ISLANDS AND ELLIPSES IN 2D DYNAMICAL SYSTEMS

Ted Paul Szylowiec

A thesis
in
The Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science (Mathematics)
Concordia University
Montréal, Québec, Canada

May 2023
© Ted Paul Szylowiec, 2023

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Ted Paul Szylowiec**

Entitled: **Islands and Ellipses in 2D Dynamical Systems**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

_____ Chair

Dr. Abraham Boyarsky

_____ Examiner

Dr. Jason Bramburger

_____ Examiner

_____ Examiner

_____ Supervisor

Dr. Pawel Góra

Approved _____

Dr Yogendra Chaubey, Graduate Program Director

_____ 20 _____ _____

Pascale Sicotte, Dean, Faculty of Arts and Science

Abstract

Islands and Ellipses in 2D Dynamical Systems

Ted Paul Szylowiec

Three main results are presented in this thesis. The first is a proof of the existence of absolutely continuous invariant measures (ACIMs) for two dimensional maps supported on *islands*, which are small, disjoint regions of $\mathbb{R}^2$. The proof is computer-assisted and uses both numerical evidence and a combinatorial method. We give examples of *weak chaos* for which ACIMs exist: within islands there is chaos, but from a distance orbits are periodic.

The second main result is a geometrical proof of the asymptotic behavior of generalized tent maps with memory which we call elliptical maps. It is proved that for certain π-rational angles, all points in the domain except for $(0, 0)$ fall into a polygonal region whose characteristics we determine. When the angles are π-irrational we prove that these points either fall in a unique ellipse, or accumulate on its boundary.

The third result is a proof that ACIMs exist for a certain range of parameters in generalized β-tent maps with memory.

The thesis begins with discussions about ACIMs and why they are interesting, followed by Tsujii's theorem and other tools, notes on computer calculations and graphics, and on weak chaos. At the end, we highlight some unanswered questions and puzzling phenomena that we encountered during our research.

Acknowledgments

Thanks foremost to Dr Pawel Góra for his guidance, for having me as his graduate student and for giving me the opportunity to collaborate on interesting research. Thanks to John T. Matthew, Br. Luc Lamontagne O.S.B and Chris Winter. Dedicated to the memory of my friend James C. Hayes.

Contents

List of Figures

List of Tables

Chapter 1

Absolutely Continuous Invariant Measures

1.1 Why ACIMs?

What are absolutely continuous invariant measures, why are they important and why are we interested in proving that they exist for dynamical systems? A dynamical system can be studied from two complementary points of view: deterministic or probabilistic. The probabilistic viewpoint can also be called the statistical or measure-theoretic viewpoint. Absolutely continuous invariant measures (ACIMs) pertain to this latter way of looking at dynamical systems.

A look at an example will provide the motivation and intuition for absolutely continuous measures and illustrate their importance. A good example for this purpose is the random-number generating dynamical system from the applications chapter of *Laws of Chaos* by Boyarsky and Góra [2, ch.13].

Consider the interval $I = [0, 1]$ and the family of maps $\tau : I \to I$ parametrized by n,

$$\tau(x) = (x + \pi)^n \pmod 1 \tag{1}$$

where n is an integer. We call τ a *map* when it is an endomorphism on I and can be iterated.

Figure 1 shows plots of τ for various n. For high enough values of n, the map τ can be used as a reasonably good random number generator. A choice is made for an initial value $x_0 \in I$, and then the orbit of x_0, which are the iterates of x_0 under

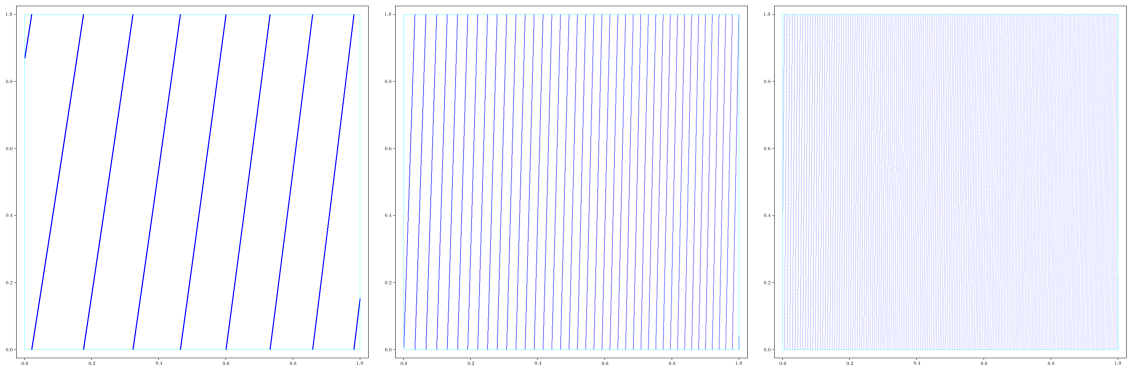

Figure 1: The map $\tau(x) = (x + \pi)^n \pmod 1$ with $n = 2$ (left), $n = 3$ (middle), and $n = 4$ (right). The reader may need to zoom in to observe the rulings when $n = 4$.

applications of τ, provide the pseudo-random numbers. Apparently this is how some calculators generate random numbers.

For example, conisider the orbits of the point x_0 shown in Figure 2 for $n = 1$ and $n = 4$. Clearly $n = 1$ won't do as a random number generator, but $n = 4$ seems like a good candidate.

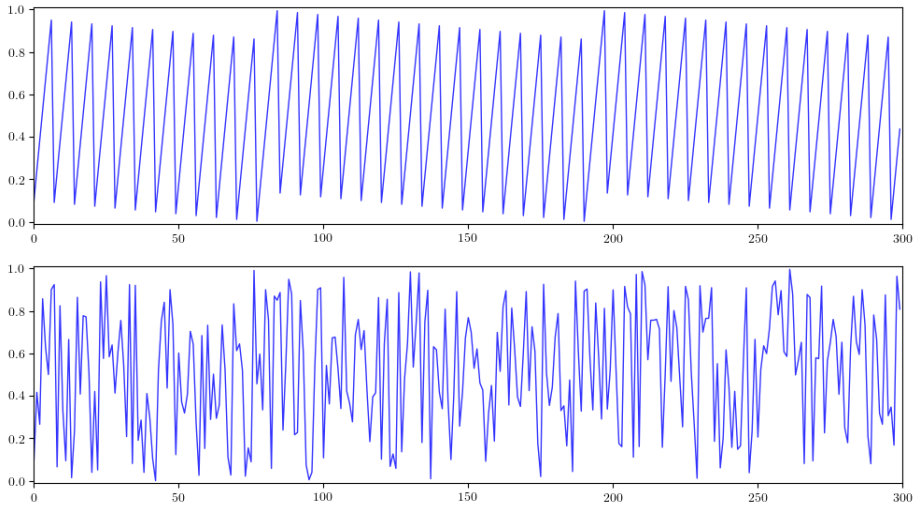

Figure 2: Orbits of the initial point $x_0 = 0.1$ under the map $\tau(x) = (x + \pi)^n \pmod 1$ with $n = 1$ (upper) and $n = 4$ (lower).

Some questions come to mind. What is meant by "random numbers" here? The numbers are generated deterministically. How can they be random? It is one of the conceptions of dynamical systems theory, that the two viewpoints, deterministic and probabilistic, do not necessarily exclude each other. At some point, which we will

illustrate shortly, the study of the deterministic behavior of a dynamical system is no longer profitable or even possible. Different tools, those of measure theory, must be used, and a statistical or probabilistic picture of the dynamical system's behavior emerges. We can have something resembling randomness, and whether this is or is not "true randomness" might not even be a sensible distinction to make.

Since τ is a map from I to I, it can be iterated, and it is natural to study the orbit of some initial point $x_0 \in I$ under iterations of τ:

$$x_0,\ \tau(x_0),\ \tau^2(x_0),\ \tau^3(x_0), \cdots . \tag{2}$$

This is the deterministic approach. It is easy to compute such orbits with computers, but there is the drawback that for a typical map τ, even a small perturbation in x_0 may lead to a wildly different orbit. Figure 3 shows the orbit of the point $x_0 = 0.1$ under τ with $n = 2$, and then, the orbit of x_0 perturbed by a tiny amount of 10^{-12}. Differences between the two orbits are already noticable after 15 iterations (the red line.) After 30 iterations the two orbits bear no resemblance to each other—they appear to be completely uncorrelated. That this happened so quickly for two initial points so close together should make us pause to reconsider what we are looking at when we look at long-term plots of orbits.

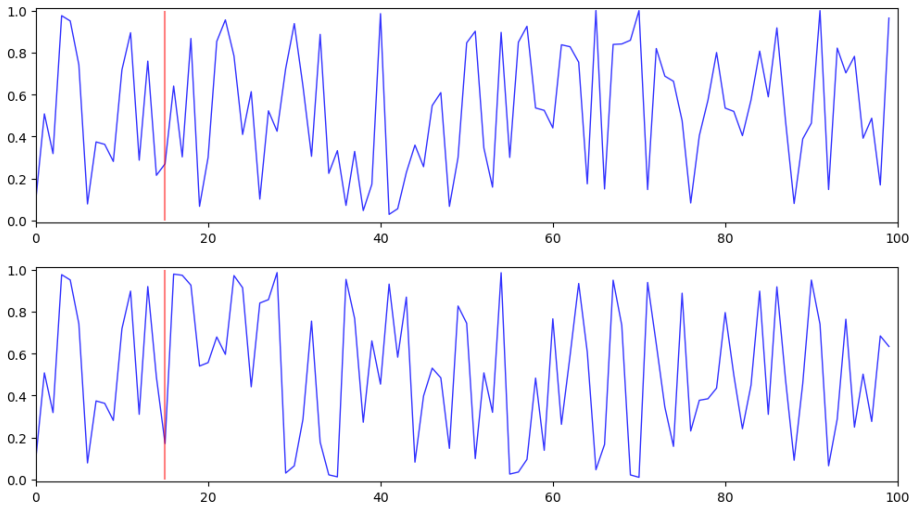

Figure 3: Orbits of $x_0 = 0.1$ under $\tau(x) = (x + \pi)^2 \pmod 1$ (upper) and in lower plot, x_0 has been perturbed by 10^{-12}. Red vertical line is at 15 iterations.

Computations for the plots in Figure 3 were done with 100-digit multiprecision numbers. 64-bit computer floats are limited to approximately 15 decimal digits of

precision, whereas we have seen that 12 digits of precision is not enough to accurately follow the orbit beyond a few handfuls of iterations of τ. Do plots created from long-term computer calculation of orbits with 64-bit floats really represent something meaningful? Perturbation upon perturbation is piled up due to the limitations of 64-bit representations of numbers. What we are looking at in long-term 64-bit orbit-plots cannot possibly resemble in exact detail the true orbit of x_0. We used multiprecision for Figure 3, but this only postpones the inevitable. Besides this, computing with hundreds or thousands of digits introduces another problem: it is orders of magnitude slower than computing with 64-bit floats.

However, we can switch to the other point of view, the probabilistic and measure-theoretic one. The long-term computed orbit is perhaps not similar to the real one in exact detail, but it may be similar in a statistical sense. This notion can be made more precise, and it can be asserted if we can prove the existence of an invariant measure, in particular an absolutely continuous invariant measure.

There are different ways that we can construct a statistical picture of the dynamics of τ. Some practical procedures are described in Lasota and Makey [17, ch.1]. One way is to choose a single initial point x_0 and then compute many terms, say N terms, of the orbit $x_0, \tau(x_0), \tau^2(x_0) \ldots$. The interval I is cut up into n disjoint bins $0 < j < n-1$ of equal width $\Delta_j = 1/n$. Then N_j, which is the number of times the orbit falls into bin j, is recorded.

Then the fraction of all orbit-points falling into bin j is $F_j = N_j/N$. This F_j is a *histogram* of the distribution of points in the orbit of x_0. Rather than F_j, if we take $f_j = nN_j/N$ as the bin fraction, then,

$$\sum_{J=1}^{n} f_j \Delta_j = \sum_{j=1}^{n} \frac{nN_j}{N} \Delta_j = n \left(\frac{1}{n}\right) \sum_{j=1}^{n} \frac{N_j}{N} = 1.$$

The area under f_j is 1, i.e, the histogram is normalized. This type of normalized histogram is called a *density* because it plays the role of a probability density. Using this density, we can construct a measure for subsets of I in the usual way: for $S \subset I$, determine how much of the density lies in S. For example, if the orbit of x_0 was uniformly randomly distributed over all I, then our measure would be essentially the same as the Lebesgue measure.

Here we have made some presuppositions concerning what our measure should look like: that it is in some sense a "good" one. It shouldn't matter where the orbit

begins, it should be possible to choose any x_0 in I, barring some pathological set of Lebesgue measure zero. It should also be possible to begin at any point in an orbit, and continue for another N iterations, and obtain a density that is similar, and with N large enough, identical to any density computed beginning somewhere else in the orbit, and hence the same measure.

A good measure should capture the long-term statistical behavior of a dynamical system and ignore temporary effects. For computations of orbits in this thesis, we typically skip the first 100,000 or 1,000,000 iterations. This is to disregard any initial, temporary phenomena that happen before the orbit settles into the attractor. A good measure should likewise ignore such impermanent phenomenon. We can think of this as analogous to the property of long-term solutions of time-dependent boundary value heat problems. There may be transient behavior along the way, but the long-term equilibrium heat distribution, which is a sort of physical analogue of our measure, does not describe such behavior.

A precise way to state the above requirement that our measure describes only long-term asymptotic phenomena, is that our measure should be τ-invariant, [1, ch.6.4], [2, ch.3]. If a subset $S \in \mathcal{B}(I)$ has measure $\mu(S)$, then we expect that $\tau(S)$ will also have the same measure. This is better expressed in terms of the preimage $\tau^{-1}(S)$. Let $(I, \mathcal{B}(I), \mu)$ be a measure space with $\mu(I) = 1$. Then μ is a *τ-invariant measure* if for any Borel set $S \in \mathcal{B}(I)$ we have

$$\mu(\tau^{-1}(S)) = \mu(S).$$

The preimages of all the orbit points in S will land in S upon the next iteration under τ. Therefore if μ is a good measure constructed from the density of orbit points, the measure of both S and $\tau^{-1}(S)$ must be the same.

A measure that is analogous to something physically reasonable appeals to intuition, but beyond that, the existence of invariant measures has important mathematical consequences. Consider two examples.

First there is the famous (and surprising) Poincare recurrence theorem. Let $(I, \mathcal{B}(I), \mu)$ be a measure space with $\mu(I) = 1$ and let μ be a τ-invariant measure. Then for any Borel set $S \in \mathcal{B}(I)$ with $\mu(S) > 0$, all points of S return infinitely often to S in their orbits under iterations of τ, except for points in a subset of S having measure zero.

The existence of an invariant measure also implies Birkhoff's ergodic theorem. Our

measure space $(I, \mathcal{B}(I), \mu)$ is *finite* because $\mu(I) = 1 < \infty$. But Birkhoff's theorem is true for wider types of measure spaces $(X, \mathcal{B}(X), \mu)$ which are σ-finite, i.e., $\mu(X)$ need not be finite, as long as X is a countable union of sets of finite measure. Suppose this is the case, and also that μ is a τ-invariant measure and f is some L^1 function. Birkhoff's theorem states that the following limit exists and is an L^1 function f^*:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j(x)\right) = f^* \quad \text{a.e.}$$

Now suppose further that, in particular, $\mu(X) < \infty$ and that τ is *ergodic*. By ergodic we mean that the effect of τ is indecomposable into separate parts of the space X, that is, if $\tau^{-1}(S) = S$ for some $S \in \mathcal{B}(X)$ then $\mu(S)$ is either zero or full measure. Then Birkhoff's theorem has the corollary

$$f^* = \frac{1}{\mu(X)} \int_X f \, d\mu \quad \text{a.e.}$$

If we think of the iterations of τ as analogous to stepping in time, then j plays the role of time. The limit expression $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(\tau^j(x)\right)$ is the time-average of f. The expression on the right-hand side in the corollary, $\frac{1}{\mu(X)} \int_X f \, d\mu$, is the space average of f. Therefore when μ is an invariant measure and τ is ergodic, the time-average of an L^1 function is equal to its space average [2, ch.3.3]. Notice that, up to a set of measure zero, these ergodic averages f^* do not depend on the initial point x chosen for starting the orbit. The existence of an invariant measure implies that certain important quantities that are computed by averaging over orbits—such as Lyapunov exponents—are well-defined [1, ch.6.5].

Aside from providing ways to compute ergodic averages, Birkhoff's theorem and its corollary have foundational implications for statistical physics.

The requirements that we have so far described for a good measure are not quite sufficient. It is possible for systems to have periodic point measures which are τ-invariant. In this case Birkhoff's theorem doesn't tell us much, and Poincare's recurrence theorem becomes trivial. We want to rule out such measures by requiring that a "good" measure be as good as the Lebesgue measure, in the sense that the measure should not be supported on sets of Lebesgue measure zero. This requirement reflects physical intuition: we don't have flux going through surfaces of zero area, or mass in solids of zero volume. Let μ and ν be two measures on a measure space $(I, \mathcal{B}(I))$. If for any $S \in \mathcal{B}(I)$ with $\mu(S) = 0$ we also have $\nu(S) = 0$, then we say that ν is

absolutely continuous with respect to μ. We can phrase our last requirement as: a good measure should be absolutely continuous with respect to the Lebesgue measure.

Thus, a good measure describing the long-term dynamics of the τ-map, is a τ-invariant measure absolutely continuous with respect to the Lebesgue measure. I.e., an *absolutely continuous invariant measure*, or ACIM for short.

One of the questions raised earlier was: *what are we looking at when we look at long term computer images of orbits?* Absolutely continuous invariant measures provide an answer to this question. In [6] and [2, ch.13.2] it is argued by Góra and Boyarsky that, while τ may have an unlimited number of invariant measures, the one revealed by computer computations of histograms is the ACIM. When an ACIM exists, the graphics created from computer histograms are faithful representations of the long-term properties of the dynamical system. Histograms made from computer-simulated orbits approach, in a sense, the true or theoretical ACIM of the τ system, and in addition, the ACIM is the only measure that is revealed by computer-generated histograms. This is especially relevant for us because this thesis contains plenty of graphical images made from the histograms of computer-simulated orbits.

A different way of constructing a statistical or probabilistic picture of τ-dynamics is discussed in Lasota and Makey [17, ch.1]. Instead of choosing an initial x_0, computing N terms of an orbit, then populating the bins of a normalized histogram to approximate a density, we can instead begin with a large number N of points, $(x)_N$ which are distributed according to some chosen initial probability density f_0. Then τ is applied to all the points $(x)_N$ and the resulting new collection of points $(\tau(x))_N$ is put into the bins of a normalized histogram, and an approximation to a new density f_1 emerges. What we have now is not a picture based on the evolution of a point as it travels along an orbit, but a picture based on the evolution of densities.

The logistic map $\tau(x) = 4x(1-x)$ provides a very pretty example which illustrates this evolution of densities clearly. The logistic map is one of the few maps for which there exists an elementary expression for the invariant density $f(x)$, discovered by Ulam and von Neuman,

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

In Figure 4, 1,000,000 initial points $(x)_N$ were randomly chosen using a uniform random generator, and an initial density f_0 was constructed from a 200-bin histogram. This initial f_0 is plotted in the upper left against the background of Ulam's invariant

density $f(x)$. With each successive application of τ, the new density rapidly evolves closer to $f(x)$ in a remarkable way. This phenomenon suggests a completely different way of producing a probabilistic picture of τ-dynamics: one based on functional analysis.

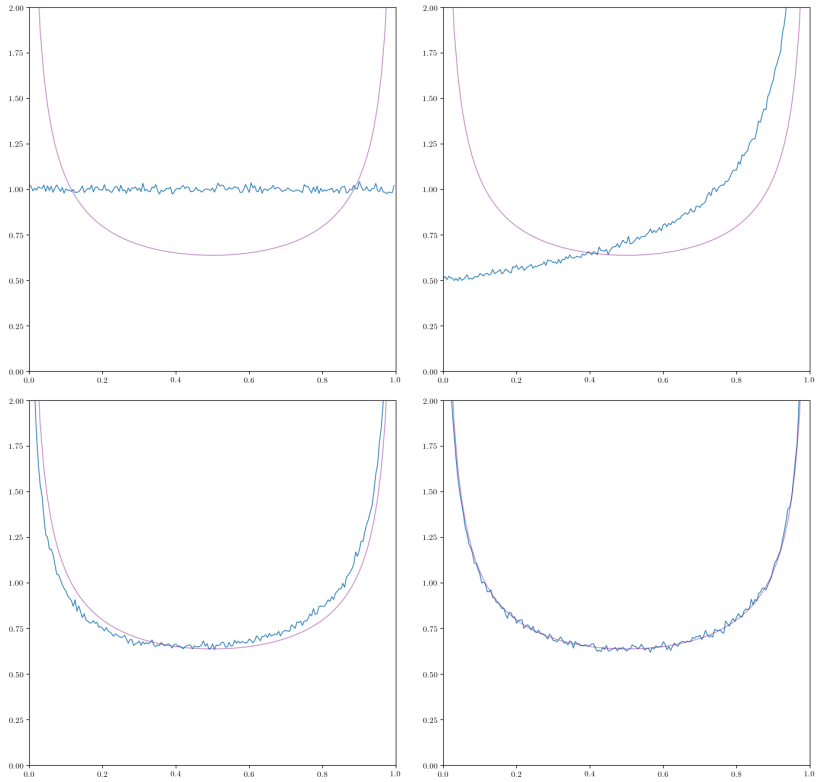

Figure 4: Evolution of logistic map density by successive applications of τ to the points in the inital distribution $(x)_N$. Shown are density histograms constructed from $(x)_N$, $(\tau(x))_N$, $(\tau^2(x))_N$ and $(\tau^5(x))_N$. The Ulam invariant density $f(x)$ is plotted in purple.

We can consider that the transformation from f_0 to f_1 is done by means of an operator P_τ, i.e., $P_\tau f_0 = f_1$. We can now study a sequence of evolving densities $f_0, f_1, f_2, \ldots$, rather than orbit points:

$$f_0,\ P_\tau f_0,\ P_\tau^2 f_0,\ \ldots\ P_\tau^k f_0, \ldots$$

If this sequence has a limit then that is a candidate for an invariant measure supported on the long-term density of the dynamical system. Fixed points of P_τ are the densities of absolutely continuous invariant measures [17, ch.4], and indeed for the logistic τ it can be shown that the Ulam density is a fixed point: $P_\tau f = f$. P_τ is called the Frobenius-Perron operator. It is a linear operator on L^1 and has many nice properties

[2, ch.4]. It is one of the main tools for discovering ACIMs in one dimension and proving their existence.

The evolution of densities for the random number generator map (1) with $n = 2$ is shown in Figure 5. We begin with a uniform density function $f_0(x) = 1$ and then determine $P_\tau^2 f_0$ and $P_\tau^5 f_0$ numerically by applying τ to 2,000,000 points sampled evenly along f_0 in 200 histogram bins. The plot on the right is already very close to the fixed-point invariant density [2, ch.1].

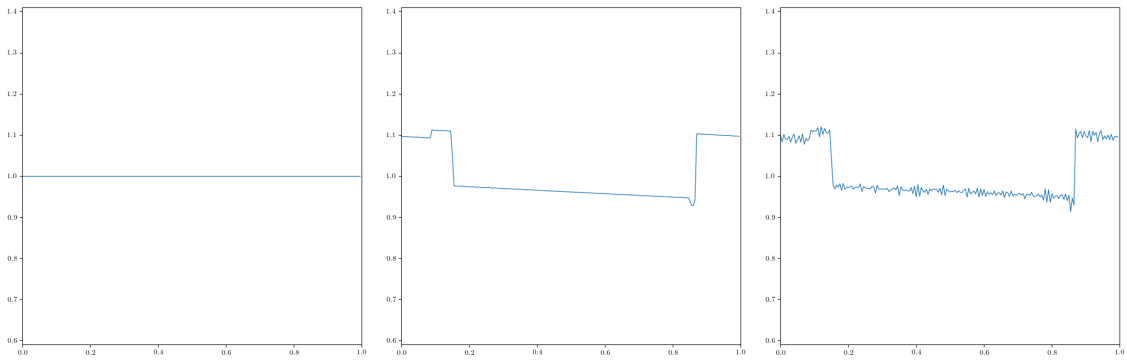

Figure 5: Initial uniform probability density of points, f_0 (left). Density after two iterations of the Frobenius-Perron operator, $P_\tau^2 f_0$ (middle). Density after five iterations, $P_\tau^5 f_0$ (right). Here $\tau = (\pi + x)^2 \pmod 1$.

