An additional constraint is that the desired form requires a rotation R that keeps the $s\left(\frac{y}{\beta} + \frac{1}{2}\right)$ term alive, so that it can be used to get an equation of the form

$$s\left(\frac{y}{\beta} + \frac{1}{2}\right) = \sum_k c_k s(\beta^k y)$$

for some constants c_k . Since $s(x) = g(x \bmod \beta/2)$, the search is limited to $0 \leq y \leq \beta/2$. The left-hand side maps $[0, 1/2]$ to $[1/2, \beta/2]$ and thus can be used to cancel out any behavior at all, including the right-hand side. Thus, it is sufficient to set

$$g(y) = \begin{cases} h(y) & \text{for } 0 \leq y \leq \frac{1}{2} \\ -\sum_k c_k h(\beta^{k+1}(y - \frac{1}{2})) & \text{for } \frac{1}{2} < y \leq \frac{\beta}{2} \end{cases}$$

This works for any h whatsoever, unless h is constrained for some other reason. The function h can be freely chosen to be anything at all, for the $n=1$ and $n=2$ cases. The $n=3$ solution places a constraint on h ; it must be of the form $h(y) = y^p$ for any complex number p . For an explicit solution, and details on this constraint, refer to the accompanying diary. It is not clear when the general form will be similarly constrained or not. Still, a one-parameter family of solutions were obtained (for fixed w), which is plenty enough to develop a countable orthonormal basis (for fixed w).

Presuming such a rotation matrix R can be found, this gives a family of solutions to $\mathcal{L}_\beta f = wf$ for any complex w with $|w| < 1$.

6.6.2 Example solution for n=1

For the $n=1$ case, $p_1(\beta) = \beta^2 - \beta - 1$ has a root $\beta = r_1 = \varphi$. The midpoints are $m_0 = \beta/2$ and $m_1 = 1/2$ so $0 < m_1 < m_0 < 1$ and $I_2 = [0, 1/2]$ and $I_1 = [1/2, m_0]$. The interval inclusion map is given by

$$B = \Theta(I_j \subseteq LI_k) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \Theta(I_j \subseteq RI_k) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

The rotation matrix R can be taken to be the identity; it is not needed. Plugging through and cranking gives the solution.

$$f(y) = \begin{cases} f_2(y) = w\beta\psi(T_\beta(y)) & \text{for } 0 \leq y \leq \frac{1}{2} \\ f_1(y) = \psi(y) & \text{for } \frac{1}{2} < y \leq \frac{\beta}{2} \\ f_0(y) = 0 & \text{for } \frac{\beta}{2} < y \leq 1 \end{cases}$$

In the process of obtaining the above, one finds the constraint

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

that must hold for all $0 \leq y \leq \beta/2$. Satisfying this constraint is easy. The map takes $[0, 1/2]$ to $[1/2, \beta/2]$, and these two intervals are non-overlapping. Thus, it is sufficient to set

$$g(y) = \begin{cases} h(y) & \text{for } 0 \leq y \leq \frac{1}{2} \\ -w\beta^2 h(\beta(y - \frac{1}{2})) & \text{for } \frac{1}{2} < y \leq \frac{\beta}{2} \end{cases}$$

for any h whatsoever, and any complex w with $|w| < 1$. This will give a valid eigenfunction with eigenvalue $\lambda = w$. So, for $\beta = \varphi$ the golden mean, this provides the general solution.

Polynomial generators An interesting case arises, if one wishes $g(y)$ to be smooth over the entire interval $0 \leq y \leq \beta/2$. This ends up constraining the spectrum. A few examples.

First consider the form $g(x) = 1$. The constraint then forces $w = -1/\beta^2$.

The linear form is $g(x) = x + b$, which must satisfy

$$0 = w\beta^2(y + b) + \frac{y}{\beta} + \frac{1}{2} + b$$

The linear form forces $w = -1/\beta^3$ and $b = -1/2(1 + w\beta^2)$. Since $\beta = \varphi$, this reduces to $b = -(1 + \varphi)/2$

The quadratic form is $g(x) = x^2 + bx + c$ and the constraint forces

$$w = -\frac{1}{\varphi^4} = 5 - 3\varphi \quad b = -(\varphi + 1) \quad c = \frac{3\varphi + 2}{4}$$

It appears that this pattern can be followed, thus obtaining an eigenfunction with eigenvalue $w = -1/\beta^{k+2}$ for polynomial generators $g(x)$ of degree k .

6.6.3 Commentary

The $n = 2$ case is very nearly identical to the above. The solution to the $n = 3$ case can be exhibited (see the diary) but only after much grief. It requires a non-trivial rotation matrix R that depends on w . It also imposes a constraint on $h(y)$, which can no longer be arbitrary, but must have $h(y) = y^p$ for arbitrary complex number p . At any rate, it is solvable.

It is hypothesized that solutions are possible for all n . The rotation matrix R is needed to resolve multiple constraints introduced by the C matrix. It seems quite reasonable to think that this is always possible, although we have no proof in hand.

The immense freedom in the $n = 1$ and $n = 2$ cases, and the one-parameter family of solutions for the $n = 3$ case suggests that one can provide a complete set of orthonormal solutions to $\mathcal{L}_\beta f = wf$ for any complex w with $|w| < 1$. Explicit formulas for the $n = 1$ and $n = 2$ cases could be brute-forced, if desired. However, explicit formulas for this set for general n remain out of reach with the given toolset.

The coherent states provide a general solution. Presumably, this aligns with the piece-wise polynomial solutions, given earlier. Understanding how these align requires a review of what happens at $w = \beta^{-k}$.

The coherent state construction is performed for the countable set of finite orbits, enumerated by $p_n(\beta) = 0$. These give a collection of β that are dense in the interval $1 \leq \beta \leq 2$. Can the construction be extended to all β ? Doing so would require some way of obtaining the interval inclusion map that the B and C matrices encode. This inclusion map is implicit in the sofic shift, but none of the above constructions work at an abstraction level that was appropriate for exposing it. There's some Borel algebra here, but the above calculations were not sufficient to make it explicit.

There are unexplored tools. For example, most real numbers x have an uncountable number of distinct β -expansions for most values of β .[\[6\]](#) This is the opposite of the so-called “univoke” expansions.[\[7\]](#) Nowhere in the above constructions was this fact made use of, or touched upon. Yet it is implicitly there: the coherent state is iterating with $T_\beta^k(x)$ but it was never questioned or debated that this may result in different iterates, and thus different coherent states. This is bad: it was implicitly assumed that the coherent state is unique and well-defined, whereas perhaps each coherent state should be taken to be an uncountable family of such states. Whether they all come together in the end, or remain distinct, is entirely unclear.

Clearly, we've pried the lid off of a jar, and peered inside. It seems to hold a bottomless collection of tasks, each engendering new and ever-more-complex hypothesis. Of course, this is the way of mathematics, and one must be careful how one spends one's time on Earth. Still, it is all a bit unnerving.

7 Islands of Stability as Arnold Tongues

The classical Feigenbaum bifurcation diagram, shown in figure 5, manifests two distinct behaviors: the “islands of stability”, in which there are periodic orbits, and the “chaotic regions”. These regions alternate and interleave as a parameter λ appearing in the iterated equation $\lambda x(1-x)$ is swept through a range of values. By contrast, the equivalent diagram for the beta-map, shown in figure 2, does not seem to have regions of stability. This is only an illusion: they are there, they are only infinitely thin. This chapter focuses on how to crowbar them open, to finite size.

In the previous chapters, it was demonstrated that there is a countable set of β values, dense in the range $1 \leq \beta \leq 2$, for which orbits are finite and terminate after a fixed number of iterations. Alternately, they can be made periodic, simply by changing a less-than sign to a less-than-or-equals sign. The β values for which orbits are periodic can be placed in correspondence with the periodic orbits of the logistic map; the β values for which orbits are chaotic correspond to the chaotic orbits of the logistic map. The open problem is to demonstrate this correspondence explicitly. This problem is not tackled here; it is just brushed up against.

The circle map $x_{n+1} = x_n + \theta + K \sin 2\pi x_n$ provides one possible mechanism for taking a set of measure zero, and crowbaring it open to a set of finite size. For $K = 0$, this is just the rotation map $x_{n+1} = x_n + \theta$ which has only periodic orbits, when θ is rational, and chaotic orbits, when θ is not rational. As one sweeps θ through a range, the subset of periodic orbits is countable, and is a set of measure zero: thus, the rotation map can be said to be chaotic for almost all θ . Setting K to a non-zero value expands the regions of periodic orbits to finite size, termed Arnold tongues. These are the mode-locking regions that are generically visible in driven oscillator systems. The perturbation by the kick K displaces what would have been chaotic orbits into mode-locked regions. For small K , this perturbation is soft, in that one might say “it shouldn’t change things much.” But even a whisper of a miniscule K is enough to convert the set of periodic orbits from a measure of zero to a finite measure.

Another possibility is to just crowbar open the periodic regions with a “hard” perturbation, localized at a point. Take the natural saw-tooth shape of the β -map, widen the middle, and insert a slanting downward line, to create a zig-zag. That is, connect the two endpoints in the middle of the beta shift, “widening” it so that it has a finite, not infinite slope, thereby converting the iterated function from a discontinuous to a continuous one. This can be constructed directly: given some “small”, real $\varepsilon > 0$, define the piecewise-linear ε -generalization of the map 4 as

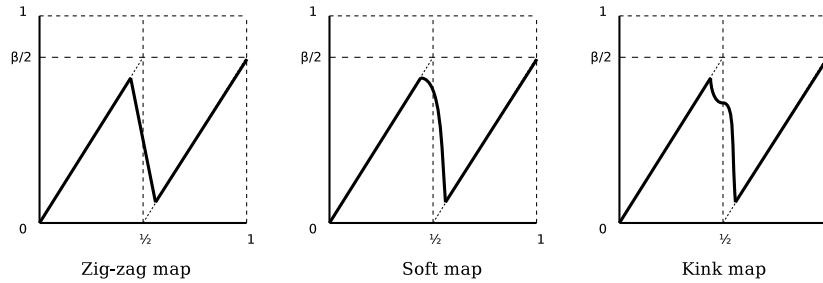
$$T_{\beta,\varepsilon}(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} - \varepsilon \\ \frac{\beta}{4} - \beta \left(\frac{1}{4} - \varepsilon\right) w & \text{for } \frac{1}{2} - \varepsilon \leq x < \frac{1}{2} + \varepsilon \\ \beta \left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} + \varepsilon \leq x \leq 1 \end{cases} \quad (32)$$

where w is just a handy notation for a downward sloping line:

$$w = \frac{2x - 1}{2\varepsilon}$$

Observe that $w = 1$ when $x = \frac{1}{2} - \varepsilon$ and that $w = -1$ when $x = \frac{1}{2} + \varepsilon$ so that w just

smoothly interpolates between +1 and -1 over the middle interval. The additional factors of $\frac{\beta}{4} - \beta (\frac{1}{4} - \epsilon) w$ just serves to insert the downward slope smack into the middle, so that the endpoints join up. The results is the zig-zag map, illustrated in the figure below



In the limit of $\epsilon \rightarrow 0$, one regains the earlier beta shift: $\lim_{\epsilon \rightarrow 0} T_{\beta, \epsilon} = T_{\beta}$, as the slope of the middle bit becomes infinite. The middle segment is a straight line; it introduces another folding segment into the map. This segment introduces a critical point only when ϵ is sufficiently large, and β is sufficiently small, so that its slope is less than 45 degrees (is greater than -1). When this occurs, a fixed point appears at $x = 1/2$. A sequence of images for finite ϵ are shown in figure 18.