1.2 One-dimensional ACIMs

The plots in Figure 5 are suggestive, but they do not prove that ACIMs exists for the τ in (1). We know that the logistic map τ admits an ACIM because it is a rare, exactly solvable case. The classic paper which established the existence of ACIMs for a wide and important class of one-dimensional maps τ was the 1973 paper by Lasota and Yorke [16]. They used the theory of functions of *bounded variation*. In the process of proving the existence of ACIMs, Lasota and Yorke derived inequalites concerning the variation of $P_\tau f$. Once an ACIM is known to exist for τ, it is from this type of Lasota-Yorke inequality that we can extract further interesting quantitative results about the density of the ACIM. We shall outline an example of this for the random generator maps (1), based on [2, ch.13.1].

Variation and *total variation* are generalizations of the concept of arclength. Let $I = [a, b]$ and $f : I \to \mathbb{R}$. Let $P = [x_0, x_1, \ldots, x_n]$ be a partition of I. Then the

variation of f with respect to P is

$$V_P(f) = \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| .$$

The *total variaton* of f, $TV(f)$, is the supremum of $V_P(f)$ over all partitions P of I:

$$TV(f) = \sup_P V_p(f).$$

A function f is of *bounded variation* if $TV(f) < \infty$. Many interesting properties about total variation and functions of bounded variation are proven in [21, ch.6.3], [2, ch.2.3] and [20, ch.8.3].

We follow the formulation of the Lasota-Yorke theorem as presented in [2, ch.5]. Let I be an interval and let $\tau : I \to I$ be a map with the following properties: (1) τ is piecewise C^1 expanding, that is, there exists a partition P such that τ restricted to any interval of P, is both C^1 and has $|\tau'(x)| \geq s$, where s is some number strictly greater than 1. (2) $1/|\tau'(x)|$ is a function of bounded variation. If τ satisfies conditions (1) and (2) then the Lasota-Yorke theorem says there exists an absolutely continuous τ-invariant measure whose associated density function is of bounded variation.

Before proceeding with the random generator example, we need a few results about functions of bounded variation.

A function of bounded variation is bounded. Suppose f is of bounded variation on $[a, b]$. Then there is an M such that $TV(f) \leq M$, and thus for all partitions P of $[a, b]$, $V_P(f) \leq M$. In particular this is true for the simple partition $P_1 = [a, y, b]$, where y is arbitrary in $[a, b]$. Then,

$$V_{P_1}(f) = |f(b) - f(y)| + |f(y) - f(a)| \leq M.$$

Since $|f(b) - f(y)| \geq 0$, it follows that $|f(y) - f(a)| \leq M$. Adding $|f(a)|$ to both sides and using the triangle inequality yields $|f(y)| \leq M + |f(a)|$. Now, it must be that $|f(a)| < \infty$ because $TV(f) \leq M$ and thus it must also be that $V_{P_0}(f) \leq M$ for the simplest partition $P_0 = [a, b]$. Therefore $f(x)$ is bounded on $[a, b]$.

If f is of bounded variation, then the bounds of $f(x)$ can be estimated from $TV(f)$. Let f be an L^1 function of bounded variation on $[a, b]$. Then there exists some y such that $|f(y)| \leq \|f\|_1/(b-a)$. Because if not, then $(b-a)|f(y)| > \|f\|_1$ for all $x \in [a, b]$, and integrating both sides of this inequality gives $\|f\|_1 > \|f\|_1$ which

is a contradiction. Combining this with the triangle inequality and the definition of $TV(f)$,

$$|f(x)| \leq |f(x) - f(y)| + |f(y)| \leq TV(f) + \frac{\|f\|_1}{b-a}. \tag{3}$$

Likewise there must be some y such that $|f(y)| \geq \|f\|_1/(b-a)$, for if not, we would obtain another contradiction similar to the one above. Combining this with the reverse triangle inequality,

$$|f(x)| \geq \big|\, |f(y)| - |f(y) - f(x)| \,\big| \geq \left| \frac{\|f\|_1}{b-a} - TV(f) \right|. \tag{4}$$

Returning to the random generator example and following the development in [2, 13.1], let $\tau(x) = p(x) \pmod 1$ where $p(x)$ maps I to the non-negative reals. Define α, β to be

$$\alpha = \inf_I |p'(x)|, \quad \beta = \sup_I \left| \frac{p''(x)}{p'(x)} \right|.$$

When $\alpha > 2$, τ admits an ACIM by the Lasota-Yorke theorem, with density f^*. An inequality for the variation of $P_\tau f$ for any function of bounded variation f is then derived in [2, 13.1]. From this inequality together with the fixed-point property of the ACIM density, $P_\tau f^* = f^*$, there follows bounds on the total variation of the ACIM density f^*:

$$TV(f^*) \leq (\beta + 2)/(\alpha - 2).$$

For $p(x) = (\pi + x)^n$, this becomes

$$TV(f^*) \leq \frac{(2\pi + n - 1)}{(n\pi^n - 2\pi)}.$$

When $n = 2$, $TV(f^*) \leq 0.54126$ and when $n = 5$, $TV(f^*) \leq 0.00675$. Because $f(x) \geq 0$, $b - a = 1$, $\|f^*\|_1 = 1$, and $TV(f^*) < 1$ when $n \geq 2$, the inequalities (3) and (4) combine to give

$$1 - TV(f) \leq f^*(x) \leq 1 + TV(f).$$

Plots of the invariant density f^* along with these theoretical bounds in red are presented in Figure 6. For $n = 2$, the bounds are $0.45874 \leq f^*(x) \leq 1.54126$, which is not a good guarantee that τ will behave well as a random number generator. However, for $n = 5$, the bounds are very narrow: $0.993 \leq f(x) \leq 1.00675$. This guarantee of near uniformity is one property that is sought-after in a good random number generator.

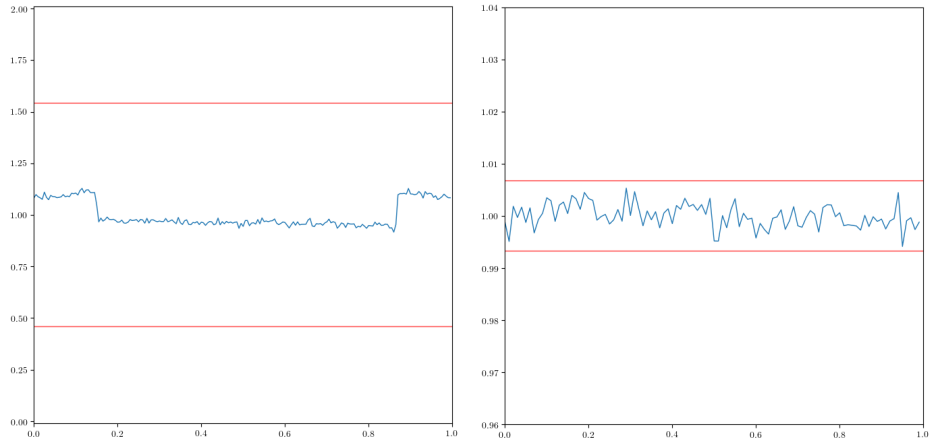

Figure 6: Invariant densities as limits of $P_\tau^k f_0$, with $\tau = (\pi + x)^n \pmod 1$ for $n = 2$ (left) and $n = 5$ (right). Theoretical bounds are shown in red.

It is impressive how, starting from the existence of an ACIM, we can prove statistical properties of long-term behavior of orbits that are geometrically unfathomably complex. Trying to prove things about the orbits of these τ maps by geometrically studying the motion of the iterates seems hopeless. And yet the probabilistic approach provides provable bounds on where the particle goes in the long run.

Despite this we will see in Chapter 3 geometrical methods re-asserting their preeminence in the study of a certain type of dynamical system.

1.3 Two-dimensional ACIMs

Our convention will be to use the symbol τ for one dimensional maps and G for two dimensional maps. Two dimensional densities are the main computational objects studied in this thesis. In this section we develop and present some concepts and propositions that will be needed for the main theorem of Chapter 2, Theorem 2.3.2. Most important of these is Tsujii's theorem which tells us when a certain type of piecewise two dimensional map G admits an ACIM.

The Lasota-Yorke theorem establishes the existence of ACIMs for a class of piecewise one dimensional maps. It is reasonable to assume something similar can be proved for two dimensions and higher, but this turned out not so straightforward. An account of the progress is given in [7].

The theorem of Tsujii [24], and simultaneously Buzzi [3], establishes the existence of ACIMs for a class of piecewise two dimensional maps to which the maps studied

in Chapter 2 belong. Some definitions are necessary before stating the theorem.

A function $f(t)$ on $\mathbb{R}$ is *real-analytic* if it has a power series expansion in t on every point in its domain. A function $f(x, y)$ on $\mathbb{R}^2$ is *real-analytic* if it has a power series expansion in x and y at every point in its domain. A *real-analytic curve* γ is the image in $\mathbb{R}^2$ of a real-analytic C^1 function $\gamma(t) : [0, 1] \to \mathbb{R}^2$ with $\|\gamma'(t)\| \neq 0$. A *real-analytic region* $X \subset \mathbb{R}^2$ is an open set which is bounded by a finite number of real-analytic curves.

G is a *piecewise real-analytic map* on $\mathbb{R}^2$ if (1) G is defined piecewise on a finite partition of real-analytic regions and (2) G on each partition region can be extended to the closure of the partition region.

$G : X \to X$, $X \subset \mathbb{R}^2$, is an *expanding map* if

$$\inf_{x \in X} \inf_{\|v\|=1} \|DG(x) \cdot v\| > c > 1. \tag{5}$$

Where $DG(x)$ is the derivative matrix of G at point x. In other words, for all vectors v, DG expands v in the sense that $\|DG \cdot v\| > \|v\|$ strictly.

If G is piecewise real-analytic then the iterates G^n are also piecewise real-analytic, and if G is an expanding piecewise real-analytic map, then so are the iterates G^n [24]. These are among the reasons for considering piecewise real-analytic maps. The Tsujii-Buzzi theorem states:

Theorem 1.3.1 *Let G be an expanding piecewise real-analytic map in $\mathbb{R}^2$. Then G admits an ACIM.*

The SVD or singular value decomposition is a tool that will allow us to determine if the maps considered in Chapter 2 are expanding. Suppose A is an $m \times n$ matrix. Then A acts from the left on column vectors $v \in \mathbb{R}^n$ as $v \to Av \in \mathbb{R}^m$ and from the right on row vectors $u^\top \in \mathbb{R}^m$ as $u^\top \to u^\top A \in \mathbb{R}^n$. It is possible to construct orthogonal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ in a natural way from A. Bases constructed this way for $\mathbb{R}^m$ and $\mathbb{R}^n$ are known as the *left singular vectors* and *right singular vectors* of A, respectively

Any square or rectangular $m \times n$ matrix A can be factored into the product of two orthogonal matrices U, $V^\top$ and a diagonal matrix Σ of positive elements called *singular values*, which are decreasingly ordered along the diagonal:

$$\begin{array}{ccccc} A & = & U & \Sigma & V^\top \\ m \times n & & (m \times m) & (m \times n) & (n \times n). \end{array} \tag{6}$$

This is called *singular value decomposition* or SVD. See Strang [23] and Kalman [11] for the justification of (6) as well as explorations of the the significance of it. In particular, Strang sees (6) as part of a "fundamental theorem of linear algebra". For our purposes, the SVD provides a way to understand the essential action of a matrix purely in terms of scaling: the othogonal matrices U and $V^\top$ only change orientation. The data that we need is in Σ.

If A is $m \times n$ then $A^\top A$ is an $n \times n$ symmetric positive-definite matrix. From (6),

$$A^\top A = \left(U\Sigma V^\top\right)^\top U\Sigma V^\top = V\Sigma^\top U^\top U\Sigma V^\top = V\Sigma^\top\Sigma V^\top.$$

This is the spectral decomposition of $A^\top A$. Because $A^\top A$ is positive-definite, $\Sigma^\top\Sigma$ will be an $n \times n$ diagonal matrix of non-negative eigenvalues of $A^\top A$. V is the matrix of associated orthonormal eigenvectors, which are a basis of $\mathbb{R}^n$. So V is the matrix whose columns are right-singular vectors. Likewise we can consider $AA^\top$ which is an $m \times m$ symmetric positive definite matrix. From (6),

$$AA^\top = U\Sigma V^\top \left(U\Sigma V^\top\right)^\top = U\Sigma V^\top V\Sigma^\top U^\top = U\Sigma\Sigma^\top U^\top.$$

This is the spectral decomposition of $AA^\top$ which is an $m \times m$ symmetric positive-definite matrix. The entries in the $m \times m$ diagonal matrix $\Sigma\Sigma^\top$ will be real and non-negative. The columns of U form an orthonormal basis of $\mathbb{R}^m$, and thus the columns of U are the left singular vectors.

Although their dimensions differ, the nonzero diagonal elements of $\Sigma^\top\Sigma$ and $\Sigma\Sigma^\top$ are always the same, because $A^\top A$ and $AA^\top$ have the same nonzero eigenvalues. Therefore the nonzero diagonal elements of Σ itself can be recovered by taking the square roots of the elements of $\Sigma^\top\Sigma$, the rest of the elements making up the $m \times n$ matrix Σ are always zero. The diagonal elements of Σ may always be arranged in decreasing order, by also rearranging the corresponding columns of V and U.

The singular value decomposition provides more information about $\mathbb{R}^m$ and $\mathbb{R}^n$. If the matrix A is square, say $m \times m$, but of rank $r < m$, then the first r columns of U (left singular vectors) forms an orthonormal basis for the column space of A, and the rest are a basis for the left nullspace. Likewise for columns of V: the first r form a basis for the row space, and the rest are a basis for the nullspace.

Practical calculations by hand for small (2×2) matrices can be done with the following method. First find the eigenvalue matrix for $A^\top A$. Then V and Σ can be read off from the diagonalization $A^\top A = V\Sigma^\top\Sigma V^\top$. Then notice that, from (6),

$AV = U\Sigma$. From this it is then easy to find U. The last expression, $AV = U\Sigma$, has an interesting interpretation. Suppose A is a square matrix. If v_k is the kth column of V, and u_k is the kth column of U. Then $Av_k = \sigma_k u_k$. This means that left-multiplication with matrix A transforms the right singular vectors into the left singular vectors, up to some some scaling factor. Indeed, it is easy to see, that because v_j, v_k are orthogonal, then Av_j, Av_k are also orthogonal:

$$Av_j \cdot Av_k = v_j^\top A^\top A v_k = \sigma_k^2 v_j \cdot v_k = 0.$$

Another way of understanding the significance of (6) is to expand the right-hand side into a linear combination of rank-1 matrices. A rank-1 matrix is the product $uv^\top$ of a column vector u and a row vector $v^\top$. The product will have dimensions $(m \times 1) \cdot (1 \times n) = m \times n$ which are the same dimensions as A, but it cannot have rank more than 1.

$$A = U\Sigma V^\top = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top + \cdots + \sigma_r u_r v_r^\top$$

where r is the rank of A. Further terms are all zero. Since the norms of the v_k and u_k are all 1, the terms are ordered in "size" or importance by virtue of the decreasing size of singular values. Depending on how quickly the singular values drop off, it is possible to construct approximations to A by taking relatively few of the largest terms of this series and ignoring the rest. An interesting exercise is to take a grayscale image, say 500×500 pixels, using that as matrix A, then decomposing it into the above series of rank-1 matrix terms. It is interesting to see that keeping only the first 50 terms is often enough to make a reasonable approximation of the original image.

A geometrical interpretation of singular values, discussed in [23] and [11], is especially clear and relevant to the case of full rank 2×2 matrices. A 2×2 full rank matrix A transforms a unit circle into an ellipse with semi-major and semi-minor axes σ_1 and σ_2. Thus if we are interested in whether a matrix is expanding, the SVD simplifies matters: the orthogonal matrices U and V don't contribute anything to the expansion, only the Σ matrix does. If the smaller singular value σ_2 is strictly greater than 1 then the larger σ_1 is also greater than 1, and every vector on the unit circle gets expanded by A because the smaller semi-axis of the image ellipse is still larger than 1. It is intuitively clear then, what the connection is between smaller singular value σ_2 and expanding matrices.

Proposition 1.3.2 *Let A be an $n \times n$ matrix of full rank and let $\|\cdot\|$ be the Euclidean 2-norm. Then*

$$\sigma_1 = \sup_{u \neq 0} \frac{\|Au\|}{\|u\|} = |||A|||, \quad \sigma_n = \inf_{u \neq 0} \frac{\|Au\|}{\|u\|}.$$

In particular, the largest singular value is the operator 2-norm of A, $|||A|||$.

Proof Consider the singular-value decomposition $A = U\Sigma V^\top$. Since U and V are orthogonal, they preserve norms: $\|Ux\| = \|x\|$ and $\|V^\top x\| = \|x\|$, Therefore,

$$\sup_{\|u\| \neq 0} \frac{\|Au\|}{\|u\|} = \sup_{\|u\| \neq 0} \frac{\|U\Sigma V^\top u\|}{\|u\|} = \sup_{\|u\| \neq 0} \frac{\|\Sigma V^\top u\|}{\|u\|}.$$

Let $w = V^\top u$. Then $Vw = VV^\top u = u$. Continuing,

$$\sup_{\|u\| \neq 0} \frac{\|\Sigma V^\top u\|}{\|u\|} = \sup_{\|u\| \neq 0} \frac{\|\Sigma w\|}{\|Vw\|} = \sup_{\|u\| \neq 0} \frac{\|\Sigma w\|}{\|w\|}$$

The expression $\|\Sigma w\|/\|w\|$, when written in full, becomes

$$\frac{\|\Sigma w\|}{\|w\|} = \frac{\left(\sum_{k=1}^n \sigma_k^2 w_k^2\right)^{1/2}}{\left(\sum_{k=1}^n w_k^2\right)^{1/2}}$$

where w_k are component of the vector w. The expression on the right-hand side is the weighted mean of the σ_k with weights w_k. Since σ_1 and σ_n are the max and min of the σ_k, this implies that

$$\sigma_n \leq \frac{\|\Sigma w\|}{\|w\|} \leq \sigma_1.$$

If we use a weight vector w such that $w_1 = 1$ and $w_{k>1} = 0$, then the weighted mean expression is σ_1. The supremum of the weighted mean expression must also be σ_1 since it cannot be larger, and we have $\sup_{\|u\| \neq 0} \|Au\|/\|u\| = \sigma_1$. Likewise regarding infimum: if we choose a weight vector w where $w_n = 1$ and $w_{k<n} = 0$, the weighted mean expression will be σ_n. Since this is the smallest possible value for the weighted mean expression, we must have $\inf_{\|u\| \neq 0} \|Au\|/\|u\| = \sigma_n$.

It is sometimes useful to have the following inequality concerning the smallest singular value σ_n of a product of matrices.

Proposition 1.3.3 *Let A and B be $n \times n$ matrices. Then $\sigma_n(AB) \geq \sigma_n(A)\sigma_n(B)$.*

Proof If either A or B is not full rank then this is true in a trivial sense, because smallest singular value of such a matrix is zero. Let $\sigma_n(M)$ be the smallest singular value of an $n \times n$ matrix M. Using Proposition 1.3.2, $\sigma_n(M) = \inf_{\|u\|=1} \|Mu\|$, and therefore

$$\begin{aligned}\sigma_n(AB) &= \inf_{\|u\|=1} \|ABu\| = \inf_{u\neq 0} \frac{\|ABu\|}{\|u\|} = \inf_{u\neq 0} \frac{\|A(Bu)\|}{\|Bu\|}\frac{\|Bu\|}{\|u\|} \geq \inf_{v\neq 0} \frac{\|Av\|}{\|v\|} \inf_{u\neq 0} \frac{\|Bu\|}{\|u\|} \\ &= \sigma_n(A)\sigma_n(B).\end{aligned}$$

From Proposition 1.3.2, we see that the infimum of stretching of vectors by matrix A is σ_2. Thus by the definition in (5), G is expanding if $\sigma_2(DG) > 1$ over all $x \in X$.

All singular values computations in Chapter 2 are done on 2×2 matrices. If we do not need U or V but just Σ, then the following derivation of direct algebraic formulas for σ_1 and σ_2 will suffice. U and V are orthogonal, therefore they preserve Frobenius norm and we have

$$\|A\|_F^2 = \|U\Sigma V^\top\|_F^2 = \|\Sigma\|_F^2 = \sigma_1^2 + \sigma_2^2.$$

Since U and $V^\top$ are orthogonal, their determinants are ± 1, and we have

$$\begin{aligned}|\det A| &= \left|\det(U\Sigma V^\top)\right| = \left|\det(U)\det(\Sigma)\det(V^\top)\right| \\ &= |(\pm 1)\det(\Sigma)(\pm 1)| = |\det \Sigma| = |\sigma_1\sigma_2| \\ &= \sigma_1\sigma_2.\end{aligned}$$

Let $p = \|A\|_F^2$, and $q = |\det A|$. We can solve the equations $p = \sigma_1^2 + \sigma_2^2$ and $q = \sigma_1\sigma_2$ for σ_1 and σ_2. Letting $r = \sqrt{p + 2q}$ and $s = \sqrt{p - 2q}$, we have the pretty solution

$$\sigma_1 = \frac{r+s}{2}, \quad \sigma_2 = \frac{r-s}{2}. \tag{7}$$

If $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ then the square of the Frobenius norm is $p = a^2 + b^2 + c^2 + d^2$ and the magnitude of the determinant is $q = |ad - bc|$, and the singular values σ_1 and σ_2 follow from (7).

We will need to determine if G^k is an expanding map. For this purpose we need some way to simplify the computation of derivative matrices of iterates of G.

It is often useful to know the derivative of an iterate τ^k of a one dimensional map at a point x_0. For example, to determine if a periodic orbit is attracting or repelling. Computing the composite iterate $\tau^k = \tau \circ \tau \circ \cdots \circ \tau$ and then taking the derivative is sure to be be a hopeless task for most τ. However, if x_0 is an orbit point, then

the chain rule can be exploited to simplify the calculation of $(\tau^k)'(x_0)$. With orbit x_0, $x_1 = \tau(x_0)$, $x_2 = \tau(x_1), \ldots$, the calculation of $(\tau^k)'(x_0)$ reduces to the product of derivatives of τ at each orbit point: $\tau'(x_{k-1}) \cdots \tau'(x_1)\tau'(x_0)$. This is proved in many texts on dynamical systems, for example [1, ch.1.4].

In two dimensions, the composite map $G^k = G \circ G \circ \cdots \circ G$ is a multivariate function and so rather than derivative, we are looking for its derivative matrix $D(G^k)$. Luckily a similar simplification works for the multivariate case: computation of the derivative matrix of G^k reduces to matrix products of DG, the derivative matrix of G, evaluated on the orbit points.

Lemma 1.3.4 *Let $X \subset \mathbb{R}^n$ and let G be a map from X to X which is differentiable at the points of the orbit of x_0, i.e., at x_0, x_1, $x_2 \ldots x_{k-1}$. Then $DG^k(x_0)$, the derivative matrix of G^k at x_0, can be calculated by*

$$DG^k(x_0) = DG(x_{k-1}) \cdot DG(x_{k-2}) \cdot \cdots \cdot DG(x_1) \cdot DG(x_0).$$

Proof In what follows, $Dg(f(x))$ means the derivative matrix of g, evaluated at $f(x)$. $D(g \circ f)(x)$ means the derivative matrix of the composite function $g \circ f$, evaluated at x.

The multivariable chain rule and its proof can be found in [22, ch.2]. Let $f : \mathbb{R}^m \to \mathbb{R}^k$ be differentiable at x and let $g : \mathbb{R}^k \to \mathbb{R}^n$ be differentiable at $f(x)$, then the composition $g \circ f : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at x, with $D(g \circ f)(x) = Dg(f(x)) \circ Df(x)$.

Now let $m = n = k$ which is our case. We then have

$$D(g \circ f) = D(g \circ f)(x) = Dg(f(x)) \cdot Df(x) \tag{8}$$

because composition of matrices is just matrix multiplication.

Let $X \subset \mathbb{R}$ and let $G^n : X \to X$ be a map. The orbit of $x_0 \in X$ up to the x_{n-1} term is

$$x_0, \quad x_1 = G(x_0), \quad x_2 = G(x_1), \quad \cdots \quad x_{n-1} = G(x_{n-2}) = G^{n-1}(x_0).$$

Using (8), we obtain for $DG^2(x_0)$,

$$DG^2(x_0) = D(G \circ G)(x_0) = DG(G(x_0)) \cdot DG(x_0) = DG(x_1) \cdot DG(x_0)$$

And again by (8), we have, for $DG^3(x_0)$,

$$DG^3(x_0) = D(G \circ G^2)(x_0) = DG(G^2(x_0)) \cdot DG^2(x_0) = DG(x_2) \cdot DG(x_1) \cdot DG(x_0).$$

The rest follows by induction.

The following lemma does not specifically pertain to maps of two dimensions or higher, but it will be used in that context. It will be the case in Chapter 2 that the existence of an ACIM for G cannot be proven directly, but the existence of an ACIM for G^p can. The following lemma allows us to deduce that G admits an ACIM whenever G^p does.

Lemma 1.3.5 *Let $G^p : X \to X$ and suppose G^p admits an absolutely continuous invariant measure supported on X. Then G also admits an ACIM supported on X.*

Proof Suppose G^p admits an ACIM μ. Then an invariant measure for G can be constructed from μ and the *push-forward* G_*. The push-forward is an operator on measures. Let $(X, \mathcal{B})$ be a measure space and let $\mathcal{M}(X, \mathcal{B})$ be the space of measures on $(X, \mathcal{B})$. G_* is a mapping from $\mathcal{M}(X, \mathcal{B})$ to $\mathcal{M}(X, \mathcal{B})$. For a measure $\mu \in \mathcal{M}(X, \mathcal{B})$, $G_*\mu$ is a new measure defined by $(G_*\mu)(S) = \mu(G^{-1}(S))$, for any $S \in \mathcal{B}$. Construct a measure ν in terms of G_*:

$$\nu = \frac{1}{p}\left(\mu + G_*\mu + G_*^2\mu + \cdots + G_*^{p-1}\mu\right).$$

Since μ is a G^p-invariant measure, then $\mu(G^{-p}(S)) = \mu(S)$ and we have

$$\begin{aligned}
\nu(G^{-1}(S)) &= \frac{1}{p}\left(\mu(G^{-1}(S)) + \mu(G^{-2}(S)) + \cdots + \mu(G^{-p}(S))\right) \\
&= \frac{1}{p}\left(\mu(G^{-1}(S)) + \mu(G^{-2}(S)) + \cdots + \mu(G^{-p+1}(S)) + \mu(S))\right) \\
&= \frac{1}{p}\left(\mu(S) + \mu(G^{-1}(S)) + \mu(G^{-2}(S)) + \cdots \mu(G^{-p+1}(S))\right) \\
&= \nu(S)
\end{aligned}$$

and ν is a G-invariant measure. The factor of $1/p$ in the construction of ν assures that it is a probablilty measure: $\nu(X) = 1$. When μ is absolutely continuous with respect to the Lebesgue measure, then so is ν. Therefore G admits an ACIM supported on X.

Chapter 2

Islands

2.1 Weak chaos

In a 2009 article titled *Birds and Frogs* [5], Freeman Dyson writes about something he calls *weak chaos*:

> "The problem of weak chaos is still unsolved fifty years later. The problem is to understand why chaotic motions often remain bounded... Chaotic behavior was never suspected in the solar system before accurate long-term integrations were done, because the chaos is weak. Weak chaos means that neighboring trajectories diverge exponentially but never diverge far. The divergence begins with exponential growth but afterwards remains bounded... I challenge you to understand the reasons why the chaos observed in a great diversity of dynamical systems is generally weak."

We will see examples of a special kind of weak chaos known as *asymptotic periodicity*. A point moving in two dimensions under iterations of a map G can, in the right circumstances, describe an orbit which is completely chaotic but bounded within small disjoint *islands*. Within the islands there is chaos, yet the islands themselves behave like perfectly periodic clockwork. Dyson continues, mentioning the famous "period three" theorem of Li and Yorke [18]:

> "The subject of chaos is characterized by an abundance of quantitative data, an unending supply of beautiful pictures, and a shortage of rigorous theorems... The theorem [period three] explains why chaos is prevalent in

> the world. It does not explain why chaos is so often weak. That remains a task for the future. I believe that weak chaos will not be understood in a fundamental way until we can prove rigorous theorems about it."

In this chapter we will indeed supply some beautiful pictures, but we also hope to address Dyson's main point which concerns proving things about weak chaos.

Systems exhibiting both chaos and periodic behavior were studied theoretically using functional analysis techniques by Lasota, Li and Yorke [15], Komornik, [13], [14] and Komornik and Lasota [12].

However, it was later in 2017, [10], that asymptotic periodicity was observed to occur naturally in the course of studies about dynamical systems with memory. In 2021, Nakamura and Mackey [19] also reported natural examples of asymptotic periodicity arising from similar iterated two-dimensional maps.

Observed from afar, this type of weak chaos appears periodic, and the periodic features are mapped to other periodic features in a straightforward, regular way. But close-up, these periodic features exhibit chaos within. In this chapter we study particular features like these which we call *islands*. Islands were first described in Góra et. al. [10] A family of two-dimensional maps with memory G_α, parameterized by α, was shown to give rise to interesting two-dimensional densities. The authors proved the existence of ACIMs for various ranges of α but also noticed an interesting phenomenon. These densities become progressively thinner as α approaches $\frac{1}{2}$. This progression can be seen in the leftmost and middle images of Figure 7.

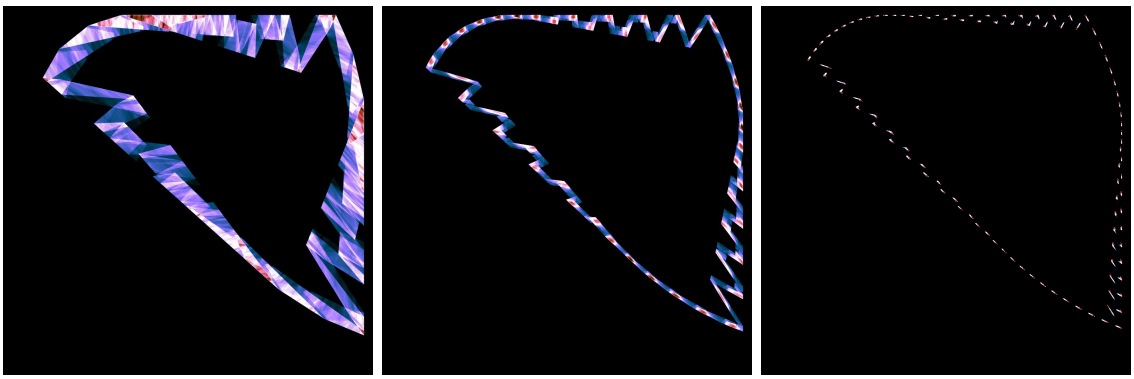

Figure 7: Densities of tent map with memory G for $\alpha = 0.45$, $\alpha = 0.471829$ and $\alpha = 0.48757$

Surprisingly, as α continues along this progression, certain values are reached for which the thinning density loses its connected support and fractures into small disjoint

two-dimensional regions, i.e., *islands*. For example, Figure 7 (right) is a disconnected density consisting of 100 small disjoint islands. Further increase in α usually results in a return to a single connected support for the density, albeit thinner, until another value of α is reached for which there are islands. Typically these special values of α for which there are islands, lie in very narrow ranges $\Delta\alpha$ of order 10^{-6} or 10^{-5} which seem to be island-forming intervals. The number of islands may be stable with perturbations of α in such an interval, or there may be several types of island orbits that are relatively stable in $\Delta\alpha$. This is what is seen numerically for the tent map with memory [10], in every case that we have examined.

The leftmost image of Figure 7 provides the far-away view where everything appears to be periodic and regular. Each island behaves as if it were a point that is mapped by G to another point-island. In fact, we can give each island a symbol and then study the symbolic trajectory of where they go under G. This symbolic trajectory of islands, which ignores finer details, is the regular behavior of the weak chaos system. We will construct and use such symbolic trajectories in this chapter.

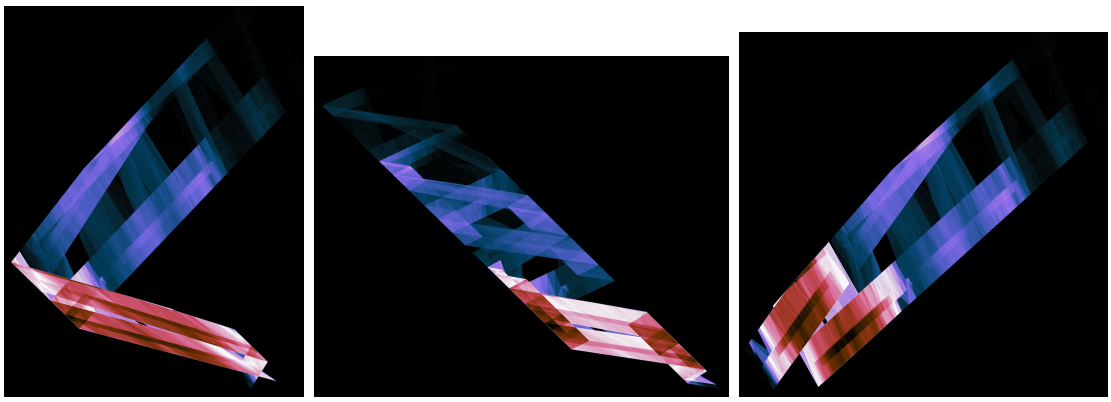

Figure 8: Islands for G with $\alpha = 0.48757$.

Zooming in closer on the tiny islands in the rightmost density of Figure 7, we obtain close-up images of individual islands, shown in Figure 8. Now we see remarkable structure and also chaos.

2.2 Computations

Most of the computations for this thesis were done on a modest notebook with a 1.6 GHz Intel i5 processor. Rendering images from histograms, making illustrations,

making diagrams of orbits or geometrical scenes with labels, proving the existence of ACIMs, etc., all these were done on the notebook computer.

However there were some computations that required more robust machines. Computations of entropy over wide ranges of α, (for example, Figures 12, 13 and 14) which required the construction of thousands of histograms—those were done on a desktop 2 GHz Xeon E5405, which happily ran those jobs for days on end. The most computationally intensive tasks involved collecting evidence for Proposition 2.3.1 and computing histograms for definitive high-resolution images of "island-0" (Figure 17 and in [9]) using 128-bit floating point arithmetic. Some of those computations took more than a week on a 3.5 GHz Xeon Gold 6144. All of these machines run Linux.

The hard-core number crunching components of these tasks were written in modern Fortran and compiled with GNU Fortran 10, which implements the Fortran 2008 standard along with much of the 2018 one. Some of the earliest computations pertaining to Proposition 2.3.1 were coded in C++, but it was decided early-on to port everything to modern Fortran. Modern Fortran and the GNU Fortran compiler are pleasant to work with: it takes some effort to write slow, unclear or incorrect code. Modern incarnations of Fortran offer convenient idioms by which one can express array computations in compact ways. For example, here is a "one-liner" that computes the entropy of a normalized histogram `h`:

```
-sum(pack(h, h > 0)*log(pack(h, h > 0)))/log(2.0_real64)
```

Programs written in Fortran were used to compute histograms for entropy plots, for the images of orbit densities, to map out the dimensions of islands, compute some of their properties, and then produce the histograms for high-resolution close-up images. Since speed is absolutely essential for these tasks, Python or another dynamic language would be unsuitable.

From the data computed about islands, another Fortran program attempts to prove the existence of ACIMs. Sometimes speed is necessary here too: we shall see there can be very many possibilities to check—too many for a program written in Python.

An interesting feature of modern Fortran compilers is the availability of 128-bit floating point numbers. Arithmetic with 128-bit floats is roughly 25× slower than regular 64-bit arithmetic on 64-bit architectures, but it's nice to have, and it was used to verify some of the 64-bit computations.

As an alternative to GNU `make`, we used the `FoBiS` Fortran build tool, available from GitHub. It is easy to use and greatly simplifies the build process for small or medium-sized Fortran projects.

Two-dimensional histograms computed by the Fortran programs are straightforward generalizations of the one-dimensional histograms of Section 1.1. Typically a two dimensional region is cut up into 1000×1000 bins, and the bins are populated with orbit point data, though for the highest resolution images this could be more like 3000×3000 bins. We chose the first way of constructing densities that was described in Section 1.1. That is, the method where one chooses a single x_0 and then applies the map repeatedly to generate a very long orbit from which the histogram is populated. The reason for this choice is that only the current position along the orbit need be kept in memory. The second method described in Section 1.1 was used to create the graphics in [19]. However, it requires that the computer remember the position of N points. If N is very large then this becomes inconvenient or even impossible. Delicate features in certain types of islands are sometimes only visible after billions of iterations. It is easier, memory-wise, to follow the evolution of 10 points for 10^9 iterations of G, than it is to follow 10^9 points for 10 iterations of the Frobenius-Perron operator. Using the first method, we are limited by computer time but not by computer memory. However, the second method of constructing densities was used to make the images of the polygonal densities in Figure 44. To produce these images, it was sufficient to take on the order of 10^6 initial points and follow them for 1500 iterations of G or less.

Histograms can contain bright spots, i.e., bins with very high values compared to the rest, possibly $10^4\times$ higher. When such histograms are rendered into images, the bright spots overpower other fainter details. this is what happens with the image of island-0 in Figure 17. The regions of highest density, near the apex of the triangular figure, are the only things visible in an image, if the image is not post-processed with an equalizing filter. The most useful equalization filters are simple functions applied to the values b in the bins: $\sqrt{b}$, $\sqrt[3]{b}$ and $\log(b)$. Island-0 was post-processed with $\log(b)$ equalization, which greatly reduces the relative peaks between high and low bins. The most radical form of equalization is binary: any bin with value $b > 0$ is set to 1 and the rest are kept at 0. This results in a two-toned image that is useful for examining the extents and outline of an island, but all other details are

lost. Sometimes histograms contain homogeneous distributions of values across bins, and it is difficult to discern details because there is not enough variation that can be translated into colors. In such cases an anti-equalization is applied, for example b^2, which increases the relative differences between bins.

Choosing a post-processing function and coloring scheme always involves trade-offs: some image features are enhanced, others toned down. But for each job of rendering an image from a histogram, we chose a post-processing equalization and particular color scheme which combined to reveal as much detail and subtle structure in islands with the best clarity we could manage.

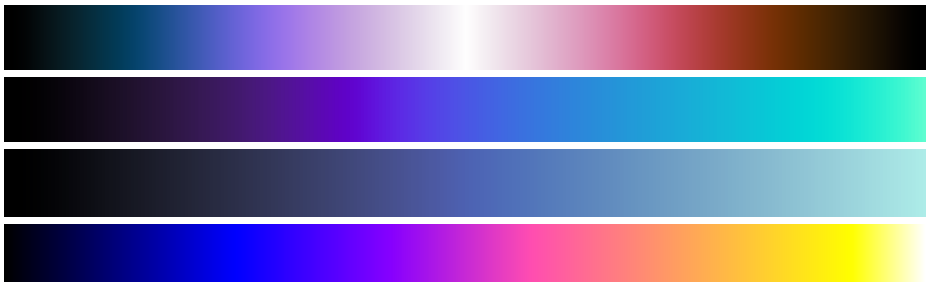

Figure 9: Color maps: `cmr.emergency` (top), `cmr.cosmic`, `cmr.sapphire` and `gnuplot2` (bottom)

The equalization computations were done with `numpy`, Python's numerical computations library. The image creation was handled by the Python plotting package `matplotlib.pyplot`, which, together with `numpy`, we also used to create most of the illustrations for Chapter 3. The final product is in the form of `png` images, some of which were adjusted for contrast using the Python `PIL` image processing library. For inclusion in this thesis, the `png` images were reduced in size by converting them to `jpeg` format, also using `PIL`.

The `matplotlib` plotting package comes with its own colormaps: some of these are good for imaging islands: `turbo`, `gnuplot2`, `cubehelix`, for example. In particular, `turbo` was the colormap used for island-0 in [9]. However, experience has shown that the islands look best when imaged against a solid black backround, which isn't the case for `turbo`. To extend the choice of colormaps which begin with black, we added some from the `cmasher` colormap library. The `cmasher` colormap `cmr.emergency`, shown in Figure 9, gives the very best results for the widest range of island images. This is a "diverging" colormap. That is, it is brightest at the center, and then moves off into red and deep red as densities get higher. It has the disadvantage of representing both

the highest and lowest densities as dark, but in practice, the `cmr.emergency` diverging colormap often gives the best images, at least when it comes to islands and densities. Sometimes other colormaps work better: `cmr.sapphire` was used for Figure 35 and Figure 36 (left), while `cmr.cosmic` was used in Figure 36 (right). A balance has to be reached, combining colormap with histogram equalization to produce the best image.

For verifying algebra computations, especially those of Chapter 3, we used the freely available `maxima` computer algebra system.

2.3 Tent map with memory

The recursion relation $f(n) = 2f(n-1)$ together with initial condition $f(1) = 1$ has a solution in the form of a sequence

$$1,\ 2,\ 4,\ 8,\ \ldots$$

or in closed-form as $f(n) = 2^{n-1}$. Note that the present value $f(n)$ of the sequence depends on the previous value $f(n-1)$. Higher-order recursions can be constructed from more than one previous value. For example, Fibonacci numbers are generated by the recursion $f(n) = f(n)+f(n-1)$ with *two* initial conditions $f(1) = 1$, $f(2) = 1$. The rest of the sequence follows by applying iterations of f:

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ \ldots$$

The first recursion is analogous to τ from Section 1.1. Since $f : \mathbb{N} \to \mathbb{N}$, one can think of f as a dynamical system map on the natural numbers, generating an orbit which are the powers of two. This orbit is analogous to the τ-orbit in (2).

What would be the dynamical system analogy to a higher recursion like the Fibonacci relationship? Beginning with two initial conditions x_0, x_1 in I, the next step could be $x_2 = T(x_0, x_1)$. However, this T, in general, maps $I \times I \to \mathbb{R}$, so it cannot be a dynamical system map because the domain and range are not the same. Following a standard trick from physics where a second-order equation is reduced to two first order equations in two-dimensional phase space by introducting $y = \dot{x}$ as the second variable, let $x = x_{n-1}$ and $y = x_n$. Then the T-recursion can be reformulated as an equivalent two dimensional function $G : (x, y) \to (y, T(x, y))$. It is necessary that $T(x, y) \in I$. This can be guaranteed if we take a convex combination of x and y, and

feed that to a dynamical system map $\tau : I \to I$. A convex combination $\alpha y + (1-\alpha)x$ is always in I for $0 \leq \alpha \leq 1$, and thus $T(x, y) = \tau(\alpha y + (1-\alpha)x)$ is always in I. With this $T(x, y)$, G becomes a map from $I \times I$ to $I \times I$, and we have a two-dimensional dynamical system with map G acting on $I \times I$:

$$G(x, y) = (y, \tau(\alpha y + (1-\alpha)x)) \,. \tag{9}$$

In the papers [8, 2016] and [10, 2017], Góra et. al. introduced $T : I \times I \to I$ and $G : I \times I \to I \times I$, calling them *maps with memory*.

A particular map $\tau : I \to I$ which provides many examples in the literature of dynamical systems is the *tent map*

$$\tau(u) = \begin{cases} 2u & u < \frac{1}{2} \\ 2(1-u) & u \geq \frac{1}{2}. \end{cases} \,. \tag{10}$$

A plot of the tent map is shown in Figure 10. Following the construction (9) but

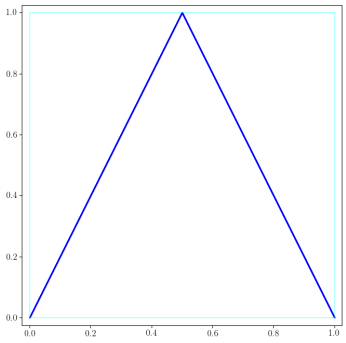

Figure 10: Tent map $\tau(x) : I \to I$.

using the tent map τ, we obtain $G(x, y)$, the two dimensional *tent map with memory*,

$$G(x, y) = (y, \tau(\alpha y + (1-\alpha)x)) = \begin{cases} (y,\, 2u) & u < \frac{1}{2} \\ (y,\, 2(1-u)) & u \geq \frac{1}{2} \end{cases} \tag{11}$$

where $u = \alpha y + (1-\alpha)x$ and $0 < \alpha < 1$. The line

$$(1-\alpha)x + \alpha y = \frac{1}{2} \tag{12}$$

which we call the *partition line*, partitions $I \times I$ into two regions: A_1 (below the partition line) and A_2 (above the partition line). Examples of partition lines for various α are shown in Figure 11. Note that when $\alpha \to 0$, the partition line becomes

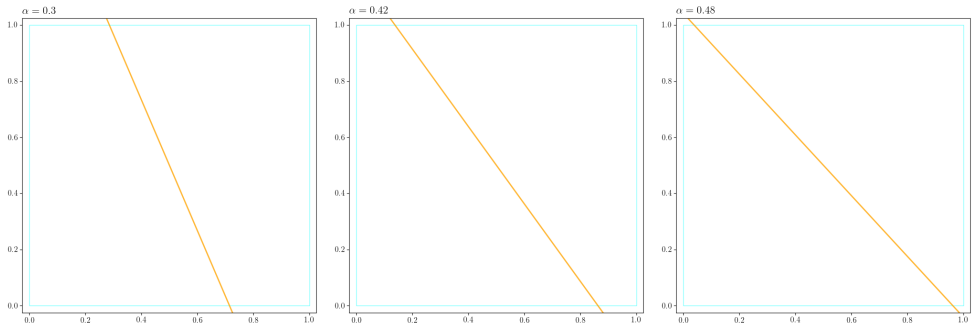

Figure 11: Partition lines for $\alpha = 0.3$, $\alpha = 0.42$ and $\alpha = 0.48$.

the vertical line $x = \frac{1}{2}$, when $\alpha \to \frac{1}{2}$ it approaches the anti-diagonal line of $I \times I$, and when $\alpha \to 1$ the partition line becomes the horizontal line $y = \frac{1}{2}$.

G is piecewise linear on the partition $\{A_1, A_2\}$, so from (11) we may define G_1 acting on A_1 and G_2 acting on A_2 as

$$\begin{aligned} G_1(x,y) &= \begin{pmatrix} 0 & 1 \\ 2(1-\alpha) & 2\alpha \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \quad u < \frac{1}{2} \\ G_2(x,y) &= \begin{pmatrix} 0 & 1 \\ -2(1-\alpha) & -2\alpha \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad u \geq \frac{1}{2}. \end{aligned} \tag{13}$$

where $u = \alpha y + (1-\alpha)x$. The derivative matrices of G in regions A_1 and A_2 can be read off from (13):

$$D_1 = D(G_1) = \begin{pmatrix} 0 & 1 \\ 2(1-\alpha) & 2\alpha \end{pmatrix}, \quad D_2 = D(G_2) = \begin{pmatrix} 0 & 1 \\ -2(1-\alpha) & -2\alpha \end{pmatrix} \tag{14}$$

Note that in region A_1, $D(G)$ is a constant matrix with respect to x, y and likewise in region A_2.

In [10], Góra et. al. studied the tent memory map (11). As α ranges from 0 to 1, G shows different kinds of dynamical behaviors such as chaotic attractors, a global point attractor and periodic attractors. The existence of ACIMs and singular invariant measures were proved for various intervals of α. There is a gradual thinning of the support for the ACIM, and then at $\alpha \to \frac{1}{2}$ there is a sudden catastrophe: the dynamics changes over to a global fixed point attractor at $(x, y) = \left(\frac{2}{3}, \frac{2}{3}\right)$. This persists for $\frac{1}{2} < \alpha < \frac{3}{4}$.

Just before the catastrophe at $\alpha = \frac{1}{2}$, the authors noticed that some values of α near $\frac{1}{2}$ produced densities that looked very different from what was regarded as

typical. At $\alpha = 0.483$ the support for a conjectured ACIM density is disconnected: it consists of 175 tiny disjoint two-dimensional regions of $I \times I$. These objects have come to be known as *islands*. This island structure was seen to persist over a narrow window of about 10^{-6}. Likewise α values of 0.4883, 0.4943 and 0.4973 produced density with supports on 106, 214 and 448 islands, respectively.