The appearance of islands of stability in the Feigenbaum attractor is due to the presence of a fixed point at any parameter value. In order to “surgically add” islands of stability to the beta transform, the middle segment interpolation must also have a critical point at “any” value of ϵ . To achieve this, consider the curve

$$D_{\beta, \epsilon}(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} - \epsilon \\ \frac{\beta}{4} - \beta (\frac{1}{4} - \epsilon) g(w) & \text{for } \frac{1}{2} - \epsilon \leq x < \frac{1}{2} + \epsilon \\ \beta (x - \frac{1}{2}) & \text{for } \frac{1}{2} + \epsilon \leq x \leq 1 \end{cases} \quad (33)$$

where the straight line has been replaced by a soft shoulder

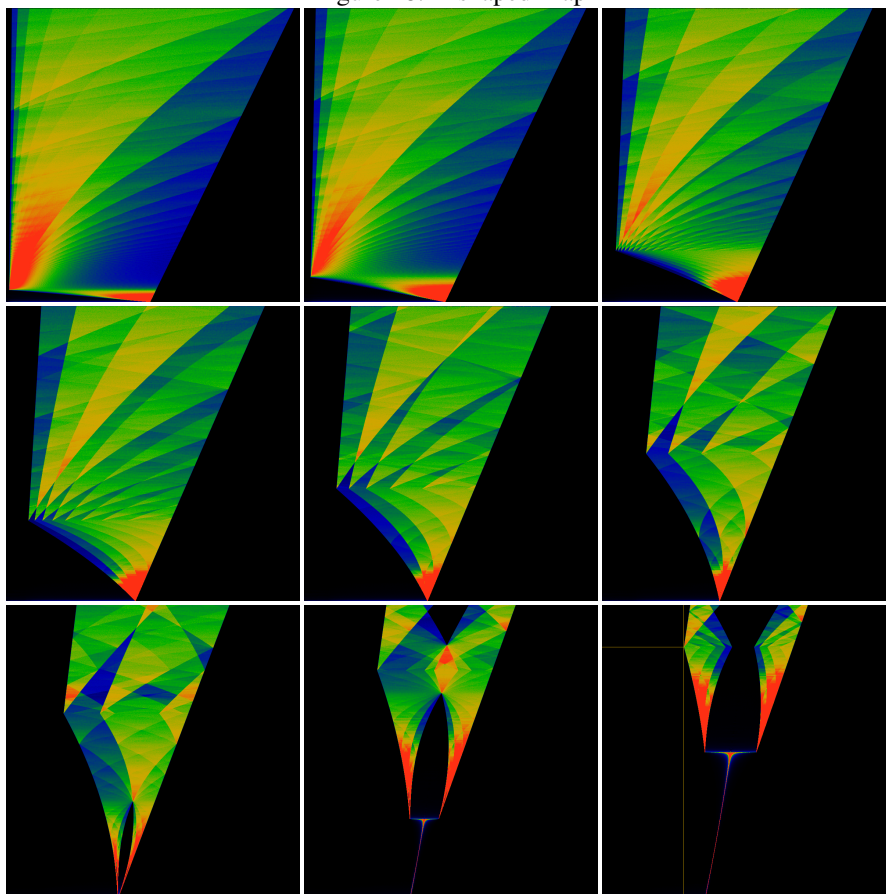
$$g(w) = 1 - 2 \cos \frac{\pi}{4} (1 + w)$$

and w is the same as before. This is scaled so that its a drop-in replacement for the straight line: $g(\frac{1}{2} - \epsilon) = 1$ and $g(\frac{1}{2} + \epsilon) = -1$. A cosine was used to create this soft shoulder, but a parabola would have done just as well. It is illustrated above, with the label “soft map”.

This map also interpolates between the left and right arms of the beta transform, forming a single continues curve. The curve is smooth and rounded near $\frac{1}{2} - \epsilon \lesssim x$, having a slope of zero as x approaches $\frac{1}{2} - \epsilon$ from above. This introduces a critical point near $\frac{1}{2} - \epsilon$. Notice that there is a hard corner at $\frac{1}{2} + \epsilon$. The interpolation is NOT an S-curve! A sequence of images for finite ϵ are shown in figure 19.

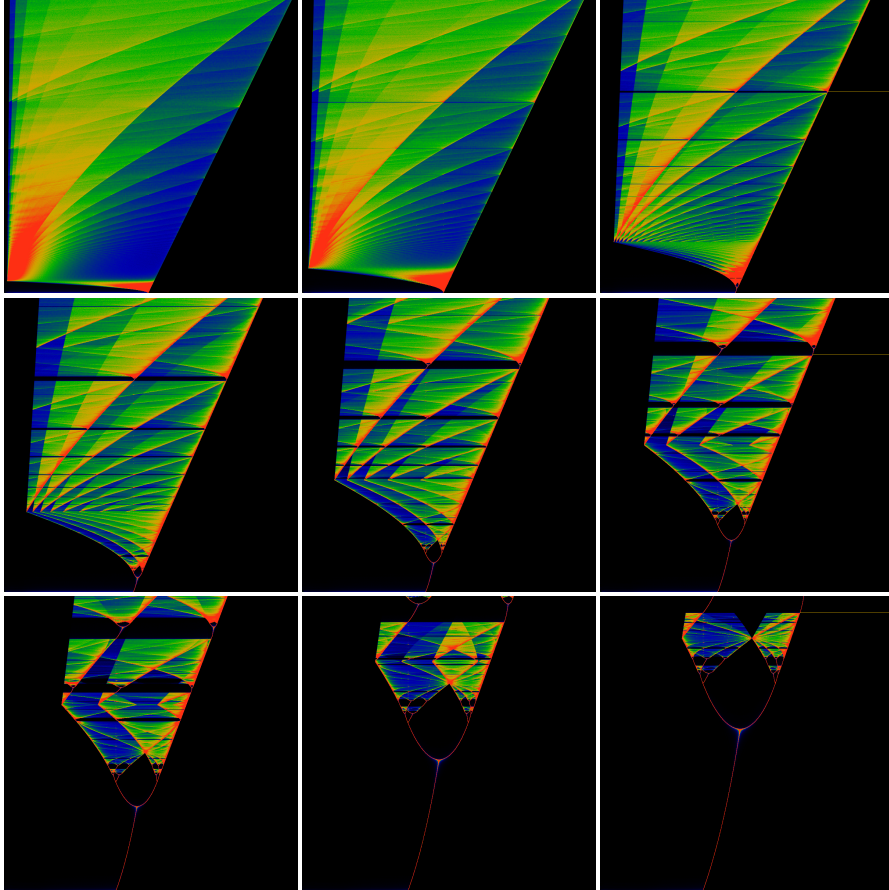
Two more variant maps can be considered. Both replace the center piece with

Figure 18: Z-shaped Map



This illustrates a sequence of iterated maps, obtained from eqn 32. Shown are $\varepsilon = 0.01, 0.02, 0.04$ in the first row, $0.06, 0.08, 0.10$ in the second row and $0.12, 0.14, 0.15$ in the third row. The image for $\varepsilon = 0$ is, of course, figure 2. The parameter β runs from 1 at the bottom to 2 at the top. Thus, a horizontal slice through the image depicts the invariant measure of the iterated map, black for where the measure is zero, and red where the measure is largest. The sharp corner at the lower-left is located $\beta = (1 + 2\varepsilon)/(1 - 2\varepsilon)$ and $x = \varepsilon(1 + 2\varepsilon)/(1 - 2\varepsilon)$. A yellow horizontal and vertical line in the last image indicate the location of this corner.

Figure 19: Critical-Z map



This illustrates a sequence of iterated maps, obtained from eqn 33. The sequence of depicted ε values are the same as in figure 18. The top row shows $\varepsilon = 0.01, 0.02, 0.04$, with $0.06, 0.08, 0.10$ in the second row and $0.12, 0.14, 0.15$ in the bottom row. The image for $\varepsilon = 0$ is, of course, figure 2. The parameter β runs from 1 at the bottom to 2 at the top. Working from bottom to top, one can see islands of stability forming in the $\varepsilon = 0.02$ and 0.04 images. The largest island, one third from the top, corresponds to $\beta = \varphi = 1.618 \dots$ the golden ratio. Moving downwards, the other prominent islands correspond to the “trouble spots” 101, 1001 and 10001, which are the Narayana’s Cows number, an unnamed number, and the Silver Ratio, at $\beta = 1.4655 \dots$ and so on. Moving upwards, one can see a faint island at the tribonacci number. Due to the general asymmetry of the map, these islands quickly shift away from these limiting values. For example, the primary island appears to start near $\beta = \delta + (2 - \delta)(\varphi - 1)$, where $\delta = (1 + 2\varepsilon)/(1 - 2\varepsilon)$. This location is indicated by a horizontal yellow line in the images in the right column. The other islands shift away in a more complicated fashion.

symmetrical sinuous S-shaped curves, but in different ways. Consider

$$S_{\beta,\varepsilon,\sigma}(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} - \varepsilon \\ \frac{\beta}{4} - \sigma \beta \left(\frac{1}{4} - \varepsilon\right) \sin \frac{\pi}{2} w & \text{for } \frac{1}{2} - \varepsilon \leq x < \frac{1}{2} + \varepsilon \\ \beta \left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} + \varepsilon \leq x \leq 1 \end{cases} \quad (34)$$

and

$$H_{\beta,\varepsilon,p,\sigma}(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} - \varepsilon \\ \frac{\beta}{4} - \sigma \beta \left(\frac{1}{4} - \varepsilon\right) \operatorname{sgn}\left(x - \frac{1}{2}\right) |w|^p & \text{for } \frac{1}{2} - \varepsilon \leq x < \frac{1}{2} + \varepsilon \\ \beta \left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} + \varepsilon \leq x \leq 1 \end{cases} \quad (35)$$

The $S_{\beta,\varepsilon}(x)$ replaces the central segment with a softly-rounded segment, containing two critical points: near $\frac{1}{2} - \varepsilon$ and near $\frac{1}{2} + \varepsilon$, where the curve flattens out to a zero slope. When $\sigma = +1$, the map as a whole is continuous. When $\sigma = -1$, the map consists of three discontinuous pieces. Different values are explored in figure 20.

The $H_{\beta,\varepsilon,p,\sigma}(x)$ replaces the central segment with a segment that has a kink in the middle, when $p > 1$. Note that $H_{\beta,\varepsilon,1,1}(x) = T_{\beta,\varepsilon}(x)$. Here, $\operatorname{sgn} x$ is the sign of x . The general shape of $H_{\beta,\varepsilon,p,\sigma}(x)$ is shown above, labeled as the “kink map”. The location of the kink in H is always centered; an off-center kink, as depicted in the figure, is explored below. The bifurcation diagrams for H are illustrated in figure 21.

To summarize: the “trouble spots” aren’t “just some periodic orbits” 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. And conversely: bifurcations can only appear here, and not elsewhere! The last sequence of images, shown in figure 21 indicate that the islands of stability need NOT consist of the period-doubling sequences seen in the Feigenbaum map. This is made explicit in figure 22, which shows a zoom by a factor of thirty.

Another interesting visualization is a Poincaré recurrence plot. The islands of stability should manifest as **Arnold tongues**. These are shown in figures 23 and 24.

To intuitively understand the location of the islands (the location of the Arnold tongues), its easiest to examine a map with a kink in it, whose location is adjustable.

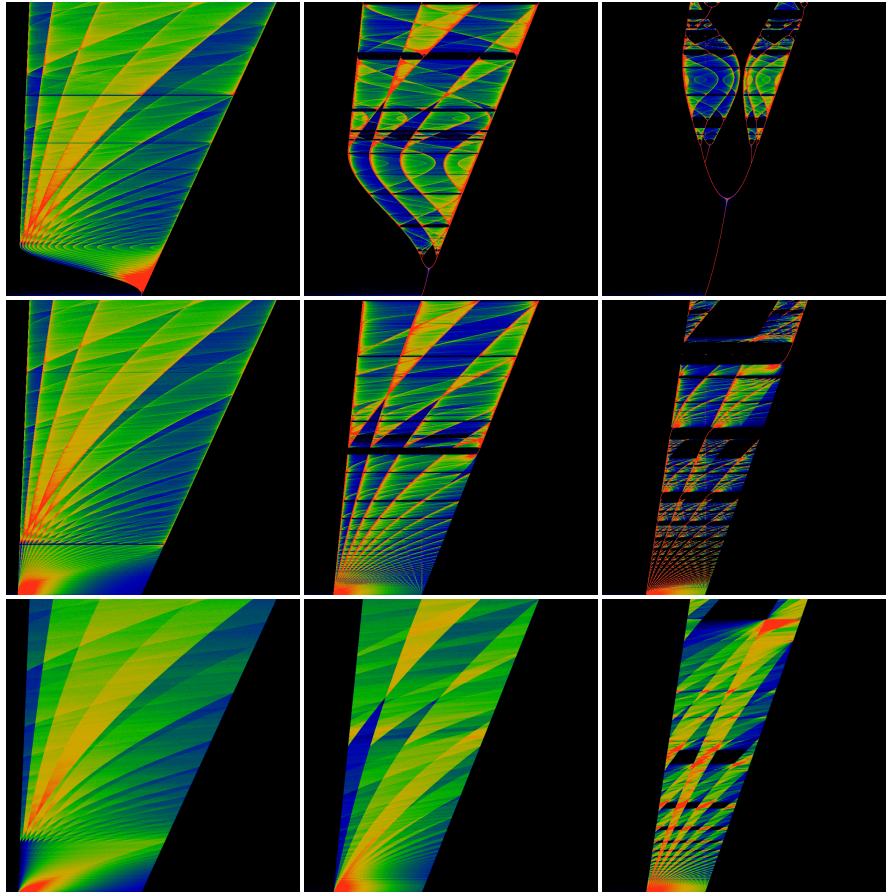
$$H_{\beta,\varepsilon,\alpha,\sigma}(x) = \begin{cases} \beta x & \text{for } 0 \leq x < \frac{1}{2} - \varepsilon \\ \frac{\beta}{4} - \sigma \beta \left(\frac{1}{4} - \varepsilon\right) h_{\alpha,p} & \text{for } \frac{1}{2} - \varepsilon \leq x < \frac{1}{2} + \varepsilon \\ \beta \left(x - \frac{1}{2}\right) & \text{for } \frac{1}{2} + \varepsilon \leq x \leq 1 \end{cases}$$