We will examine all of these α values, but first we consider the question of how to discover these values. It isn't so easy, because as mentioned above, they seem to exist in very narrow windows or intervals of island-formation. The authors of [10] conjectured that there are many more α-windows where islands exist. This turned out to be true, and a practical tool for the discovery of more of these α-windows of island formation is entropy.

Entropy is calculated directly from the histogram before it is rendered into an image. We use a standard partition-entropy definition, such as in [4, ch.8]. First the $m \times n$ histogram is normalized to produce a density, then the bins containing zero probability density are ignored. The entropy of the density is then

$$-\sum_{i=1,\, p_i \neq 0}^{mn} p_i \log_2(p_i).$$

Fortran code to compute entropy of a normalized histogram was showcased in Section 2.2. For entropy computations, histograms were fixed at 1000×1000 bins, and one histogram was computed for each of 5000 values of $0 < \alpha < 1$. Figure 12 shows the resulting entropy plot.

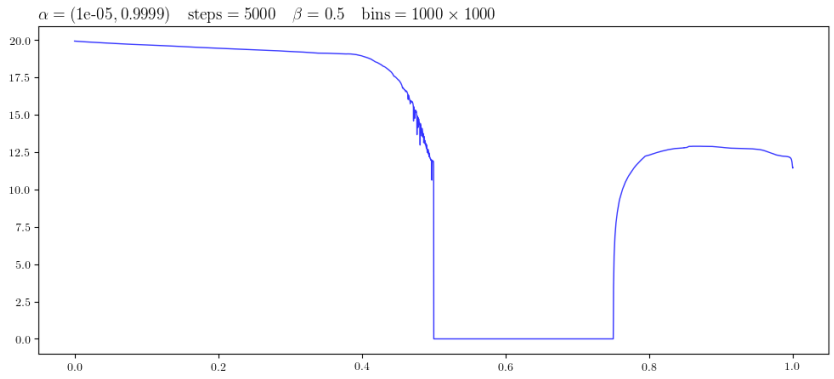

Figure 12: Entropy plot for the full range $0 < \alpha < 1$.

The qualitative changes in the entropy plot mirrors the changes in the dynamics of G as described in [10]. As α approaches approximately 0.4, there is a gentle

decrease in entropy from about 20 to 18. This corresponds to the shrinking of the attractor—it starts off covering the majority of the $I \times I$ region, but slowly collects into mostly region A_2, above the partition line. At about $\alpha = 0.4$ there is a much sharper decrease in entropy, this corresponds to what is seen in images: this is the region of α where the support for the density becomes very thin. At $\alpha = \frac{1}{2}$ there is a catastrophe: everything except $(0, 0)$ is attracted to the global fixed point at $(\frac{2}{3}, \frac{2}{3})$ in A_2. The entropy drops to near-zero and stays that way for as long as the global attractor exists.

A peculiar detail is visible in Figure 12. Just before the point of catastrophe at $\alpha = \frac{1}{2}$, the plot no longer looks smooth. For a small interval just prior to $\alpha = \frac{1}{2}$, the plot exhibits anomalous downward spikes. This critical region is roughly $0.470 < \alpha < 0.498$, and it is here that we find tiny windows of α where islands form.

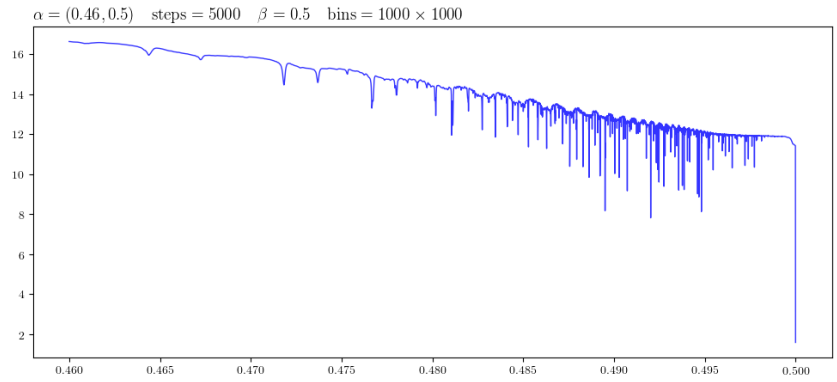

Figure 13: Entropy for $0.46 < \alpha < 0.5$

Figure 13 is a zoom on $0.46 < \alpha < 0.5$. This figure is another computation involving 5000 histograms.

Zooming in further, again with a resolution of 5000 histograms, we have Figure 14. All the island-forming α values mentioned in [10], and many more, are found here. It is tempting to conjecture that the sudden, sharp downward drops in entropy are associated with the formation of islands. This turns out to be partly true, or rather, true enough. Every case of island formation that we have looked at is associated with a sharp drop in entropy. But not every sharp drop in entropy indicates islands. Often it does, but it also gives false positives. Nevertheless, it is good enough. We search through the thousands of points of entropy data for minima—values of α for which the entropy has suddenly dropped. Then images are made with these α, and

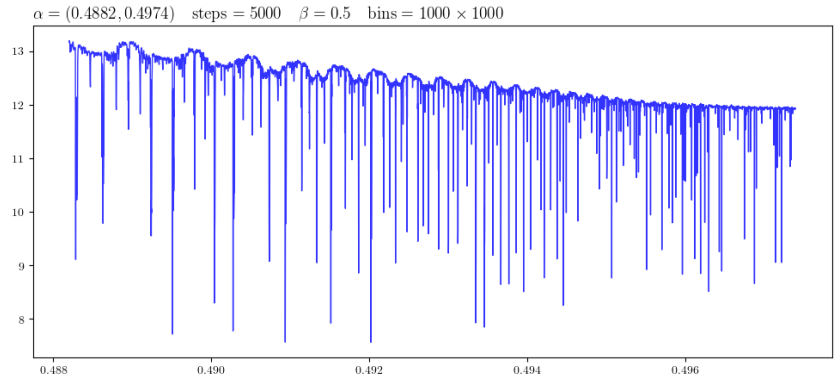

Figure 14: Entropy for $0.4882 < \alpha < 0.4974$.

the images reveal the presence of islands. Entropy is a good qualitative indicator of what is going on, broadly, with the dynamics of G. Using entropy plots, we have discovered many new values of α for which there exist islands supporting ACIMs.

In Chapter 4 we will consider another method of broadly describing the behavior of 2D memory maps over ranges of α and how to use it to discover regions of island formation.

The principal example of islands presented in [10] and then studied further in the paper [9], is the system of 175 islands at $\alpha = 0.493$. This value of alpha corresponds to one of the entropy drops in the critical inteval of Figure 14. An entropy plot of a very thin neighborhood of approximate width 10^{-5} around 0.493 is shown in Figure 15. This *entropy well* is typical for all tent map and β-tent map island systems that we will encounter. There is a sharp drop in entropy, followed by a rise, and again a drop, into a kind of plateau. The same pattern is seen in other entropy wells. The width of the entropy well in Figure 15 is approximately 6.5×10^{-6}. The red interval has a with of $\Delta\alpha \sim 3.2161 \times 10^{-7}$ with extents

$$0.4929999849246231 \leq \alpha_{\text{red}} \leq 0.4930003065326633.$$

The green interval has a width of $\Delta\alpha \sim 2.01 \times 10^{-6}$ with extents

$$0.49300309045230 \leq \alpha_{\text{green}} \leq 0.4930051005025125.$$

From the red interval, 100 values of α were chosen and their densities all revealed 175 islands. It is the same for the green interval: density images for all 100 α samples chosen revealed systems of 175 islands. It seems then, that 175 is a stable number of

islands across the red and green intervals. The value $\alpha = 0.493$ belongs to the red interval.

For the $\alpha = 0.493$ system, numerical studies were done using both 64-bit and 128-bit floating point arithmetic. Islands were examined for at least 4 trillion iterations of G. The islands maintained their shape, character and individuality. They remained disjoint. They did not diffuse or spread. They remained in the same bounding boxes that were intitially computed for them with less than a billion iterations of G. Nor was it the case that islands were bright spots of a fainter connected underlying density: we ruled this out by going to extremes with iterations. The only difference that higher iterations made was to bring out finer details in the islands images, such as the triangular umbra or shadow of the island in Figure 17. These tests of island existence for $\alpha = 0.493$ were the most numerically intensive computations that we did. Some of these computer experiments took a week or more.

The process of computing images of islands and other properties such as region index (defined shortly) involves computing bounding boxes for islands and also mapping one island onto another to obtain complete sets of data for all islands. This was done many times for a large number of examples and different types of G with different underlying τ. All of these computations support the same conclusions about island existence as above, with added evidence that islands are *coherent*, i.e., all the points of one island are mapped by G to another island, and if there are n islands then G^n is an map from an island onto itself. Therefore we are confident in proposing the following.

Proposition 2.3.1 *Computer calculations establish the existence of islands. Islands are small regions of support in the plane. They remain disjoint and intact, maintaining the same general shape, size and character with applications of G. Islands are coherent: G maps the entirety of one island onto another, and if there are n islands then G^n maps an island to itself.*

Once this proposition is accepted, the main result of this chapter, published in [9], follows.

Theorem 2.3.2 *When $\alpha = 0.493$, the two-dimensional tent map with memory G admits an ACIM supported on* 175 *islands.*

Proof If we accept Proposition 2.3.1 then the theorem follows by a combinatorial argument. A schematic for the complete orbit with labelled islands is presented Figure 16. The islands themselves are, for $\alpha = 0.493$, too small to be seen—they are represented by dots. At this distance from the underlying orbit details, all we see is a periodic "trajectory" of island labels **0**, **1**, **2**, up to **174**, where G maps island **0** to island **1**, and then to **2** and so on until **174** is mapped back to **0**. The label **0** is given to the island near the upper-right corner of the $I \times I$ box, at the beginning of the long arc of islands going down to the lower-right. It will be possible to identify the **0** island in other trajectory diagrams that will be presented for other systems. Otherwise, the choice of island **0** location is arbitrary.

The orange line in Figure 16 is the partition line (12). Each island can be assigned a *region index* depending on its position with respect to the partition line and the regions A_1, A_2. An island has region index `1` if it lies completely in A_1, below the partition line. If it lies completely in A_2, above the partition line, then it has region index `2`. If an island is partly in A_1 and partly in A_2, then it has region index `3`. Region indexes are denoted in Figure 16 by colors of the island dots. Green means region index `1`. Red means region index `2` and orange means region index `3`. In this manner, the region indexes of the islands in the trajectory of any island can be assembled into their own symbolic trajectory. The region index trajectory of island **0** under iterations of G is

```
221222221223221222221223221222221223221222221223221
222221223221222221223221222221223221222221223221222
221223221222221223223222222222222222222222222222222
22222222222222222223
```
. (15)

G is piecewise linear on the partition, so there are only two possible derivative matrices, (14). Region index `1` corresponds to D_1 and region index `2` corresponds to D_2. Region index `3` could be either D_1 or D_2. Suppose we make a choice of D_1 or D_2 for each of the indexes `3` in (15). By using Lemma 1.3.4, this possible value for $D(G^{175})$ can be computed by simply taking a string of derivative matrices corresponding to region indexes and forming the product. If the smaller singular value σ_2 is then computed over all possible such $D(G^{175})$ expressions, and the minimum is shown to be greater than 1, then by Proposition 1.3.2 we will have established that G^{175} is an expanding map.

For $\alpha = 0.493$ there are 12 islands with region index `3`. Thus there are $2^{12} = 4096$

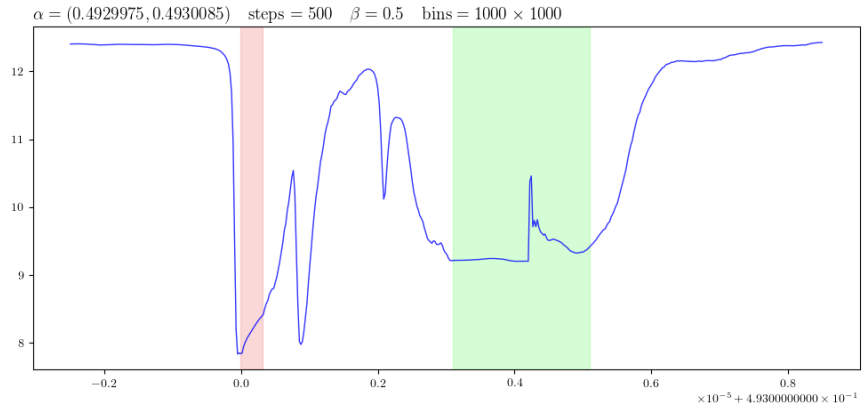

Figure 15: Entropy well and island formation intervals. Sudden drop in entropy beginning near $\alpha \sim 0.493$. Number of islands remains stable across red and green intervals.

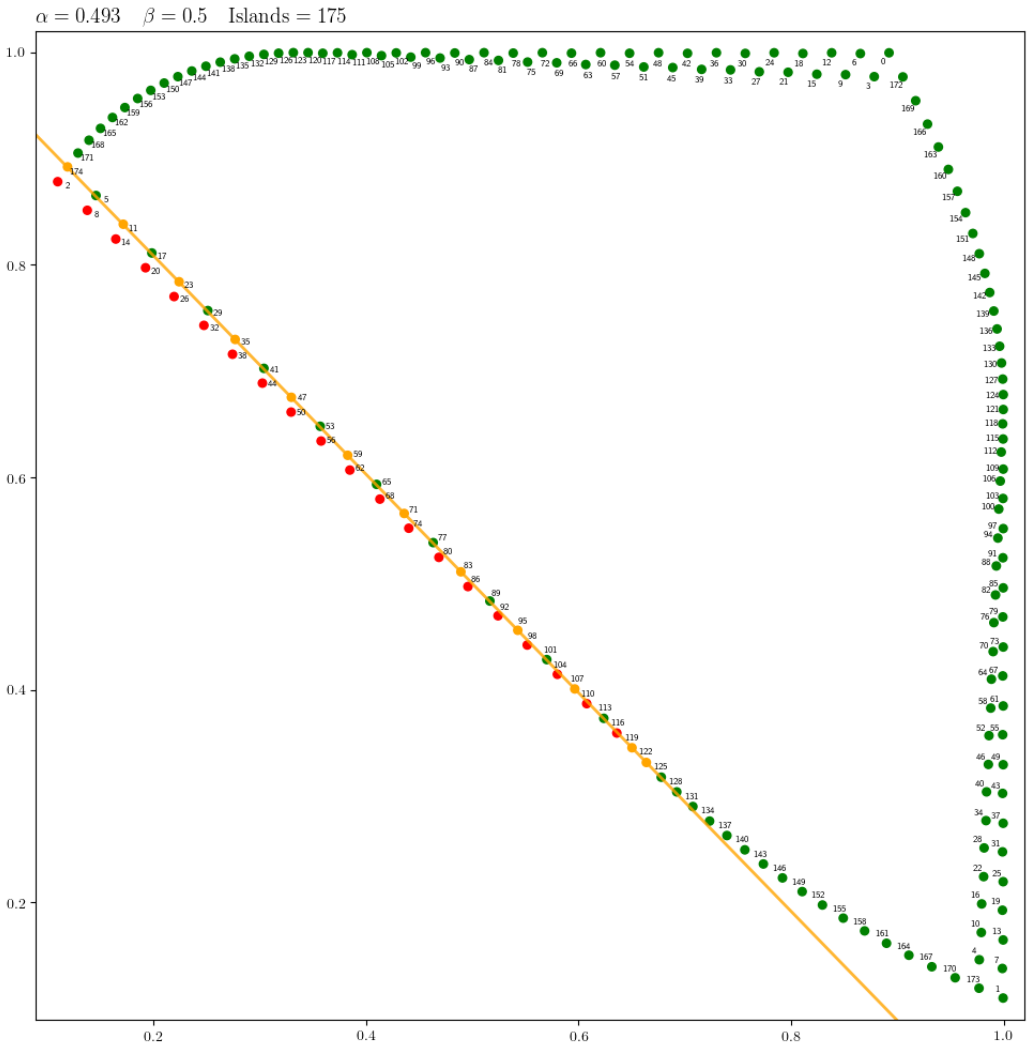

Figure 16: Region index and island trajectory for $\alpha = 0.493$: 175 island system.

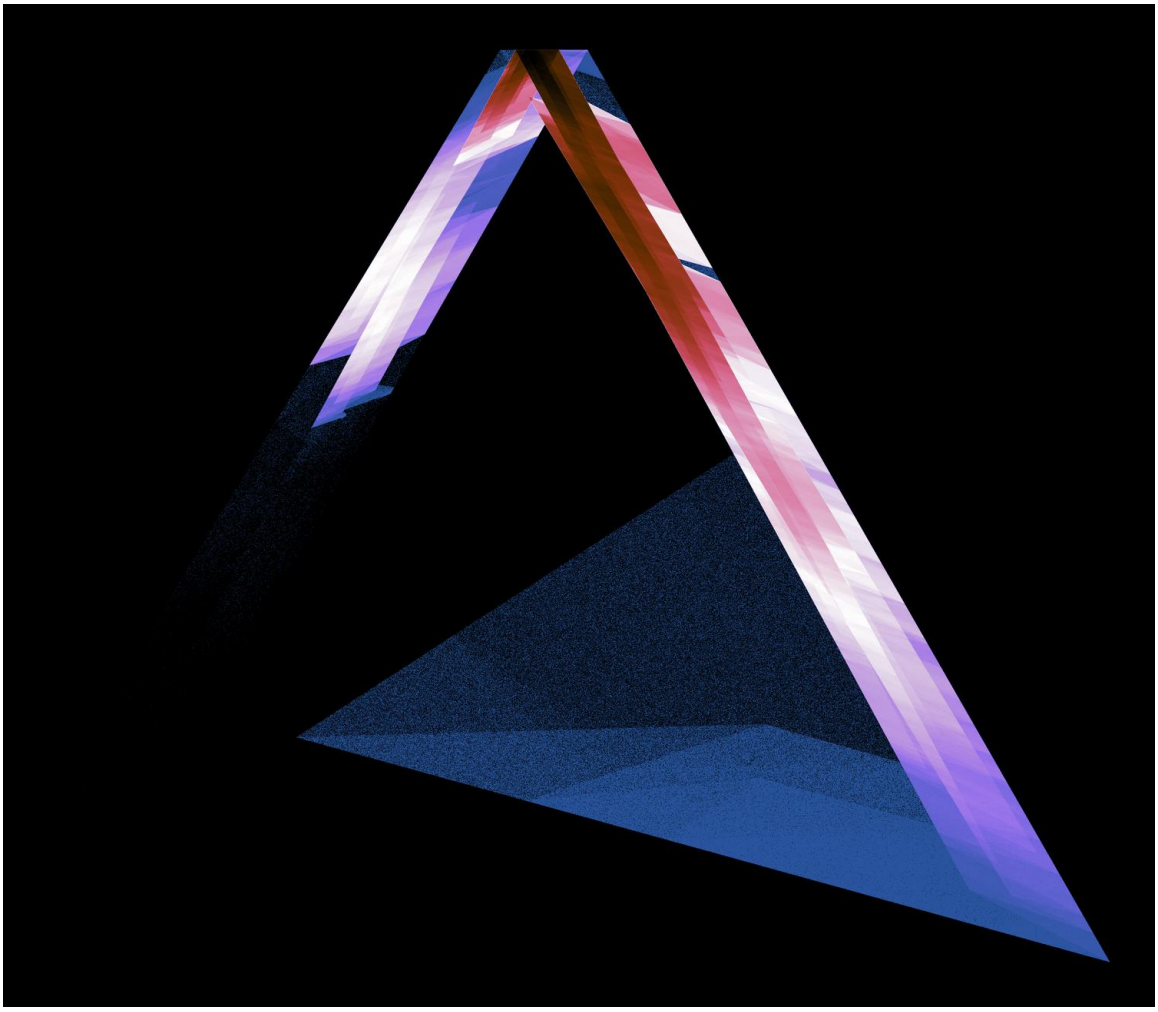

Figure 17: Island **0**. Image computed from 1.05 trillion iterations of G using 128-bit arithmetic.

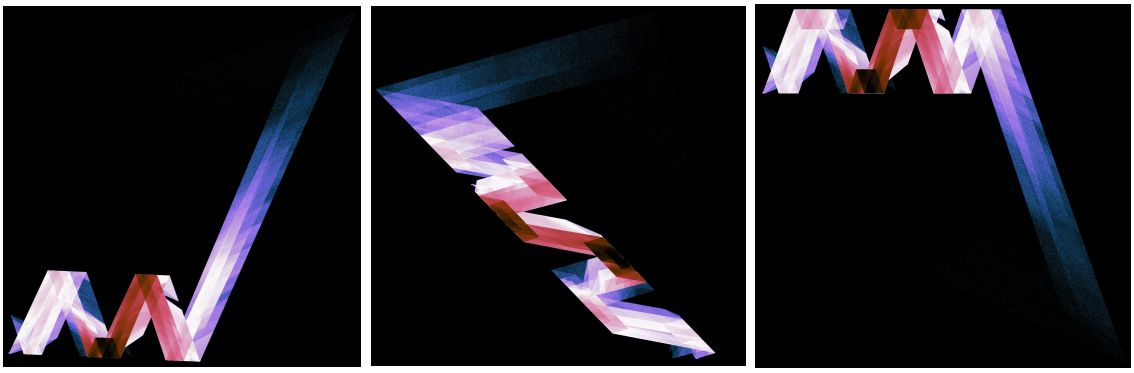

Figure 18: Close-ups of islands **27**, **71** and **65** in the $\alpha = 0.493$ system of 175 islands.

possible region index trajectories to check, and 4096 singular values to compute. Using the σ_2 computation method of Section 1.3, we find that the minimum σ_2 over all 4096 possible region index trajectories is 1.154116744823.

G is defined on a real-analytic partition and it is piecewise linear, which means it is also piecewise real-analytic. The composition G^{175} is therefore also piecewise real-analytic on some real-analytic neighborhood of island $\mathbf{0}$. Since the minimum σ_2 is greater than 1, G^{175} is piecewise expanding on this real-analytic neighborhood. Therefore by invoking Theorem 1.3.1 (Tsujii-Buzzi), G^{175} admits an ACIM supported on island $\mathbf{0}$, and likewise on every other island. Taking X in Lemma 1.3.5, to be the union of all the islands, we have that G itself admits an ACIM supported on all the islands.

Note that it is not necessary that G^{175} be expanding on *all* islands. Expanding on *one* island is sufficient for the construction of Lemma 1.3.5 to go through.

Figure 17 is the definitive image of island $\mathbf{0}$. It was created from a 3000×3000 histogram and 6 billion iterations of G^{175}—a total of 1.05 trillion iterations of G—all with 128-bit arithmetic. The `cmr.emergency` colormap was used, and log-equalization was applied to the histogram. This image contains finer details (e.g. the shadow) which are not clearly visible in the images of Figure 18 without resorting to binary image equalization and binary colormaps.

Now we examine two values of α in the green interval of the entropy well in Figure 15. The island trajectory diagrams will be omitted, since they are similar to the one in Figure 16, but it is interesting to see what the islands look like.

At $\alpha = 0.493003157454$ there are 175 islands, 22 of which have region index `3`. Island $\mathbf{0}$ for this system has region index trajectory

```
221223221223221223221223221223221223221223221223221
223221223221223221223221223221223221223221223221223
221223221223221223223222222222222222222222222222222 .
22222222222222222223
```

There are $2^{22} = 4{,}194{,}304$ possible matrix products to examine. The minimum σ_2 computed over all of them is 1.136615387348. Therefore by the same arguments as those of Theorem 2.3.2, G admits an ACIM supported on these islands.

When $\alpha = 0.493004028476$ there are again 175 islands, 18 of which have region

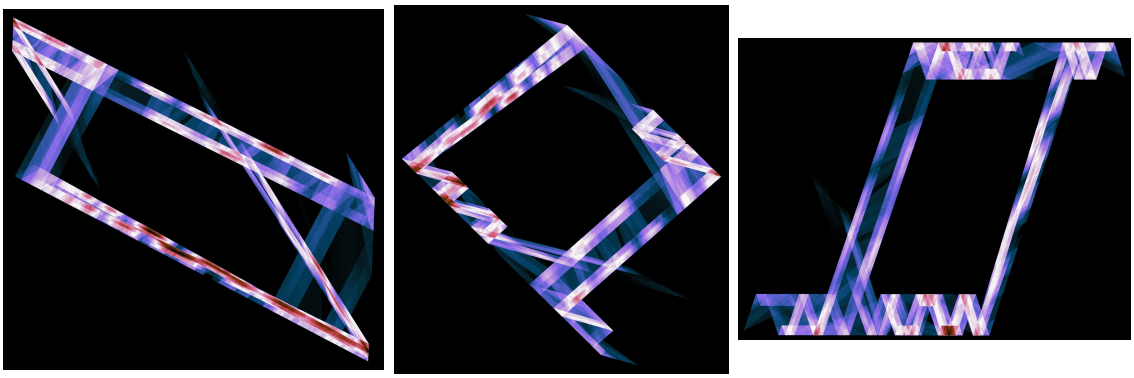

Figure 19: System of 175 islands in green interval. Close-ups of islands 10, 26 and 42. $\alpha = 0.493003157454$.