with

$$h_{\alpha,p}(x) = \begin{cases} \alpha + (1 - \alpha) |w|^p & \text{for } x < \frac{1}{2} \\ \alpha - (1 + \alpha) |w|^p & \text{for } \frac{1}{2} \leq x \end{cases}$$

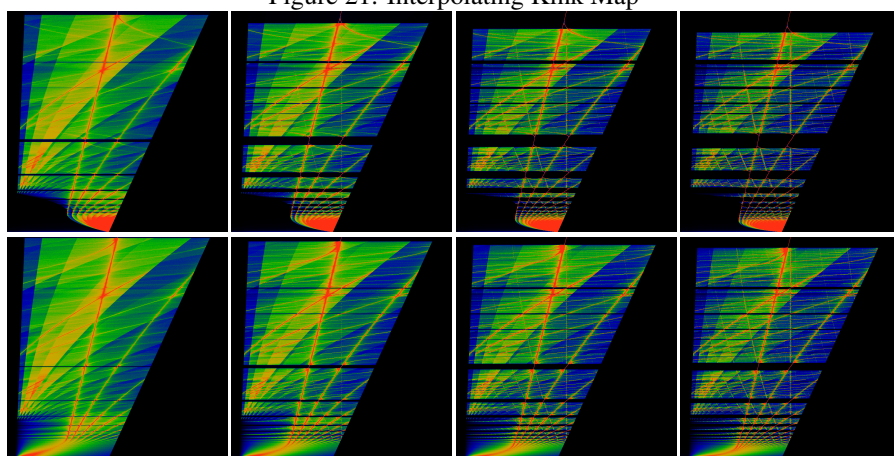
As before, $h_{\alpha,p}(x)$ is designed to interpolate appropriately, so that $h_{\alpha,p}\left(\frac{1}{2} - \varepsilon\right) = 1$ and $h_{\alpha,p}\left(\frac{1}{2} + \varepsilon\right) = -1$. The location of the kink is now adjustable: $h_{\alpha,p}\left(\frac{1}{2}\right) = \alpha$. Iterating on this map results in figures that are generically similar to those of figure 21, except that this time, the location of the islands is controllable by the parameter α . Roughly,

Figure 20: Interpolating Sine Map



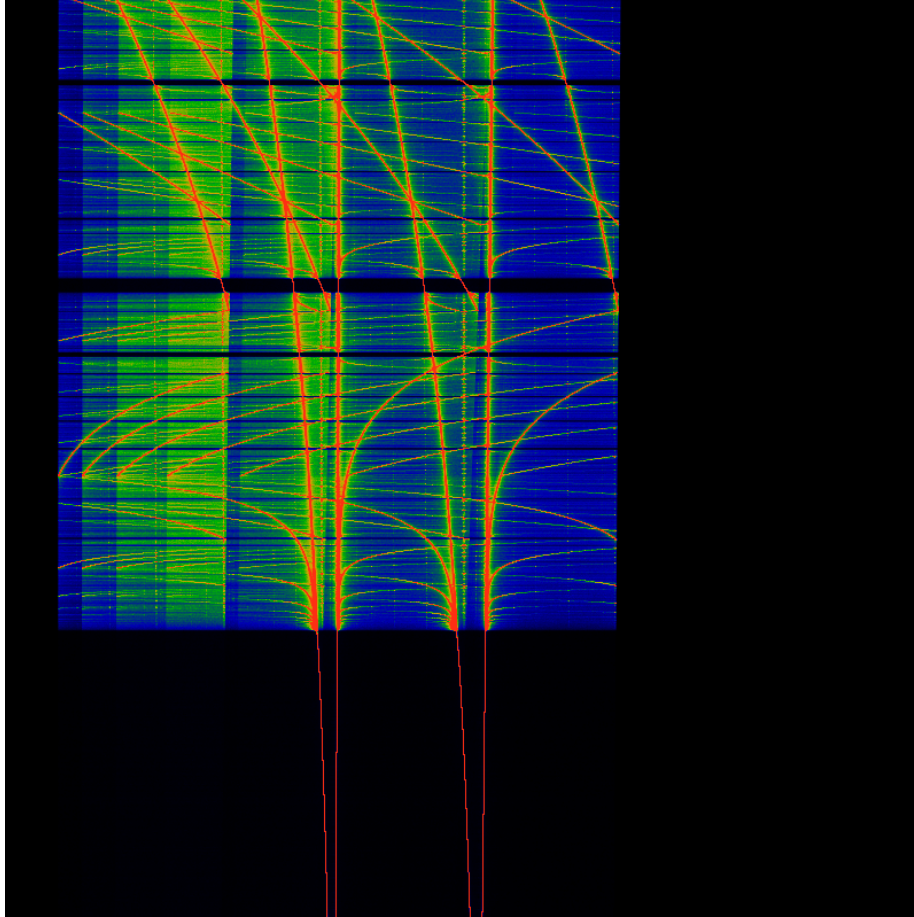
This illustrates a sequence of iterated maps, obtained from eqn 34. The sequence in the upper row shows $\varepsilon = 0.04, 0.10$ and 0.15 ; with $\sigma = +1$. The upper row is much like the sequence shown in figure 19, except that its made sinuous, thanks to symmetrical S-shape. The middle row shows the same ε values, but for $\sigma = -1$. The bottom row shows eqn 35 with $p = 1$ and $\sigma = -1$; thus, because $p = 1$ gives a straight-line segment in the middle, this bottom row is directly comparable to the zig-zag map. It should make clear that the islands appear in the middle row due to critical points in the S-curve, and not due to the tripartite map. The lower right diagram exhibits islands, but only because the middle segment has a slope of less than 45 degrees, resulting in a critical point at the middle of the map. As usual, the parameter β runs from 1 at the bottom to 2 at the top.