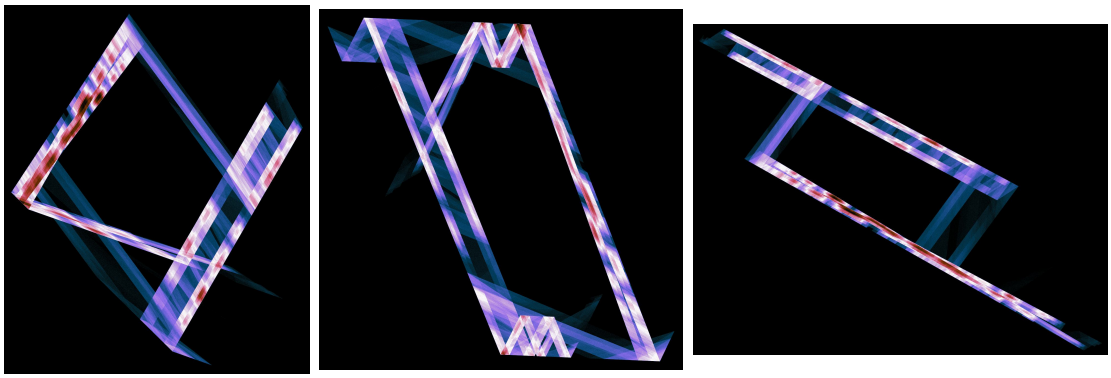

Figure 20: Close-ups of islands 2, 21, 161 in green interval system. $\alpha = 0.493003157454$.

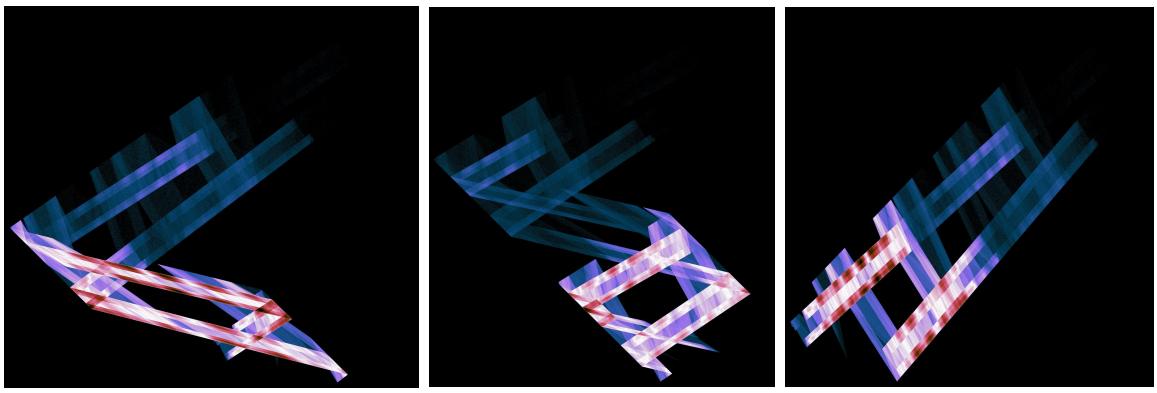

Figure 21: Islands 32, 38 and 168 in another green interval system of 175 islands. $\alpha = 0.493004028476$.

index `0`. Island **0** has region index trajectory

```
221221221221221221221223221223221223221223221223221
223221223221223221223221223221223221223221223221223
221223221223221223223222222222222222222222222222222 .
2222222222222222222222.
```

The smaller singular value for the $2^{18} = 262{,}144$ possible matrix products are computed, and the minimum is found to be 1.136523859914. Thus, G for this value of α admits an ACIM supported on these islands. Images of islands the first green interval system are shown in Figures 19 and 20. Some islands in the second green interval example are shown in Figure 21.

What is the smallest value of α for which there are islands? It is unknown if there is a definitive answer. However, the smallest island-producing α which we were able to find by searching through entropy data was $\alpha = 0.48757$. Figure 22 shows the associated entropy well. This α gives rise to a system of 100 islands with 14 of them having index `3`. The island trajectory is shown in Figure 23. One hundred islands is not that many—it results in a clearer plot of the island trajectory with less density of details. The region index trajectory for island **0** is

```
221223221223221223221223221223221223221223221223221
2232212232212232232232222222222222222222222222223   .
```

There are $2^{14} = 16{,}384$ σ_2 singular values to compute. The minimum of these is 1.069230611395, therefore G admits an ACIM on these islands. Images of these islands are found in Figures 26 and 27.

The paper [10] mentions a system of 106 islands at $\alpha = 0.4883$. We will examine this presently. This value of α lies at the lower end of the critical entropy region of Figure 14. It's interesting that small α tend to produce densities with relatively fewer islands. The entropy well near $\alpha = 0.4883$ is shown in Figure 24 and the island trajectory diagram in Figure 25. There are 16 islands of region index `3` and therefore $2^{16} = 65536$ possibilities to check. Doing so for island **0** results in a minimum σ_2 of 0.991107643093, which is inconclusive. However, we can get what we need by looking at island **1**. The region index trajectory for island **1** is

```
212212212232212232212232212232212232212232212232232
232223222322232223222322232222222222222222222222222 .
2222
```

The minimum σ_2 over all possibilities is 1.452561553082 and thus an ACIM exists.

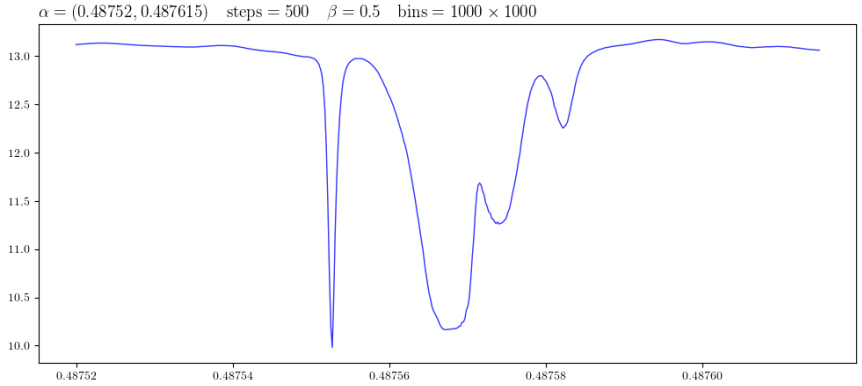

Figure 22: Entropy well around $\alpha \sim 0.48757$.

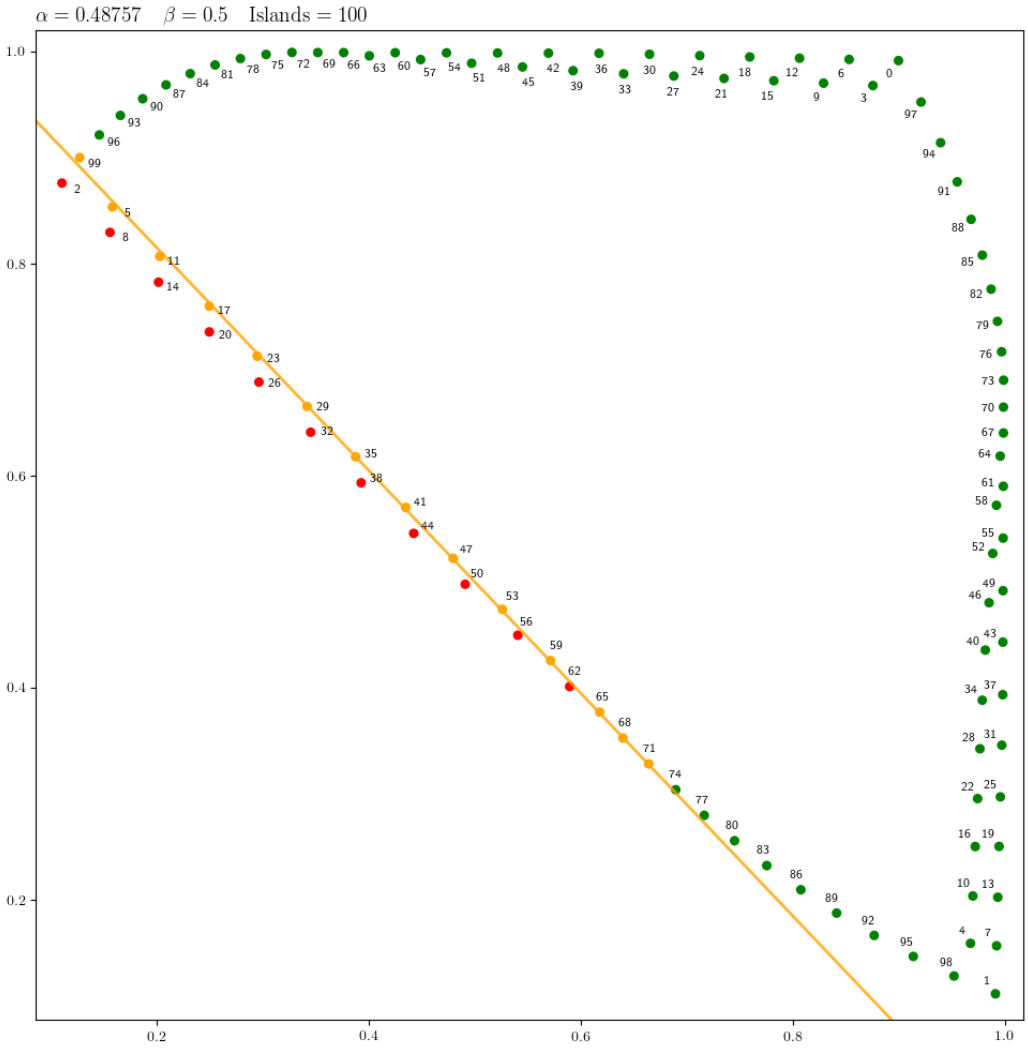

Figure 23: Island trajectory for $\alpha = 0.48757$: system of 100 islands.

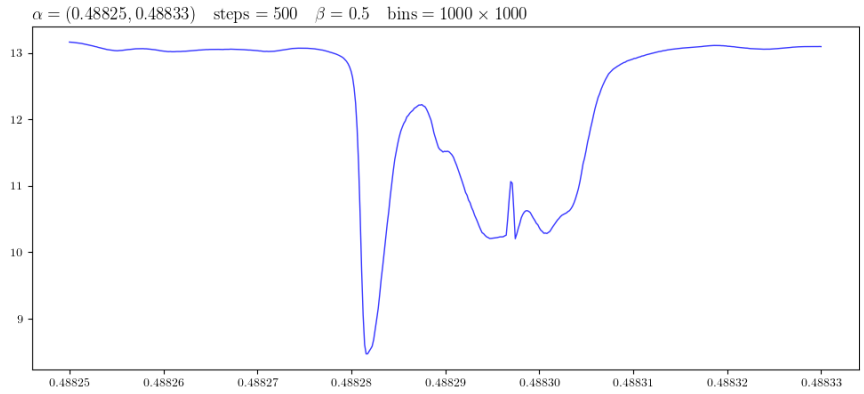

Figure 24: Entropy well around $\alpha \sim 0.4883$.

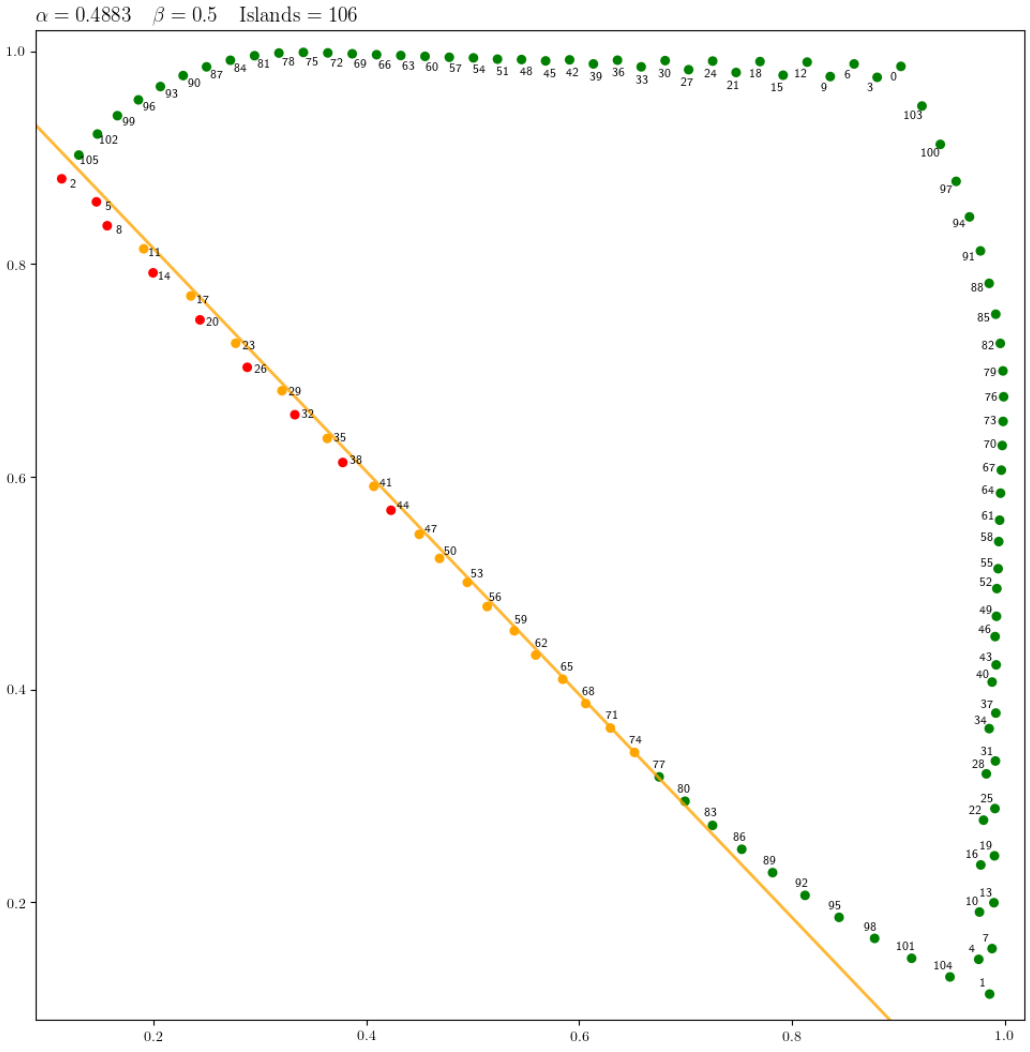

Figure 25: Island trajectory for $\alpha = 0.4883$: system of 106 islands.

Figure 26: Islands **20**, **21** and **41** in the 100 island system at $\alpha = 0.48757$.

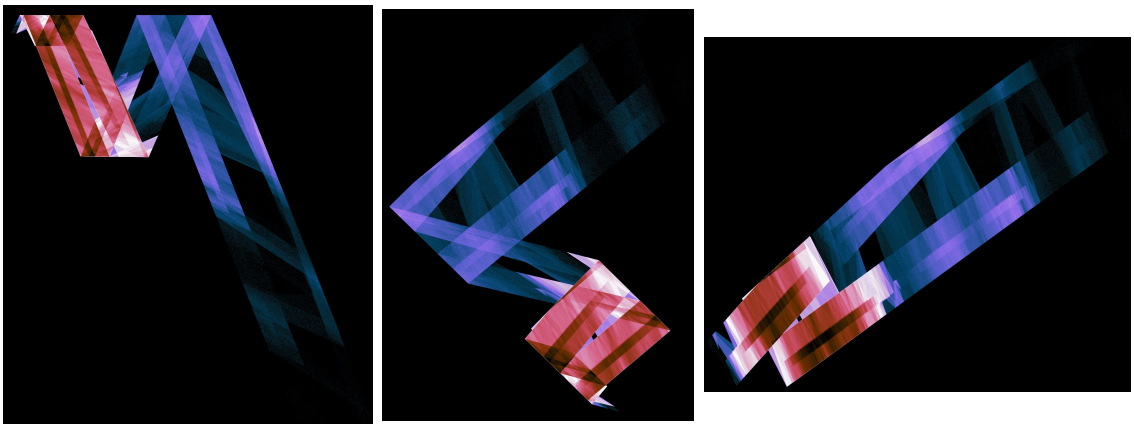

Figure 27: Islands **12**, **14** and **93** in the 100 island system at $\alpha = 0.48757$.

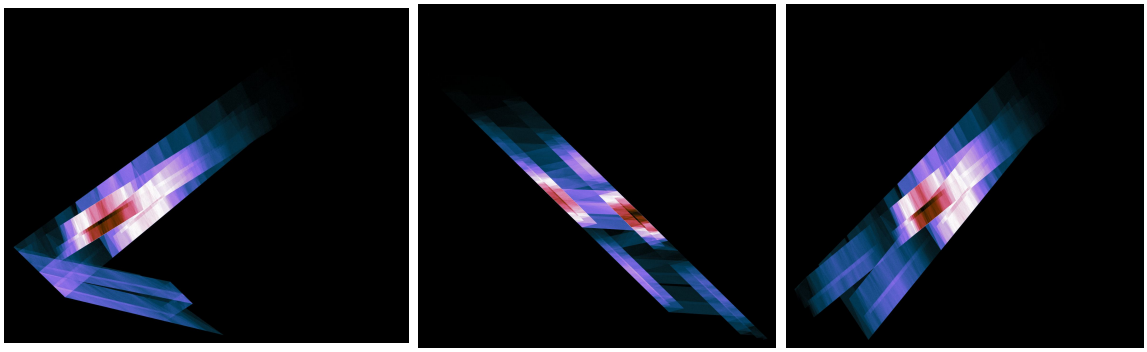

Figure 28: Close-ups of islands **20**, **68** and **102** in the 106 island system at $\alpha = 0.4883$.

One final example from the family of tent maps with memory G, taken from the paper [10]. The claim is made that there is a system of 214 islands at $\alpha = 0.4943$. Figure 29 shows the entropy well around α. There are 26 islands of region index `3`. This means $2^{26} = 67{,}108{,}864$ possibilities to examine, which is much more computation than the previous examples. Luckily we do not have to look very far to find a good island. The symbolic region index trajectory for island **0** is

```
22122322122322122322122322122322122322122322122321
22322122322122322122322122322122322122322122322123
22122322122322122322122322122322122322122322322222 .
22222222222222222222222222222222222222222222222222
2222222222
```

The minimum σ_2 over all possibilities is 1.113417236455, and therefore G admits an ACIM on these 214 islands. Figures 30 and 31 showcase the remarkable appearance of these islands.

2.4 β-tent map with memory

Section 2.3 introduced the tent map with memory. In this section we propose a generalization. The modified or skew tent map τ_β is a generalization of (10). Instead of peaking at $\frac{1}{2}$, τ_β peaks at any $0 < \beta < 1$:

$$\tau_\beta(u) = \begin{cases} u/\beta & u < \beta \\ \dfrac{1-u}{1-\beta} & u \geq \beta. \end{cases}$$

Plots of τ_β for various β are shown in Figure 32.

When $\beta = \frac{1}{2}$, $\tau_\beta(u)$ becomes the usual tent map $\tau(u)$. Using τ_β, a two-dimensional dynamical system map $G_\beta : I \times I \to I \times I$ is constructed in the same way as in Section 2.3:

$$G_\beta(x,y) = (y, \tau_\beta(\alpha y + (1-\alpha)x)) = \begin{cases} (y,\, u/\beta) & u < \beta \\ \left(y,\, \dfrac{1-u}{1-\beta}\right) & u \geq \beta \end{cases}$$

where $u = \alpha y + (1-\alpha)x$ and $0 < \alpha < 1$. We will call this the β-tent map with memory. Analogously to 12, there is a partition line which cuts $I \times I$ into two regions, A_1 and A_2, except that now the the partition line depends on β:

$$(1-\alpha)x + \alpha y = \beta. \tag{16}$$

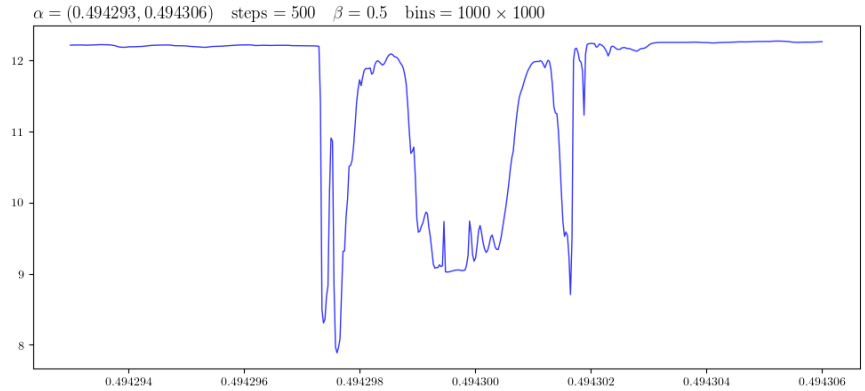

Figure 29: Entropy well around $\alpha \sim 0.4943$.

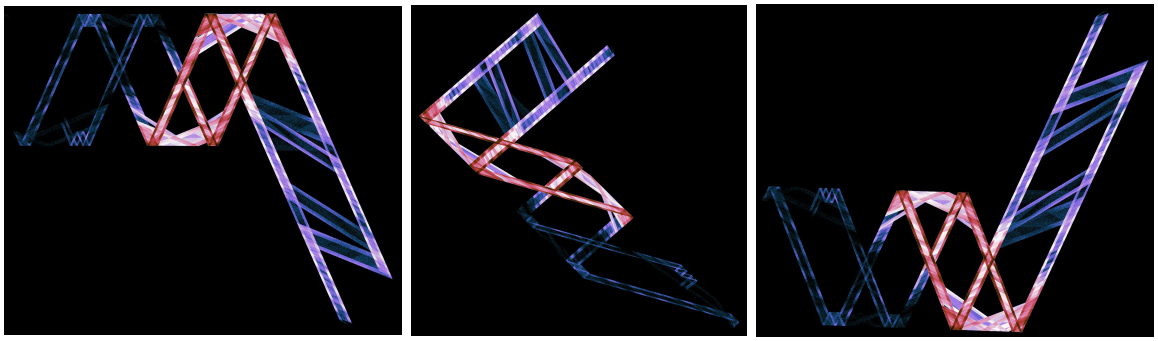

Figure 30: Islands **24**, **26** and **27** in the 214 island system at $\alpha = 0.4943$.

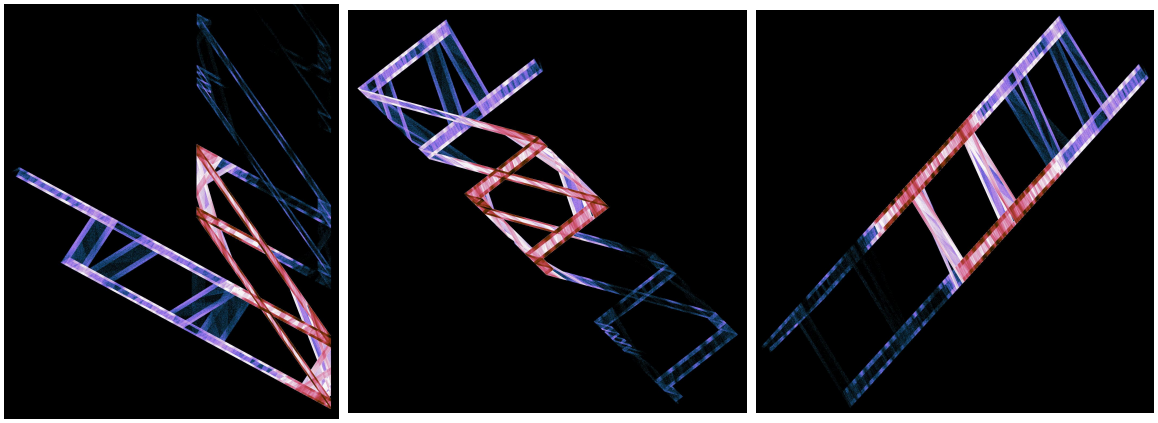

Figure 31: Close-ups of islands **25**, **32** and **207** in the 214 island system at $\alpha = 0.4943$.