Figure 21: Interpolating Kink Map



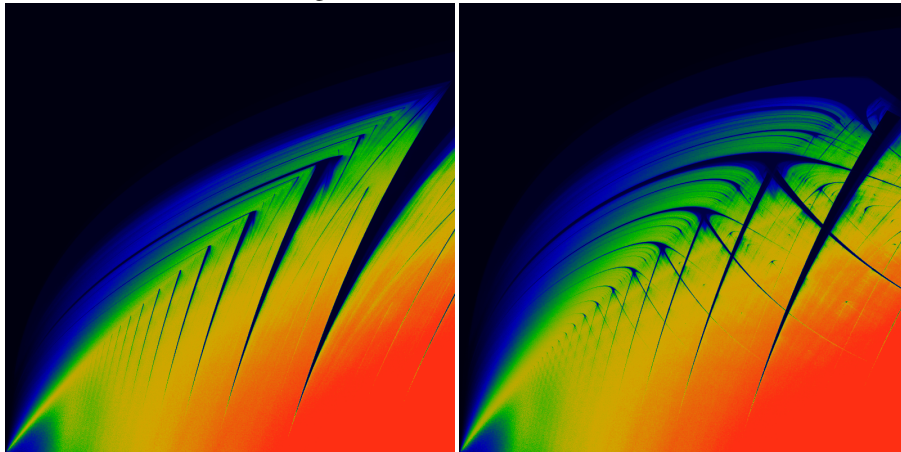
This illustrates a sequence of iterated maps, obtained from eqn 35. All eight images are held at $\varepsilon = 0.04$. The top row has $\sigma = +1$ (and thus the map is continuous) while the bottom row has $\sigma = -1$ (and thus the map has three disconnected branches. Left to right depicts the values $p = 2, 3, 4, 5$. As usual, the parameter β runs from 1 at the bottom to 2 at the top. In all cases, islands appear, and numerous common features are evident. Perhaps most interesting is that the islands do NOT contain period-doubling sequences. The primary sequence of islands, starting from the central largest, proceeding downwards, are located the inverse powers of two, viz at $\beta = \sqrt[p]{2}$. Why are the islands located at inverse powers of two, instead or, for example, the golden means? The short answer: it depends on the location of the kink in the map, as explored in the main text.

Figure 22: No Period Doubling



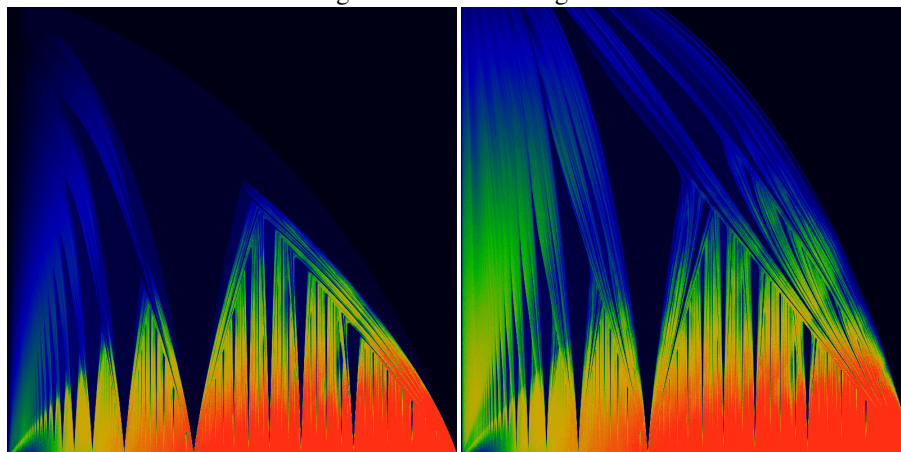
This figure is a zoom, confirming a lack of period doubling in the map $H_{\beta,\varepsilon,p,\sigma}(x)$ of eqn 35. The explored region is $0 \leq x \leq 1$, viz no zoom in the horizontal direction. Vertically, the image is centered on $\beta = 1.45$, having a total height of $\Delta\beta = 0.015625$. This uses the quintic kink, so $p = 5$ and $\sigma = +1$, making the the continuous variant. The value of $\varepsilon = 0.04$ makes this directly comparable to other images.

Figure 23: Poincaré recurrence



The above visualize the Poincaré recurrence times for the map $D_{\beta, \epsilon}(x)$ of eqn 33 on the left, and the map $S_{\beta, \epsilon, 1}(x)$ of eqn 34 on the right. In both cases, the parameter β runs from 1 to 2, left to right. The parameter ϵ runs from 0 to 0.2, bottom to top. The Poincaré recurrence time is obtained by iterating on the maps, and then counting how many iterations it takes to get near an earlier point. The color coding is such that yellow/red indicates large recurrence times; green is intermediate time, blue a short time, and black corresponds to n less than 3 or 4 or so. The vertical black spikes are the Arnold tongues; they correspond to parameter regions which lie in an island of stability. That is, the recurrence time is low, precisely because the the point x is bouncing between a discrete set of values. The yellow/red regions correspond to chaos, where the iterate x is bouncing between all possible values. The largest right-most spike is located at $\beta = \varphi = 1.618\dots$, with the sequence of spikes to the left located at the other primary golden means (*viz.*, 1.3803\dots and the silver mean 1.3247\dots and so on). As noted earlier, the general curve of that spike appears to follow $\beta = \delta + (2 - \delta)(\varphi - 1)$, where $\delta = (1 + 2\epsilon)/(1 - 2\epsilon)$. The dramatic swallow-tail shapes in the right-hand image are identical to those that appear in the [classic iterated circle map](#).

Figure 24: Arnold Tongues



The above visualize the Poincaré recurrence times for the map $H_{\beta, \epsilon, p, \sigma}(x)$ of eqn 35. The parameter β runs from 1 to 2, left to right. The parameter ϵ runs from 0 to 0.2, bottom to top. The power p is held fixed at $p = 5$. The left image shows $\sigma = -1$; the right shows $\sigma = +1$. The Poincaré recurrence time is obtained by iterating on $H_{\beta, \epsilon, p, \sigma}(x)$ and counting how many iterations it takes until $|x - H_{\beta, \epsilon, p, \sigma}^n(x)| < 0.009$. The shapes depicted are not sensitive to the recurrence delta 0.009; this value is chosen primarily to make the colors prettier. The color coding is such that yellow/red indicates large recurrence times n ; green is intermediate time, blue a short time, and black corresponds to n less than 3 or 4 or so. The vertical blue spikes are the Arnold tongues; they correspond to parameter regions which lie in an island of stability. That is, the recurrence time is low, precisely because the the point x is bouncing between a discrete set of values. The yellow/red regions correspond to chaos, where the iterate x is bouncing between all possible values. The central spike is located at $\beta = \sqrt{2}$ with the sequence of spikes to the left located at $\sqrt[k]{2}$ for increasing k . In that sense, the large black region dominating the right side of the figures corresponds to $\beta = 2$. These correspond to the black bands in figure 21.

to first order, the primary series of islands are located at $\sqrt[k]{2/(1-\alpha)}$; as before, these islands do not allow period-doubling to take place.

To get islands with period doubling, one needs to recreate the “soft shoulder” of eqn 33, but at a variable location.

Thus, the above presents a general surgical technique for controlling both the general form of the chaotic regions, the location of the islands of stability, and what appears within the islands.

Conjectures are fun! 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 beta shift $T_{\beta,\epsilon}$. 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.

Essentially, the claim is straight-forward: topologically, all chaotic parts of a map correspond to folding (as per Milnor, 1980’s on kneading maps), into which one may surgically insert regions that have cycles of finite length. The surgical insertion can occur only at the discontinuities of the kneading map. It almost sounds trivial, expressed this way; but the algebraic articulation of the idea would be worthwhile.

8 Conclusion

The idea of analytic combinatorics takes on a whole new meaning in the computational age. Historically, the ability to provide an “exact solution” in the form of an analytic series has been highly prized; the ultimate achievement in many cases. Being able to expression a solution in terms of the addition and multiplication of real numbers is very comforting. Every school student eventually comes to feel that arithmetic on the real numbers is very natural and normal. It’s more than that: Cartesian space is smooth and uniform, and all of differential geometry and topology are founded on notions of smoothness.

The inner workings of computers expose (or hide!) a different truth. The most efficient algorithm for computing $\sin(x)$ is not to sum the analytic series. Arbitrary precision numerical libraries open the rift further: neither addition nor multiplication are simple or easy. Both operations have a variety of different algorithms that have different run-times, different amounts of memory usage. In the effort to minimize space and time usage, some of these algorithms have grown quite complex. The root cause of the complexity is bewildering: it is the use of the binary digit expansion to represent a real number. Computers use the Cantor space $\{0, 1\}^{\mathbb{Q}}$ or at least a subset thereof, under the covers.

Different representations of the real numbers potentially offer different algorithms and performance profiles. One could represent reals by rationals, but then several other issues arise. One is that the rationals are not evenly distributed across the real number line: rationals with small denominators cluster about in a fractal fashion. This is easily exhibited by considering continued fractions. As a result, one promptly gets stuck in a quagmire of trying to understand what a “uniform distribution” should be. Binary expansions are more “obviously” uniform. A more basic issue is that, if working with rationals, one must somehow accomplish the addition or multiplication of 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.

Compare this situation to that of iterated functions and fractals. At first impression, these seem pathological in almost every respect: differentiable nowhere, unbounded and nonuniform: somehow they feel like the quintessential opposite of the analytic series, of the smoothness of Cartesian space, of the smoothness of addition and multiplication. The place where these two worlds come together is that both are attempts to approach countable infinity, and both are attempts to harness the first uncountable infinity. The real number number is an infinite string of binary digits. The analytic series is an infinite sum. The iterated function is recursively infinite. The historic labor of finding “exact solutions” to problems can perhaps be better views as the discovery of correspondences between finite structures (“the problem to be solved”) and infinite structures (“the solution”).

The situation here is more easily illustrated in a different domain. The hypergeometric series was presented and studied by Gauss; then 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 current situation is that there is no known algorithm that can enumer-

ate all such relations; there is no systematic way to classify them. There is an interplay between infinite series and algorithmic relationships between them. Stated a different way: hypergeometric series have a class of self-similarities, and the identities relating them are expressions of that self-similarity. What is that class of self-similarities? For the hypergeometric series, it remains unknown.

For Cantor space, that place where we represent real numbers, the situation is much better. The Cantor space itself has the structure of an infinite binary tree; the tree and its subtrees are obviously self-similar; the class of similarities is described by the dyadic monoid. The dyadic monoid embeds naturally into the modular group; this in turn is a gateway to vast tracts of modern mathematics. The recursive aspects, the shadow that the Cantor space seems to leave behind everywhere appears to be “explained” by Ornstein theory.

Yet, the picture remains incomplete. The β -transform provides a simple, silly model for multiplying two real numbers together: β and x . The “extra complication” of taking mod 1 after multiplication just reveals how complex multiplication really is. After all, mod 1 is just the subtraction of 1; how hard can that be? Moving in one direction: the fastest, most efficient-possible algorithm for multiplying two numbers is not known. Moving in another direction, the simple iterated maps, shown in figures 2, 4 and 5 are obviously not only self-similar, but also are surely topologically conjugate to one-another, and in all cases are presumably described by the dyadic monoid; likewise the Mandelbrot set and its exterior. Yet the details remain obscure.

The meta-question is: what is the correct framework by which one can best understand the interplay between symmetries, infinite series, infinite recursion and algorithms? Until modern times, mathematical practice has reified addition and multiplication into oracular operations that magically obtain “the right answer”. Modern computers have put a lie to this: the theory of numerical methods has made clear that addition and multiplication are necessarily algorithmic operations performed on finite truncations of infinite series. What other algorithms are hiding nearby, and what is their relationship to analytic series?