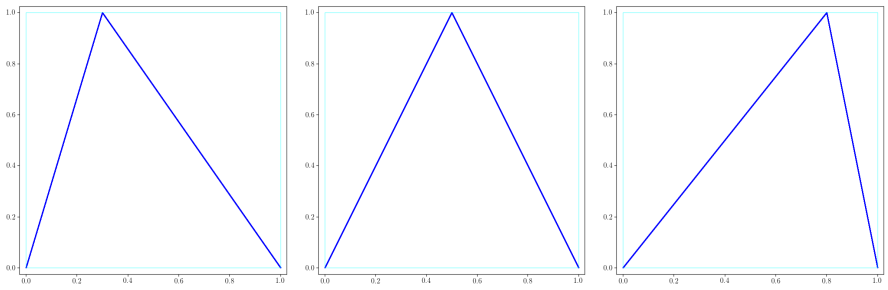

Figure 32: β-tent map $\tau(x)$ for $\beta = 0.3$, $\beta = 0.5$ and $\beta = 0.8$.

As with the tent memory map, G_β is piecewise linear on A_1, A_2. To simplify the notation, let G refer to G_β in this section—it being understood that G is defined for some β and α. The general practice will be to use the symbol G to refer to the 2D dynamical system currently being discussed. In matrix form, G_1 on A_1 and G_2 on A_2 are

$$\begin{aligned} G_1(x,y) &= \begin{pmatrix} 0 & 1 \\ \frac{1-\alpha}{\beta} & \frac{\alpha}{\beta} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}, \quad u < \beta \\ G_2(x,y) &= \begin{pmatrix} 0 & 1 \\ \frac{\alpha-1}{1-\beta} & \frac{\alpha}{\beta-1} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{1-\beta} \end{pmatrix}, \quad u \geq \beta. \end{aligned} \tag{17}$$

In regions A_1 and A_2, G has constant determinant matrices $D_1 = DG_1$ and $D_2 = DG_2$. These can be read from (17):

$$D_1 = \begin{pmatrix} 0 & 1 \\ \frac{1-\alpha}{\beta} & \frac{\alpha}{\beta} \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 1 \\ \frac{\alpha-1}{1-\beta} & \frac{\alpha}{\beta-1} \end{pmatrix}.$$

Now we will show that for a given β, there is a range of α, $0 < \alpha < \alpha_1$ for which G admits ACIMs. This proof is similar to that in [10, thm.4.1].

Consider the map $G^2 : I \times I \to I \times I$. By Lemma 1.3.4, the derivative matrix $D(G^2)$ along the orbit is the product of two derivative matrices. But there are only two possible derviative matrices, D_1 and D_2. Therefore, for any orbit point, $D(G^2)$ is one of four matrix products: D_1D_2, D_1D_2, D_2D_1 or D_2D_2. For $\beta = 0.3$ (a typical case), the smaller singular values of these matrix products, as functions of α, are shown in Figure 33.

What is the value of α that guarantees all four of the σ_2 possibilities are greater than 1? Call this α_1. It can be found by doing a numerical bisection to get the solution of $\sigma_2(\alpha) = 1$ for each of the four curves in Figure 33, and then taking the

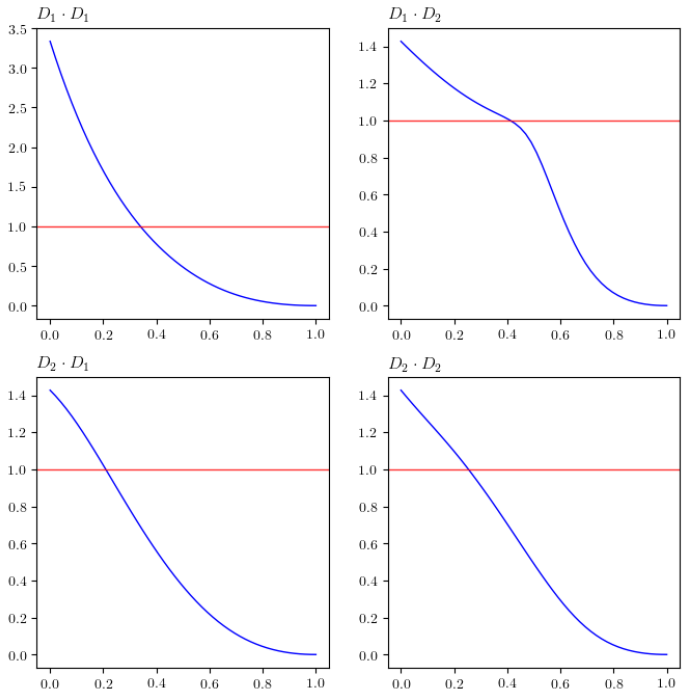

Figure 33: For fixed $\beta = 0.3$, plots of the four possible $\sigma_2(D_i \cdot D_j)$. Here α varies along the x-axis.

minimum. Since $\beta = 0.3$ is arbitrary, the same method can be used to find α_1 for any $0 < \beta < 1$. The result is Figure 34 which is a plot of $\alpha_1(\beta)$.

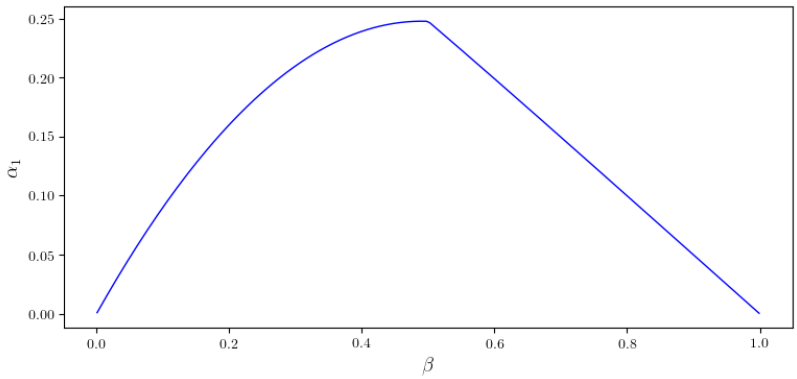

Figure 34: Plot of α_1 as a function of β.

Theorem 2.4.1 *G admits an ACIM for $0 < \alpha < \alpha_1(\beta)$, $0 < \beta < 1$.*

Proof We first show that for each matrix product $D_i D_j$, $i, j = 1, 2$, the functions $\sigma_2(\alpha)$ are monotone decreasing for any fixed $0 < \beta < 1$. Recalling from (7) that the

product, q, of the singular values is the absolute value of the determinant, we have the following $q(\alpha)$ expressions for the four matrix products,

$$\frac{(\alpha-1)^2}{\beta^2}, \quad \frac{(\alpha-1)^2}{|\beta^2-\beta|}, \quad \frac{(\alpha-1)^2}{|\beta^2-\beta|}, \quad \frac{(\alpha-1)^2}{(\beta-1)^2}.$$

Each of these has the form $\frac{(\alpha-1)^2}{Q}$, where Q is positive, so $q(\alpha)$ is a monotone decreasing function in all four cases. Therefore the product $\sigma_1(\alpha)\sigma_2(\alpha)$ is monotone decreasing. Both $\sigma_1(\alpha)$ and $\sigma_2(\alpha)$ are positive and $\sigma_1(\alpha) \geq \sigma_2(\alpha)$, therefore $\sigma_2(\alpha)$ must be monotone decreasing.

Since $\sigma_2(\alpha)$ is monotone decreasing, then $\sigma_2(\alpha) \geq \alpha_1$ for all $0 < \alpha < \alpha_1$. For these values of α, G^2 is piecewise expanding on a real-analytic neighborhood of the attractor. G^2 is also piecewise real-analytic on the same neighborhood, therefore by Theorem 1.3.1, G^2 admits an ACIM for any $0 < \alpha < \alpha_1$. Then by Lemma 1.3.5, G itself admits an ACIM for any $0 < \alpha < \alpha_1$.

The maximum of $\alpha_1(\beta)$ in Figure 34 is at $\beta = \frac{1}{2}$. This corresponds to the case of the tent memory map of Section 2.3. Bisection search gives a maximum of $\alpha_1 \approx 0.2476$, which implies the result of [10, thm.4.1]. Figures 35 and 36 are some examples of attractors which support ACIMS, as determined by Theorem 2.4.1.

The extra degree of freedom in the β-tent map with memory allows us to find new intervals of island formation and new phenomena. Some of this new phenomena is studied in Chapter 3. As an example of island formation, we examine a density consisting of 203 islands found at $\alpha = 0.49$, $\beta = 0.4942$. The island trajectory is interesting and sufficiently unlike previous ones, that it has been included as Figure 37. Island **0** has region index trajectory

```
221223221223221223221223221223221223221223221223221
221222221223221223221223221223221223221223222222222
222222222222222222222222222222222222222222222222222 .
122122122122122122122122322122322322322322222222
222222222222222222222222222222222222222222222222223
```

There are 12 islands with region index `3`, making a total of 4096 possibilities to examine. The minimum σ_2 over all these is 1.712122889006. Thus, G admits an ACIM supported on these islands. Images of the islands are found in Figure 38

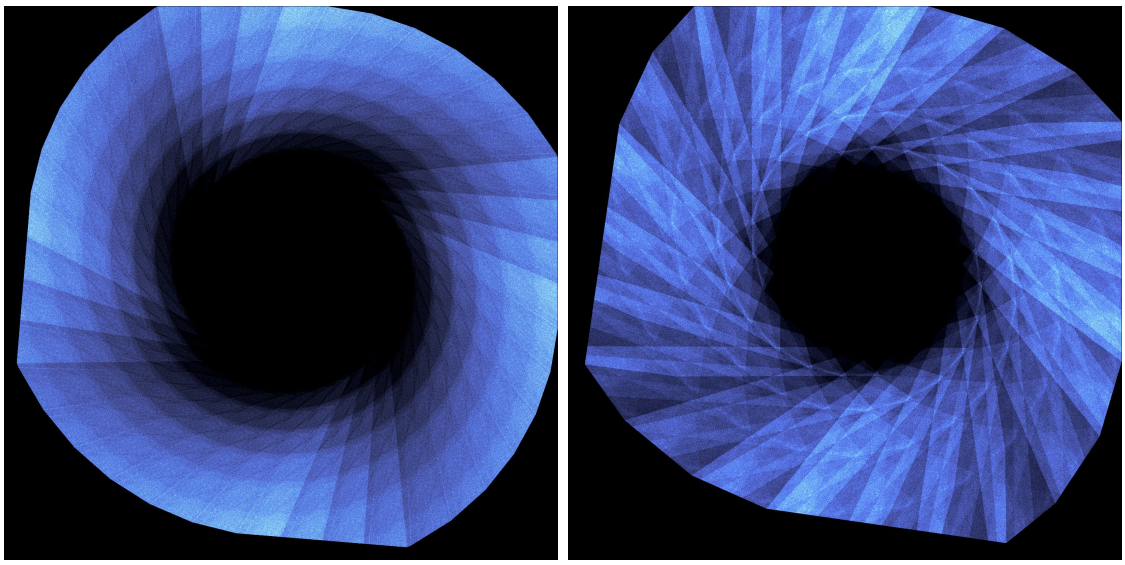

Figure 35: Densities of attractors supporting ACIMs for β-tent map with memory. Left: $\alpha = 0.75667$, $\beta = 0.08247$. Right: $\alpha = 0.122295$, $\beta = 0.143571$.

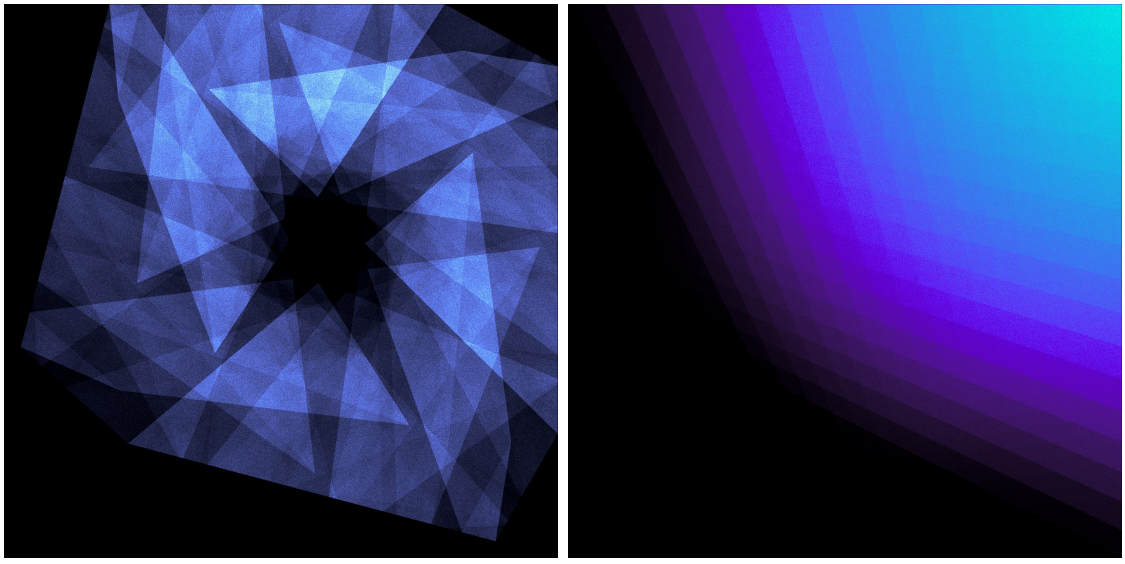

Figure 36: Densites of attractors supporting ACIMS for β-tent map with memory. Left: $\alpha = 0.19503$, $\beta = 0.265776$. Right: $\alpha = 0.03105$, $\beta = 0.93788$.

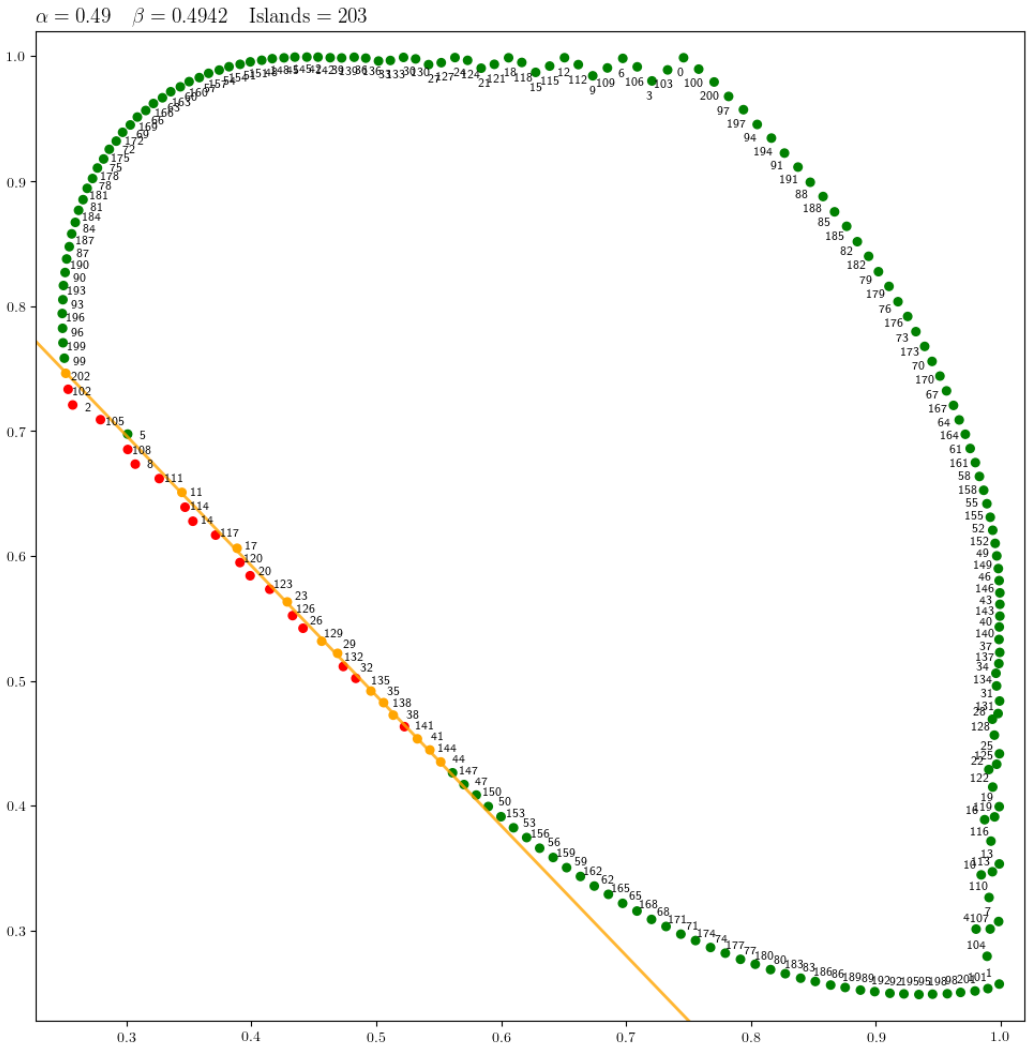

Figure 37: Island trajectory, $\alpha = 0.49$, $\beta = 0.4942$

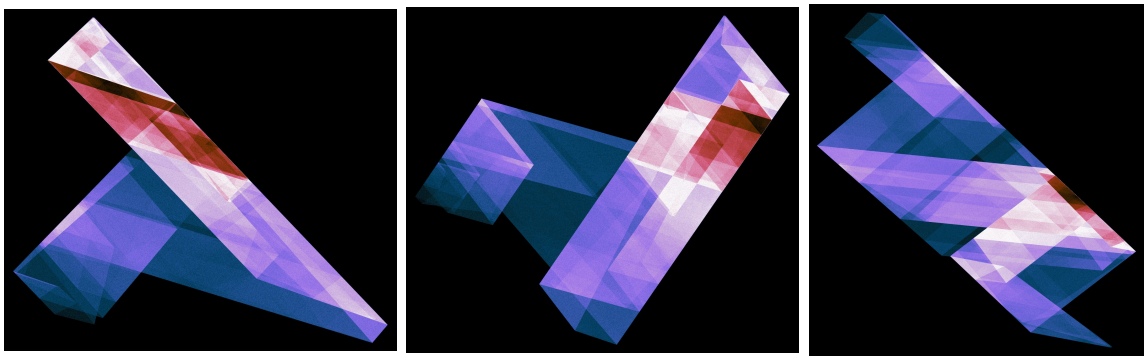

Figure 38: Close-ups of islands **29**, **126** and **200** in the $\alpha = 0.49$, $\beta = 0.4942$ system of 203 islands.

2.5 Nakamura map

In the paper [19], Nakamura and Mackey also reported the phenomenon of islands, but for a different kind of map. We will consider map $\tilde{S}$ from their paper. Here we will call it simply S. The Nakamura map S is defined as

$$S(x,y) = \begin{cases} (ax+y+1,\, bx) & x < 0 \\ (ax+y+1,\, -bx) & x \geq 0 \end{cases}$$

The vertical line $x = 0$ plays the part of the partition line. There are two regions, A_1 to the left of the partition line, and A_2 to the right. S is piecewise linear on these partitions. In matrix form,

$$\begin{aligned} S_1(x,y) &= \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x < 0 \\ S_2(x,y) &= \begin{pmatrix} a & 1 \\ -b & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x \geq 0. \end{aligned} \tag{18}$$

From (18), the derivative matrices D_1 and D_2 in regions A_1 and A_2 can be read off. The domain of interest over which the histograms and images are calculated is roughly $[0, 1.3] \times [0, -1.3]$. We can assume that the domain can be bounded by real-analytic arcs and that S, being linear, is piecewise real-analytic on this domain.

We will study two examples of Nakamura map islands taken from the paper. The islands in these systems are so large that they are easily visible in density images. The first is at $a = 0.284$ and $b = 1.02$. The density image and island trajectory diagram are shown in Figure 39. Note the vertical orange line at $x = 0$, this is the Nakamura map partition line. There are 22 islands, three of them having region index `3`. This is a mere 8 possibilities to evaluate. The region index trajectory for island **0** is:

`2223222222223222322222`.

Minimum σ_2 over the 8 possibilities is 1.069751635741. S satisfies the conditions necessary for applying Theorem 1.3.1 (Tsujii-Buzzi), and thus by the method of Theorem 2.3.2, S admits an ACIM on all 22 islands. Images of the islands are shown in Figue 40.

The second example is S with $a = 0.3015$ and $b = 1.02$. There are 31 islands. The density image and trajectory diagram are shown in Figure 41. There are only four

islands of region index 3. Island **0** has region index trajectory

223223222232221222222223222222.

The minumum singular value σ_2 over the 16 possibilities is 1.033290606356. Thus, an S invariant ACIM is supported on these 31 islands. Images of some of the islands are found in Figure 42.

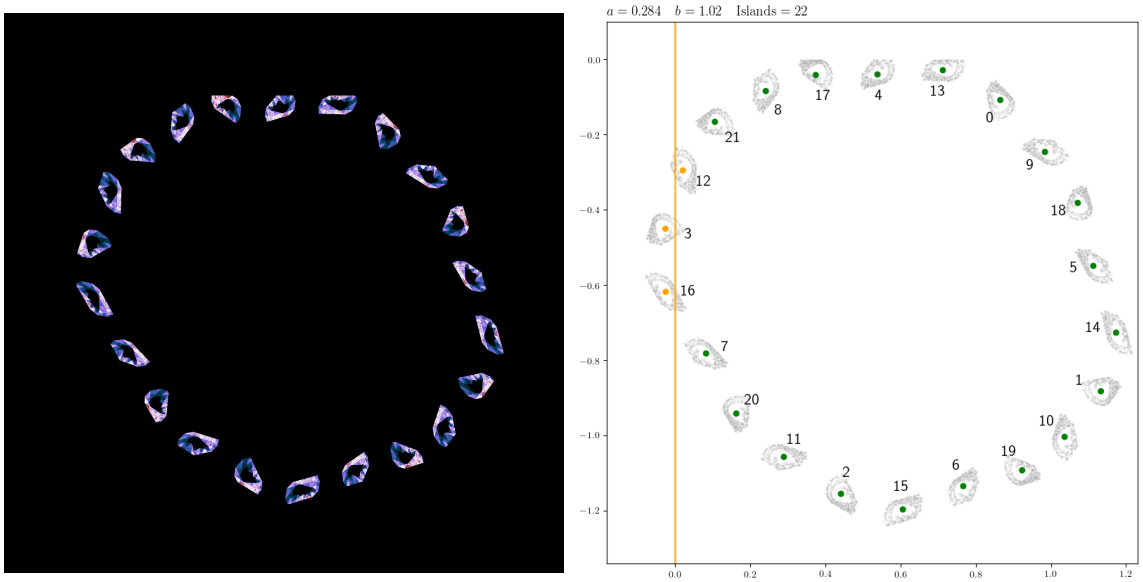

Figure 39: Density (left) and trajectory (right) for the Nakamura 22 island system, $a = 0.284$, $b = 1.02$.

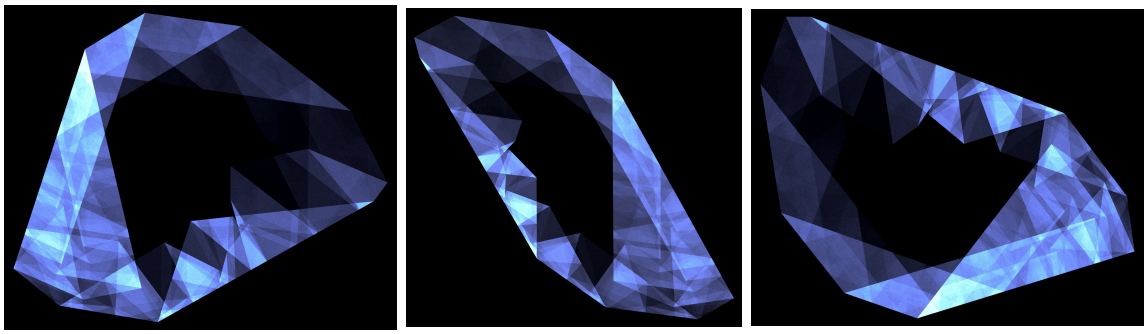

Figure 40: Islands **3**, **16** and **19** of the Nakamura-22 system.