References

- [1] Georges de Rham, “On Some Curves Defined by Functional Equations (1957)”, in *Classics on Fractals*, edited by Gerald A. Edgar, Addison-Wesley, 1993, pp. 285–298.
- [2] A. Rényi, “Representations for real numbers and their ergodic properties”, *Acta Math Acad Sci Hungary*, 8, 1957, pp. 477–493.
- [3] W. Parry, “On the β -expansion of real numbers”, *Acta Math Acad Sci Hungary*, 11, 1960, pp. 401–416.
- [4] 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>.
- [5] A.O. Gel’fond, “A common property of number systems”, *Izvestiya Akad Nauk SSSR Seriya Matematicheskaya*, 23, 1959, pp. 809–814, URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=im&paperid=3814&option_lang=eng.
- [6] Nikita Sidorov, “Almost every number has a continuum of β -expansions.”, *The American Mathematical Monthly*, 110, 2003, pp. 838–842, URL <http://www.maths.manchester.ac.uk/nikita/amm.pdf>.
- [7] Martijn de Vries and Vilmos Komornik, “Unique Expansions of Real Numbers”, *ArXiv*, arXiv:math/0609708, 2006, URL <https://www.esi.ac.at/static/esiprpr/esil810.pdf>.
- [8] 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>.
- [9] P. Erdős and V. Komornik, “Developments in non-integer bases”, *Acta Math Hungar*, 79, 1998, pp. 57–83.
- [10] Nikita Sidorov, “Universal β -expansions”, *Arxiv*, 2002, URL <https://arxiv.org/abs/math/0209247v1>.
- [11] Boris Adamczewski, et al., “Rational numbers with purely periodic β -expansion”, *Bull London Math Soc*, 42, 2010, pp. 538–552, URL http://adamczewski.perso.math.cnrs.fr/AFSS_BLMS.pdf.
- [12] K. Schmidt, “On periodic expansions of Pisot numbers and Salem numbers”, *Bull London Math Soc*, 12, 1980, pp. 269–278.
- [13] F. Blanchard, “Beta-expansions and Symbolic Dynamics”, *Theoretical Comp Sci*, 65, 1989, pp. 131–141.

- [14] 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.
- [15] Shigeki Akiyama, “Finiteness and Periodicity of Beta Expansions - Number Theoretical and Dynamical Open Problems”, , 2009, URL <http://math.tsukuba.ac.jp/~akiyama/papers/proc/BetaFiniteCIRM.pdf>.
- [16] Louis-Sébastien Guimond, et al., “Arithmetics on beta-expansions”, *Acta Arithmetica*, 112, 2001, pp. 23–40, URL https://www.researchgate.net/profile/Edita_Pelantova/publication/259299735_Arithmetics_of_beta-expansions/links/5434e32a0cf294006f736e7c/Arithmetics-of-beta-expansions.pdf.
- [17] Bernat Julien, “Arithmetics in β -numeration”, *Discrete Mathematics and Theoretical Computer Science*, 9, 2006, URL <http://www.iecl.univ-lorraine.fr/~Julien.Bernat/arithbetanum.pdf>.
- [18] M. Hbaib and Y. Laabidi, “Arithmetics in the set of beta-polynomials”, *Int J Open Problems Compt Math*, 6, 2013, URL <http://www.i-csrs.org/Volumes/ijopcm/vol.6/vol.6.3.1.pdf>.
- [19] W. P. Thurston, “Groups, tilings and finite state automata”, *AMS Colloquium Lectures*, 1989.
- [20] Valérie Berthé and Anne Siegel, “Tilings associated with beta-numeration and substitutions”, *Integers: Electronic Journal of Combinatorial Number Theory*, 5, 2005.
- [21] Sh. Ito and H. Rao, “Purely periodic β -expansions with Pisot unit base”, *Proc Amer Math Soc*, 133, 2005, pp. 953–964.
- [22] Shigeki Akiyama, “Beta expansion and self-similar tilings”, , 2017, URL <http://cloud.crm2.uhp-nancy.fr/pdf/Manila2017/Akiyama.pdf>.
- [23] Valérie Berthé and Anne Siegel, “Purely periodic β -expansions in the Pisot non-unit case”, *ArXiv*, arXiv:math/0407282, 2004, URL <https://hal.archives-ouvertes.fr/hal-00002208/document>.
- [24] Jakob Grue Simonsen, “Beta-Shifts, their Languages, and Computability”, *Theory of Computing Systems*, 48, 2011, pp. 297–318, URL <http://www.diku.dk/~simonsen/papers/j12.pdf>.
- [25] Leopold Flatto, et al., “The Zeta Function of the Beta Transformation”, *Ergodic Theory and Dynamical Systems*, 14, 1994, pp. 237–266.
- [26] Daniel J. Thompson, “Irregular sets and conditional variational principles in dynamical systems”, , 2010, URL https://people.math.osu.edu/thompson.2455/thesis_thompson.pdf.

- [27] 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>.
- [28] Hans-Peter Blatt, et al., “On a generalization of Jentzsch’s theorem”, *J Approx Theory*, 159, 2009, pp. 26–38, URL <https://core.ac.uk/download/pdf/210808388.pdf>.
- [29] Kevin Hare, et al., “Three Series for the Generalized Golden Mean”, *Arxiv*, 2014, URL <https://arxiv.org/abs/1401.6200>.
- [30] 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>.
- [31] Solomon W. Golomb, “Irreducible polynomials, synchronizing codes, primitive necklaces and cyclotomic algebra”, *Proceedings Conf Combinatorial Mathematics and its Applications*, 1969, pp. 358–370.
- [32] Christophe Reutenauer, *Free Lie Algebras*, London Mathematical Society Monographs, 1993.
- [33] S. Duzhin and D. Pasechnik, “Automorphisms of necklaces and sandpile groups”, *ArXiv*, arXiv:math/1304.2563, 2013, URL <https://arxiv.org/pdf/1304.2563>.
- [34] Kevin Cattell, et al., “Fast Algorithms to Generate Necklaces, Unlabeled Necklaces, and Irreducible Polynomials over $GF(2)$ ”, *Journal of Algorithms*, 37, 2000, pp. 267–282, URL <http://www.cis.uoguelph.ca/~sawada/pub.html>].
- [35] Linas Vepstas, “The Bernoulli Map”, , 2004, URL <https://www.linas.org/math/bernoulli.pdf>, self-published on personal website.