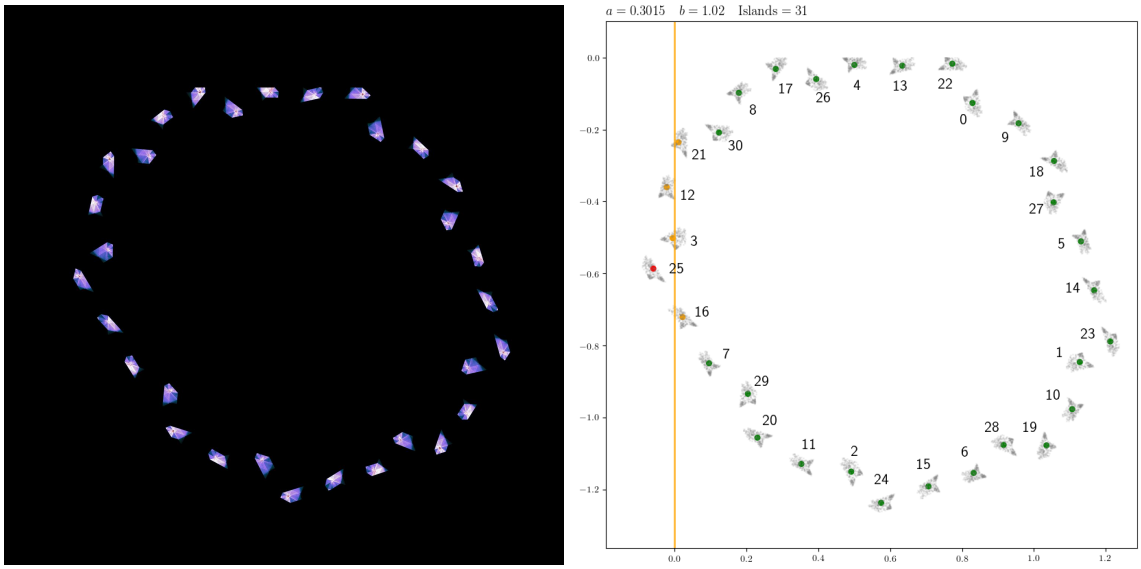

Figure 41: Density (left) and trajectory for Nakamura-31 system, $a = 0.3015$, $b = 1.02$

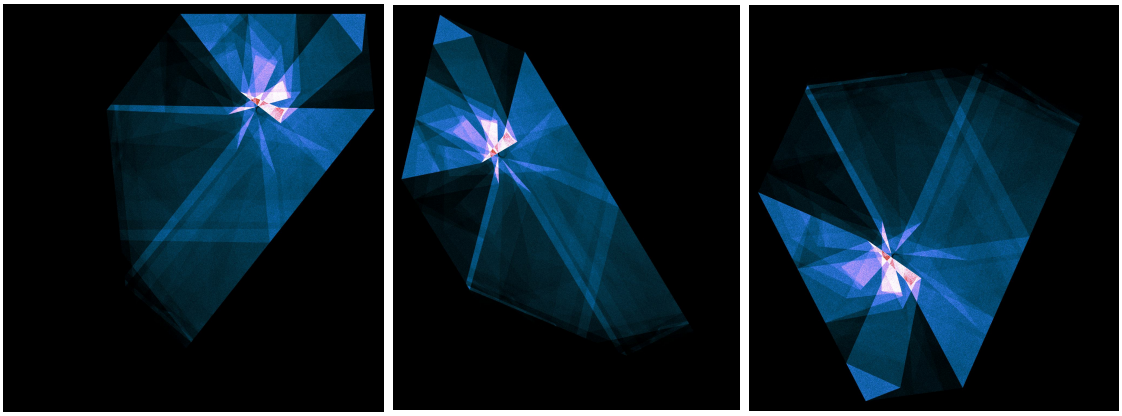

Figure 42: Islands **17**, **25** and **27** of the Nakamura-31 system.

Chapter 3

Ellipses

3.1 Elliptical dynamical system

The family of β-tent memory maps G_β was proposed and studied in Section 2.4. G_β depends on two parameters, α and β. In this chapter we examine what happens when $\alpha = \beta$, and specifically $0 < \alpha = \beta < \frac{1}{2}$.

Set $\beta = \alpha$ in (16) to obtain the new partition line:

$$(1-\alpha)x + \alpha y = \alpha. \tag{19}$$

This again partitions $I \times I$ into two regions. As before, let A_1 be the region below the partition line, and A_2 the region above. Then, we may piecewise define the elliptical map with memory G by setting $\beta = \alpha$ in (17):

$$\begin{aligned} G_1(x,y) &= \begin{pmatrix} 0 & 1 \\ \frac{1-\alpha}{\alpha} & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}, \quad u < \alpha \\ G_2(x,y) &= \begin{pmatrix} 0 & 1 \\ -1 & \frac{\alpha}{\alpha-1} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{1-\alpha} \end{pmatrix}, \quad u \geq \alpha. \end{aligned} \tag{20}$$

G_1 is the elliptical memory map G acting on A_1, G_2 is G acting on A_2 and $u = (1-\alpha)x + \alpha y$.

Density plots of the orbits of points under G are shown in Figure 43 for $\alpha = 0.3$ and $\alpha = 0.48$. Approximately 400 initial points were taken outside the boundary of the largest ellipse. Under the action of G, these initial points eventually make their way into A_2 and settle into the concentric elliptical orbits shown in the figures.

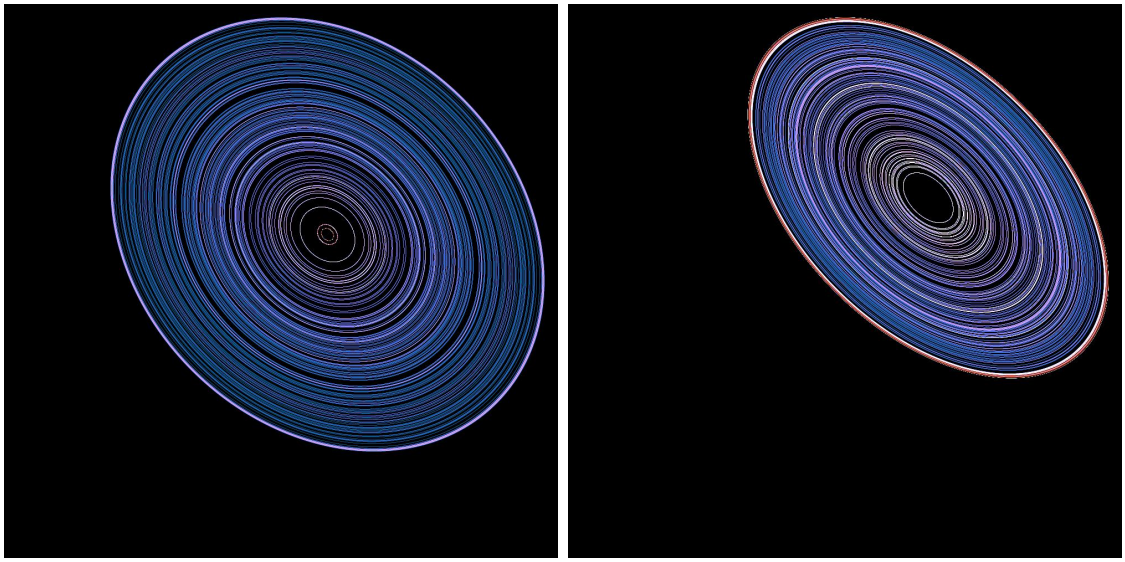

Figure 43: Densities for $\alpha = 0.3$ and $\alpha = 0.48$. 400 initial points were followed for 10^6 iterations.

For some values of α, the densities take on radically different appearances. With $\alpha = \frac{3-\sqrt{5}}{2}$ and $\alpha = 0.272601699536533$, initial points settle into periodic orbits in region A_2. A few million initial points chosen from outside the largest polygon were iterated under G, creating the polygonal density images in Figure 44, These polygons have 10 and 25 sides respectively.

Exactly why this happens and for what values of α does it happen, along with other questions about the dynamics of G, are the subject of this chapter.

3.2 Geometry of E_m and Π_m

The geometry of the $I \times I$ region with respect to the map G is shown in Figure 45 (left). $ABCD$ is the boundary of $I \times I$. BR is the partition line (19). Note that unlike previous partition lines, the partition line for the elliptical memory map always goes through point B. The anti-partition line DS is the reflection of the partition line BR over the diagonal AC. The anti-partition line always goes through point D.

The *maximal ellipse* E_m is defined to be the ellipse with center O on the diagonal AC, tangent to the boundary lines BC, CD and tangent to the partition and anti-partition lines BR, SD. This is enough to determine E_m uniquely.

By symmetry, E_m must be inclined at $-45°$ to the x-axis. Since the description

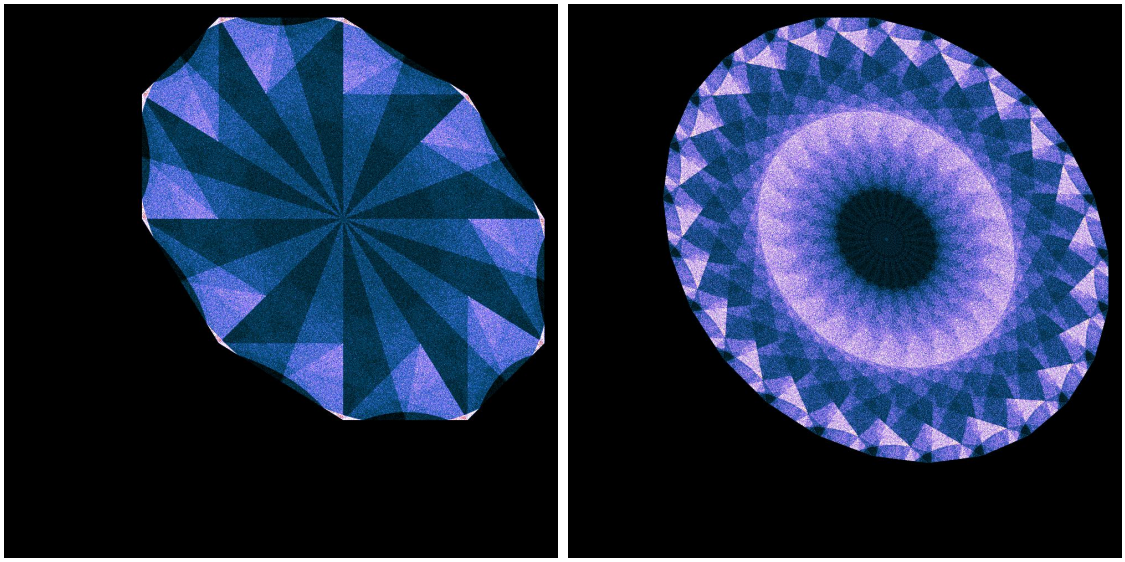

Figure 44: Polygonal densities for $\alpha = (3-\sqrt{5})/2 = 0.3819660112501051$ (left) and $\alpha = 0.272601699536533$ (right). Densties were made by following 5×10^6 and 3×10^6 initial points for 1500 iterations.

of E_m only depends on the slope of the partition line, $\frac{\alpha-1}{\alpha}$, all geometric properties of E_m and of the scene in Figure 45 can be determined by analytic geometry in terms of only α. We present the results of these analytic geometry computations. The slope of the anti-partition line DS is $\frac{\alpha}{\alpha-1}$. Points R and S are determined from the intersections of BR and DS with the box $ABCD$,

$$R = \left(\frac{\alpha}{1-\alpha}, 0\right) \quad S = \left(0, \frac{\alpha}{1-\alpha}\right).$$

The point X, which is the intersection of the partition and anti-partition lines, is simply (α, α). Let $c = \frac{1}{2-\alpha}$. The center O of the maximal ellipse E_m is at (c, c). The squares of the semi-major and semi-minor axes of E_m are respectively

$$a^2 = \frac{(1-\alpha)}{2(2-\alpha)}, \quad b^2 = \frac{(\alpha-1)(3\alpha-2)}{2(2-\alpha)^2}. \tag{21}$$

In terms of the following level-curve function

$$f(x,y) = \frac{(x-y)^2}{2a^2} + \frac{(x+y-2c)^2}{2b^2}, \tag{22}$$

the equation of E_m is $f(x, y) = 1$.

The four tangent lines BR, BC, CD and DS upon which the construction of E_m is based, are tangent at points T_0, T_1, T_2 and T_3 respectively:

$$T_0 = \left(\frac{\alpha}{2(1-\alpha)}, \frac{1}{2}\right), \; T_1 = \left(\frac{1}{2}, 1\right), \; T_2 = \left(1, \frac{1}{2}\right), \; T_3 = \left(\frac{1}{2}, \frac{\alpha}{2(1-\alpha)}\right).$$

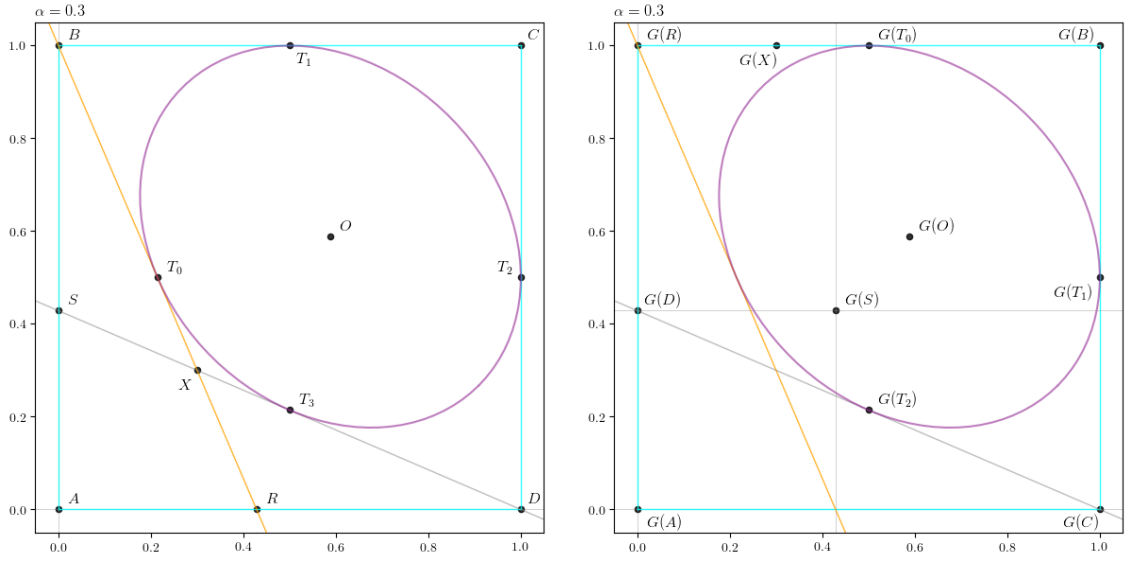

Figure 45: Left: maximal ellipse E_m for $\alpha = 0.3$. BR is the partition line, SD is the anti-partition line. Center is at $O = (c, c)$. Right: images of the points in left figure under G.

Note that the x-coordinate of T_0 is always the midpoint of segment AR, while the y-coordinate of T_3 is always the midpoint of AS.

We now consider where the geometrical points of Figure 45 (left) are mapped by G. That is, the images of these points under G. There are two fixed points of G in $I \times I$: A and O. Point A (in region A_1) is a fixed point of G_1 and O (in region A_2) is a fixed point of G_2. The tangent points T_1, T_2 and T_3 are iterated images of T_0:

$$T_1 = G(T_0), \;\; T_2 = G_2(T_1), \;\; T_3 = G_2(T_2).$$

Some of the bounding box points A, B, C, D are mapped by G in the clockwise direction: $G(B) = C$, $G(C) = D$, while D is mapped to the other end of the partition line, and likewise R to the other end of the anti-partition line. Point S is mapped into region A_2: $G(S) = \frac{1}{1-\alpha}(\alpha, \alpha)$, forming a square with points A, S and R. Finally X is mapped vertically, $G(X) = (\alpha, 1)$. These images under G are shown in Figure 45 (right).

When R and S divide AD and AB in the golden mean, then $G(S)$ coincides with O. This special case happens when $\alpha = \frac{3-\sqrt{5}}{2}$ and is featured in Figure 44 (left) and Figure 46 (right).

Proposition 3.2.1 *The maximal ellipse $f(x, y) = 1$ is invariant under G_2 and so are the level-curve ellipses $f(x, y) = k < 1$ which lie inside the maximal ellipse.*

Proof Suppose (x, y) is a point on E_m. Let $(x', y') = G_2(x, y) = (y, \frac{\alpha y+(1-\alpha)x-1}{\alpha-1})$. We show that (x, y) and (x', y') are on the same level-curve. Let $p = \frac{4}{(\alpha-1)(3\alpha-2)}$. Then

$$\begin{aligned} f(x', y') &= p \left[\begin{array}{r} (\alpha-2)(\alpha-1)y^2 - \alpha(\alpha-2)xy + (\alpha-2)y + \\ +(\alpha-2)(\alpha-1)x^2 + (\alpha-2)x + 1 \end{array} \right] \\ &= p \left[(x-y)^2 \frac{(\alpha-2)}{\alpha-1} + (x+y-2c)^2 \frac{(\alpha-2)^2}{(\alpha-1)(3\alpha-2)} \right] \\ &= \frac{(x-y)^2}{2a^2} + \frac{(x+y-2c)^2}{2b^2} = f(x, y). \end{aligned}$$

This is true for the maximal ellipse level-curve $f(x, y) = f(x', y') = 1$, but it is also true for any other level curve in region A_2, e.g., $f(x, y) = f(x', y') = k < 1$. Therefore E_m is invariant under G_2, as well as all level-curves inside E_m.

Proposition 3.2.2 *G_2 is conjugate to a rotation.*

Proof Let Γ^{-1} be the transformation that brings the maximal ellipse into canonical position, i.e., center on $(0, 0)$ and semi-axes parallel to the x and y axes. Γ^{-1} is translation from (c, c) to the origin followed by a rotation of $+45°$. Γ is then a rotation through $-45°$ followed by translation from $(0, 0)$ to (c, c).

$$\Gamma(x, y) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ c \end{pmatrix}, \quad \Gamma^{-1}(x, y) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x-c \\ y-c \end{pmatrix} \quad (23)$$

Define the scaling transformation matrices Λ, Λ^{-1}:

$$\Lambda = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad \Lambda^{-1} = \begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix}$$

where a, b are the semi-axes of E_m. Define the map K to be the conjugate of G,

$$K = \Lambda^{-1}(\Gamma^{-1} \circ G \circ \Gamma)\Lambda.$$

While G leaves E_m invariant, K leaves the unit circle invariant. Consider K_2, which is the map K restricted to the image of region A_2, i.e., $K_2 = \Lambda^{-1}(\Gamma^{-1} \circ G_2 \circ \Gamma)\Lambda$ on $A_2' = \Lambda^{-1}\Gamma^{-1}A_2$,

$$K_2(x, y) = \frac{1}{2(\alpha-1)} \begin{pmatrix} \alpha & \frac{(\alpha-2)b}{a} \\ \frac{(2-3\alpha)a}{b} & \alpha \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

The K_2 matrix is orthogonal with determinant 1. Therefore K_2 is a pure rotation. The sines and cosines of the angle of rotation can be read off from the matrix elements of K_2, but we give a derivation of the angle of rotation based on geometry.

We call the scene in Figure 45 (left), with its geometrical elements, the *standard scene*. Transforming the geometrical elements of the standard scene by $\Lambda^{-1}\Gamma^{-1}$ results in the *isometric scene* of Figure 46. Let primes denote isometric scene equivalents of the points in Figure 45. Then B' is the point B mapped into the isometric scene by $\Lambda^{-1}\Gamma^{-1}$:

$$B' = \Lambda^{-1}\Gamma^{-1}B = \frac{1}{\sqrt{2}}\left(-\frac{1}{a}, \frac{\alpha}{(\alpha-2)b}\right).$$

Referring to Figure 46 (left), since $K_2 T_0' = T_1'$, then the rotation angle θ must be the clockwise rotation through $\angle T_0' O' T_1'$. This angle depends only on the position of B', since E_m', the mapping of E_m into the isometric scene, is the unit circle, which is fixed. Since $\angle O' T_0' B'$ is a right angle, $\cos\frac{\theta}{2} = \frac{1}{O'B'}$. Then $\cos\frac{\theta}{2} = \sqrt{\frac{\cos\theta - 1}{2}} = \frac{1}{O'B'}$ and

$$\cos\theta = \frac{2}{(O'B')^2 - 1} = \frac{2}{{B_x'}^2 + {B_y'}^2} - 1 = \frac{\alpha}{2(\alpha-1)}.$$

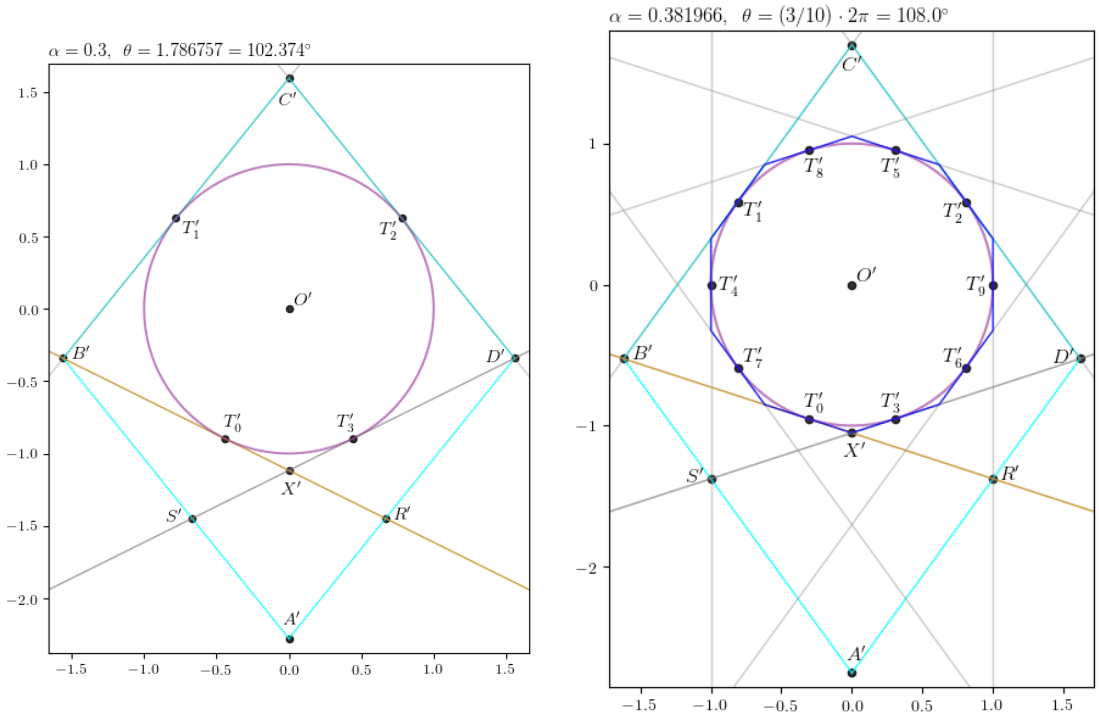

Figure 46: Isometric scene for general α (left) and periodic α (right). Maximal polygon Π_m is shown in blue.

All points on the maximal ellipse and inside it, travel on their corresponding level-curve ellipses. Their trajectories are either periodic or dense in their ellipses, depending on the nature of the conjugated rotation angle θ.

The rotation angle $\theta(\alpha)$ is an increasing function of α. As $0 < \alpha < \frac{1}{2}$, θ goes from $\theta(0) = \frac{1}{4} \cdot (2\pi)$ to $\theta\left(\frac{1}{2}\right) = \frac{1}{3} \cdot (2\pi)$. Let $\frac{1}{4} < r < \frac{1}{3}$ be a real number, either rational or irrational. Let $\theta = r \cdot (2\pi)$. We say that θ is *π-rational* if $r = \frac{p}{q}$ with $\gcd(p, q) = 1$, i.e., when r is rational. And we say θ is *π-irrational* when r is irrational. When $r = \frac{p}{q}$, θ is π-rational, and the orbits of points on E_m or on a level curve inside E_m are periodic with period q.

Figure 46 (right) shows an example of the π-rational periodic case, in the isometric scene. There will be $q - 1$ images $T'_1, T'_2, \ldots T'_{q-1}$, of the point T'_0 under applications of K. Together with T'_0 we will call these q points *tangent points*. The order in which the tangent points appear adjacently around E'_m can be determined. Let T'_n be a tangent point. Then the next adjacent tangent point in the clockwise direction is T'_m, where $m = n + k \bmod q$ and $pk = 1 \bmod q$. The same holds for E_m and the tangent points in the standard scene.

Tangent points are midpoints of the sides of the *maximal polygon* Π'_m, shown in Figure 46 (right) in the isometric scene. Π'_m is constructed by drawing tangent lines at adjacent tangent points on E'_m and computing their intersections. There are q such intersections and they form the vertices of a regular q-sided polygon. Since the standard and isometric scenes are related by linear transformations, these considerations are equally true for Π_m, the maximal polygon in the standard scene.

Proposition 3.2.3 *Let θ be π-rational. I.e., $\theta = \frac{p}{q} \cdot (2\pi)$ with $\gcd(p, q) = 1$. Then the maximal polygon Π_m is invariant under G and points inside Π_m are periodic.*

Proof Consider the isometric scene. The maximal polygon is entirely in region A'_2. K_2 in region A'_2 is a clockwise rotation with period q. Since Π'_m is a regular polygon with q sides, such a rotation leaves Π'_m unchanged. The conclusion must hold for Π_m in the standard scene.

3.3 Asymptotic behavior of points outside E_m

Proposition 3.3.1 *For $0 < \alpha < \frac{1}{2}$, the image of S under G is always in A_2.*

Proof $G_1(S) = \frac{1}{1-\alpha}(\alpha, \alpha)$ and this point ranges between $(\alpha, \alpha) = X$ to $(1, 1) = C$ as $0 < \alpha < 1/2$, therefore $G_1(S)$ is on the diagonal $y = x$ in region A_2.

Proposition 3.3.2 *Any point in* $A_1 \setminus \{(0, 0)\}$ *eventually enters* A_2.

Proof Exclude $(0, 0)$ because it is a fixed point. For any point $(x, y) \in A_1 \setminus \{(0, 0)\}$, its image $G_1(x, y) = (y, \frac{1-\alpha}{\alpha}x + y)$ has modulus strictly larger than that of (x, y) itself

$$y^2 + \left(\frac{1-\alpha}{\alpha}x + y\right)^2 > y^2 + \left(\frac{1-\alpha}{\alpha}x\right)^2 > y^2 + x^2$$

for $0 < \alpha < 1/2$. Thus, any point in $A_1 \setminus \{(0, 0)\}$ eventually enters region A_2.

From the level-curves function (22), the quantity $f(x, y) - 1$ can be used as a signed distance of (x, y) from the maximal ellipse. Then the quantity

$$D(x, y) = f(x, y) - f(G(x, y)) \tag{24}$$

can be considered as a "movement" function, measuring how the distance of (x, y) from the maximal ellipse changes under the action of G. When $D(x, y) > 0$ then G either moves the point (x, y) closer to the maximal ellipse, or into it.

Proposition 3.3.3 *For any point in* A *we have* $D(x, y) \geq 0$. *The inequality is strict for points in* A_1 *except for the origin and points on the partition line.*

Proof For $(x, y) \in A_2$, we have $D(x, y) = 0$ because every level curve concentric with E_m is invariant under G_2.

We will show that $D(x, y) \geq 0$ for points in A_1, so either their distances from E_m decrease or they enter E_m. Substituting the semi-axes (21) in (22), (24) and simplifying, we have

$$D(x, y) = \frac{8\left(\left(x - \frac{y}{2}\right)\alpha - \frac{x}{2}\right)(\alpha - 2)((x - y + 1)\alpha - x)}{3\alpha^4 - 5\alpha^3 + 2\alpha^2}.$$

When $0 < \alpha < \frac{1}{2}$, the denominator is positive. The factor $8(\alpha - 2)$ is negative. The factor $((x - y + 1)\alpha - x)$ is zero on the partition line, and since it is a linear function, it is positive for smaller y, i.e., for points in A_1. The factor $h(x, y) = \left(\left(x - \frac{y}{2}\right)\alpha - \frac{x}{2}\right)$ is again a linear function. Its values on the vertices of A_1 are

$$h(0, 0) = 0, \quad h(0, 1) = -\frac{\alpha}{2}, \quad h\left(\frac{\alpha}{1-\alpha}, 0\right) = -\alpha.$$

All are non-positive, so h is non-positive on region A_1. Altogether this shows that $D(x, y)$ is non-negative on A_1. This means that the distance of $G_1(x, y)$ to the maximal ellipse is smaller than or equal to this distance for the original point (x, y). In addition, h is strictly negative for points in $A_1 \setminus \{(0, 0)\}$ below the partition line, thus $D(x, y) > 0$ for these points.

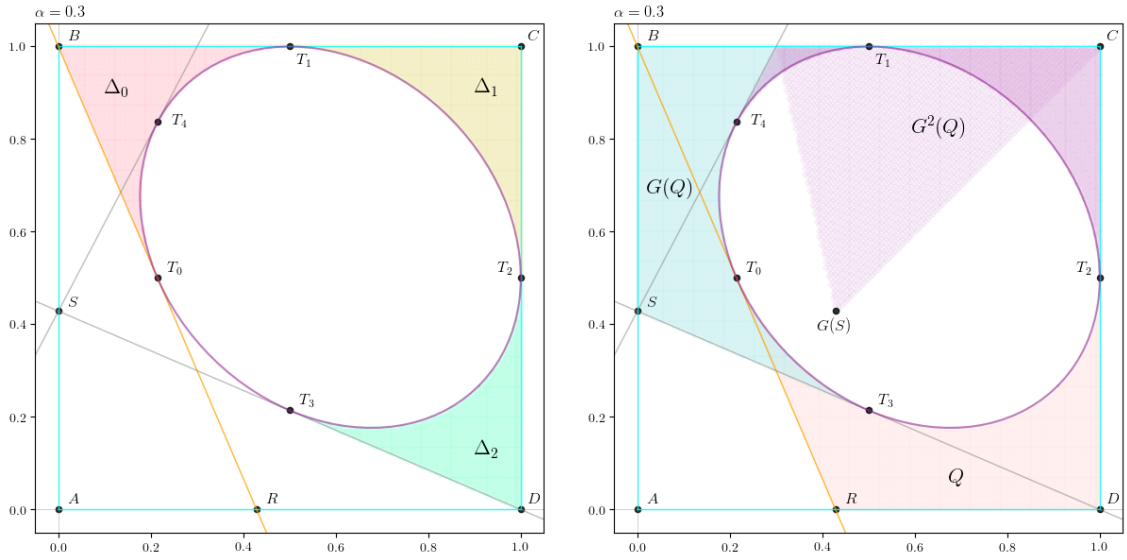

Figure 47: Left: triangle-like regions Δ_0, Δ_1 and Δ_2. Right: quadrangle-like region Q and its images under G.

Proposition 3.3.4 *If $x \in A_2$ visits A_1 then x returns to A_2 with the next application of G.*

Proof Let Δ_0 and Δ_1 be the triangle-like regions $\triangle T_0T_1B \setminus E_m$ and $\triangle T_1T_2C \setminus E_m$ shown in Figure 47 (left). Let Q be the quadrangle-like region $\square T_0T_2DR \setminus E_m$ shown in Figure 47 (right). Then $G(\Delta_0) = \Delta_1$ and $G_2(\Delta_1) = \Delta_2 = \triangle T_1T_2C \setminus E_m$ and $G_2(\Delta_2) \subset Q$. Thus it is enough to study the future of the points in Q.

The triangle-like regions Δ_0, Δ_1 and Δ_3 are shown in various colors in Figure 47 (left). The quadrangle-like region Q and its images $G_2(Q)$ and $G(G_2(Q))$ are shaded in various other colors in Figure 47 (right). The vertices of these regions can be read from Figure 45 and the way these regions are mapped can be determined from the fact that G_1 and G_2 map intersections to intersections and tangents to tangents:

$$G_2(T_0) = T_1, \;\; G_2(T_2) = T_3, \;\; G_2(D) = S, \;\; G_2(R) = B$$
$$G(G_2(T_0)) = T_2, \;\; G(G_2(T_2)) = T_4, \;\; G(G_2(D)) = G(S), \;\; G(G_2(R)) = C.$$

Thus $G_2(Q) = \square T_1 T_3 SB \setminus E_m$ and $G(G_2(Q))$ is the region with vertices T_2, T_4, $G(S)$, C and bounded by the boundary BC of $ABCD$.

Because T_3 is on E_m, then by Proposition 3.2.1, $G(T_3)$ must also be on E_m and is therefore in A_2. By Proposition 3.3.1, $G(S)$ is in A_2. Therefore all the vertices of $G(G_2(Q))$ are in A_2, and this in turn shows that any point of A_2 which visits A_1 returns to A_2 after one step.

The following lemma characterizes the periodic points of the map G.

Lemma 3.3.5 *A point $x \in I \times I \setminus \{(0,0)\}$ can be periodic if and only if it never visits the region A_1. In particular, x is periodic if and only if it never visits A_1 and θ is π-rational.*

Proof Points in A_1 eventually enter A_2, so it is enough consider points of A_2. Let $x \in A_2$ and consider two cases. (1) There is some k such that $G^k(x) \in A_1$. Then by Proposition 3.3.3, $D(G^{k+1}(x)) < D(G^k(x))$ and the value of D further along on the orbit never increases again, which implies that x is not periodic. (2) For all k, $G^k(x)$ is in A_2. Then the orbit is actually $G_2^k(x)$, so x is either periodic or non-periodic depending on whether the rotation angle θ is π-rational or not.

Some properties of the Δ-regions of Figure 47 (left) are more easily studied in the isometric scene. We dispense with the use of primes to denote geometrical features transformed into the isometric scene by $\Lambda^{-1}\Gamma^{-1}$. The presence or absence of primes will understood by context. In the isometric scene, E_m will refer to the maximal ellipse which is the unit circle.

When θ is π-rational, there are five special Δ-regions such that their vertices touch the boundaries of $ABCD$. These are shown in Figure 48 (left). If we imagine ourselves to be observers at the center of E_m, then the Δ-regions can be given labels based on the tangent lines that make up their left edges. Thus we have Δ_0, Δ_1, Δ_2, Δ_3 and Δ_{q-1} which always lies between Δ_2 and Δ_3.

Let θ be π-rational with $r = \frac{p}{q}$, $\gcd(p,q) = 1$. Then $T = \{T_0, T_1, \ldots T_{q-1}\}$ is the set of all distinct images of T_0. The lines which make up the edges of a Δ-region are tangent lines at two points in T, such that there are exactly $p-1$ other points of T lying adjacently between them. We can say that these tangent lines are a *tangent line pair of order* $p-1$. We may also refer to the tangent points on which these tangent

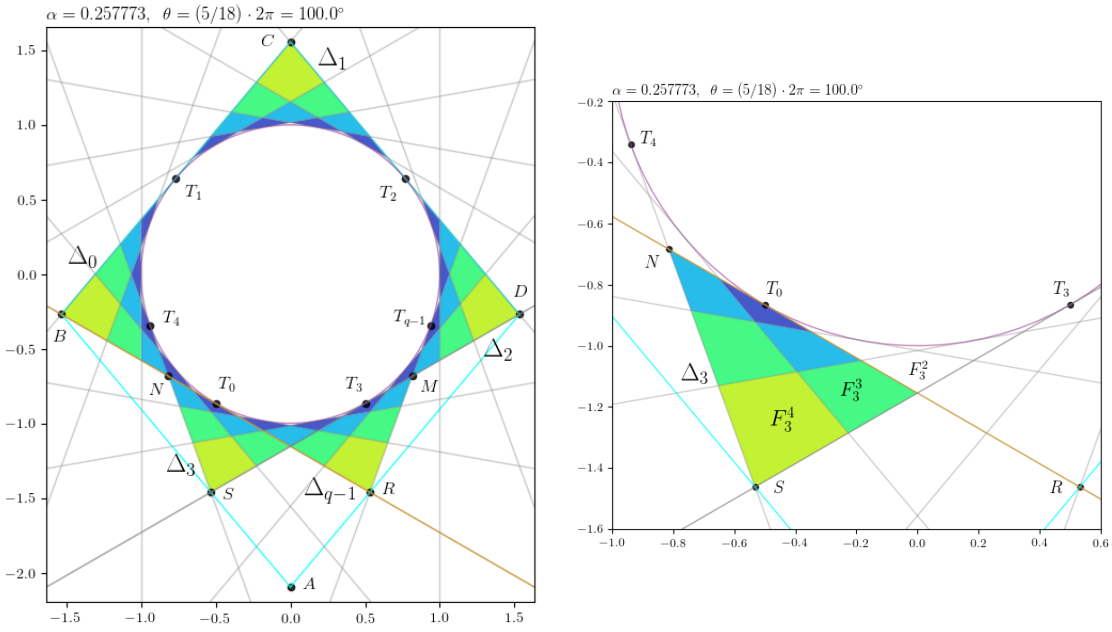

Figure 48: Left: the five Δ-regions decomposed into faces F_k^n. Right: zoom on the faces of Δ_3 lying below the partition line in region A_1. Faces of higher order are assigned lighter colors.

lines lie as a *tangent point pair of order* $p-1$, or simply as *tangent pair of order* $p-1$ to refer to both. In particular, tangent line pairs of order 0 intersect at the vertices of Π_m and constitute the sides of Π_m.

With the notion of tangent pairs of order n we can decompose each Δ-region into disjoint smaller regions which we will call *faces*. Figure 48 (left) shows a decomposition of the five Δ-regions into disjoint colored faces.

A tangent line pair of order 0 together with the maximal ellipse E_m form a *face of order* 0. Points inside E_m are defined to be order 0. Faces of order 0 lie in A_2 and they are in the maximal polygon Π_m. Alternatively we can consider Π_m itself to be the single face of order 0.

A face of order 1 is a region bounded by a tangent line pair of order 1 and does not contain any faces of order 0. A face of order n is bounded by tangents of order n and does not contain any face of order less than n.

A face of order $n \geq 2$ can be constructed in the following way. Begin with a tangent pair of order n. There will be n adjacent tangent points lying on E_m in between the tangent points on which the tangent line pair of order n lie. From these n inner tangent points, the choice of a pair of tangents points of order $n-2$ is unique.

The original tangent line pair of order n along with the new tangent line pair of order $n-2$ constitute the boundaries of a face of order n. For $n \geq 2$, faces of order n are always quadrilaterals.

It is sometimes useful to have a labelling scheme for faces, especially when trying to draw them by computer programs. One scheme is as follows. Viewed by an observer at the center of E_m, a face of order n can be given a unique label based on the left tangent line of the tangent line pair of order n used in its construction. If this line is tangent at point $T_k \in T$ then the face can be given the label F_k^n. Some labels of faces are shown in Figure 48 (right). A point x is a *point of order* n if it is in some F_k^n, with smallest n chosen if the point happens to be on the boundary between faces. An easy way to see the order of a point is to find the *minimum* n such that a tangent pair of order n encloses x against E_m. For example, the points R, S, B, C, D are all of order 4. S is in F_3^4, D is in F_2^4, N is in F_0^1 and R is in F_{q-1}^4.

Proposition 3.3.6 *Let F be the face of order 1 centered on T_0. Then $G(F)$ is in Π_m.*

Proof Refer to Figure 49. Face F is the light red triangular face with base centered on T_0 and with vertex V. The image of F, $G(F)$, is shown in light blue, with vertex $G(V)$ and base centered on T_1. Region $\Delta_3 \cap A_1$ or $\triangle SNX$ is shown in light gray. Its image, $G(\triangle SNX)$ is the triangle $\triangle G(S)G(N)G(X)$. Note that, as mentioned earlier, in the standard scene $G(S)$ lies on the same horizontal line as S, i.e., on the line $y = \frac{\alpha}{1-\alpha}$.

The line OT_0 is shown in red. Vertex V lies on this line. That is to say, V, T_0 and O are collinear. This is evident in the isometric scene. In the isometric scene, Π_m is a regular polygon. The sides of Π_m are on the tangent points. Thus the side of Π_m which is the base of F in the isometric scene, has midpoint T_0. Vertex V is constructed from two tangent lines erected on the two tangent points neighboring T_0. Therefore in the isometric scene, F is an isoceles triangle and V, T_0, O are collinear. Linear transformations Γ and Λ preserve collinearity, so these three points must be collinear in the standard scene as well.

The map G_1 is also a linear transformation and thus preserves collinearity. If we extend G_1 to the whole plane, then it maps the light red line VT_0O to the light blue line $G_1(V)T_1G(O)$. The point $G(O)$ lies above BC so it is not visible in Figure 49.

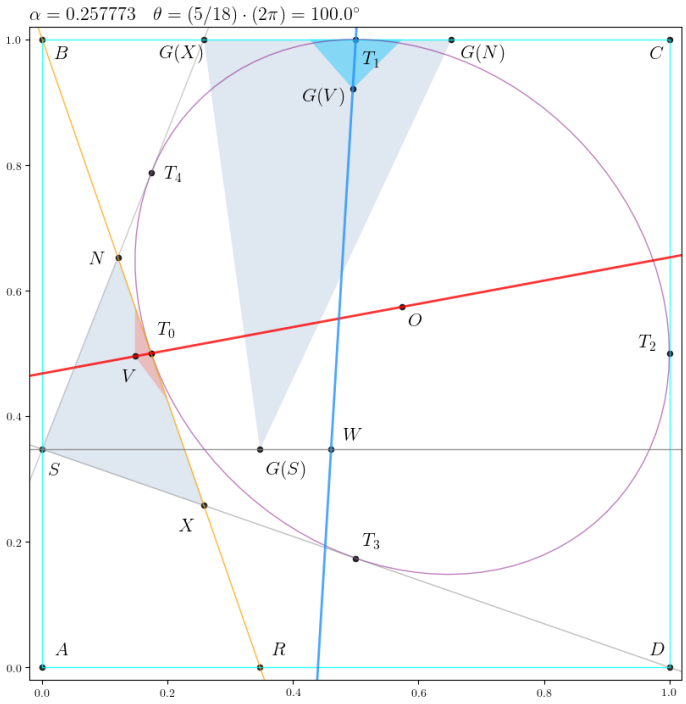

Figure 49: The face of order 1 on T_0 and its vertex V.

Extending this blue line downward, it meets $y = \frac{\alpha}{1-\alpha}$ at point W, which will be examined shortly.

Because F is a face of order 1, it is a triangular face, not a quadrilateral face. In the isometric scene, the base of F, lying on point T_0, is a side of the regular polygon Π_m. Therefore if the vertex V is moved inside E_m, then the triangle made with the base of F will lie completely within Π_m. Therefore it is sufficient to prove that the image $G(V)$ of the vertex V is always inside E_m. However, working with V directly may be algebraically intractable so we seek another point that is simpler.

Consider the point W in Figure 49, which is the intersection of $T_1G(V)$ and the horizontal line $y = \frac{\alpha}{1-\alpha}$ going through S.

Figure 50 shows the limiting behavior of W as $0 < \alpha < \frac{1}{2}$. As $\alpha \to 0$, the blue line approaches the vertical line through T_1, and S approaches A, and the point W seems to approach T_3 from above, suggesting that W remains in E_m. As $\alpha \to \frac{1}{2}$, the blue line approaches the line AT_1, while S approaches B and W seems to approach T_1 from below, again suggesting that W remains in E_m.

If it can be shown that W is always in E_m then it would imply that the image of V is also always in E_m for the following reason. Since $F \subset \Delta_3 \cap A_1$ then $G(F) \subset$

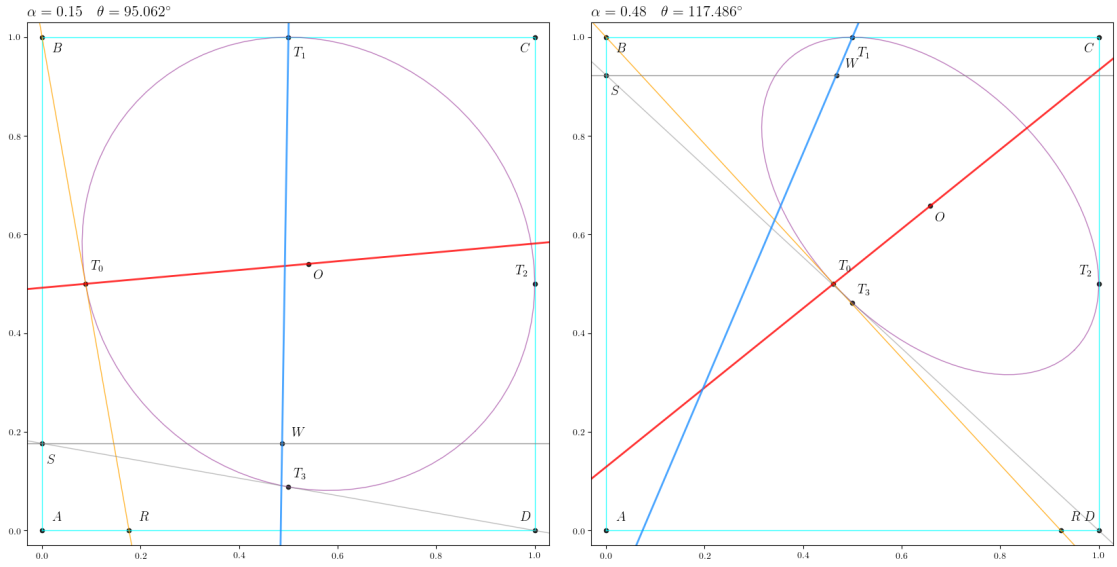

Figure 50: Limiting behavior of the blue line T_1W. Left, $\alpha = 0.15$. Right, $\alpha = 0.48$.

$G(\Delta_3 \cap A_1)$ and it must be that $G(V)$ is inside $G(\Delta_3 \cap A_1)$. But the line T_1W crosses the boundary of $G(\Delta_3 \cap A_1)$ (except in the degenerate case where $W = G(S)$, but this amounts to the same thing), therefore $G(V)$ must lie on the line $T_1G(V)$ in the segment between T_1 and W. Therefore if W is inside E_m, so is $G(V)$.

The point W is found by computing the intersection of the lines $G_1(OT_0)$ and $y = \frac{\alpha}{1-\alpha}$:

$$W = (W_x, W_y) = \left(-\frac{\alpha^3 + 2\alpha^2 - 3\alpha + 1}{2\alpha^3 - 6\alpha^2 + 6\alpha - 2}, \frac{\alpha}{1-\alpha} \right).$$

Putting W in the level-curves function (22), we obtain

$$f(W_x, W_y) = \frac{(3\alpha^3 - 8\alpha^2 + 5\alpha - 1)(\alpha^4 - 6\alpha^3 + 13\alpha^2 - 9\alpha + 2)}{(\alpha - 1)^6(3\alpha - 2)}. \tag{25}$$

Expression (25) is plotted in Figure 51. As $\alpha \to 0$ and $\alpha \to 1/2$, the expression approaches 1 from below, meaning that W approaches E_m from the inside. The minimum is at $\alpha \sim 0.37405812$, not far from the golden mean case of $\alpha = (3-\sqrt{5})/2 \sim 0.381966$. For all $0 < \alpha < 1/2$, W is in E_m.

We return to the topic of faces and prove some propositions about them which will be of use in proving the main theorem of this chapter, Theorem 3.3.10.

Faces have important properties. (1) K_2 does not change the order n of a face F_k^n. (2) Faces are periodic with respect to the rotation K_2. That is, there exists an

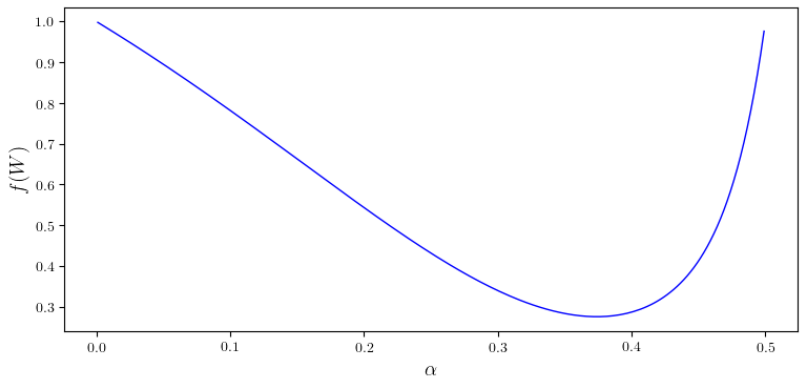

Figure 51: Plot of the expression in (25) for $0 < \alpha < \frac{1}{2}$.

m such that a face $F^n_{k_1}$ can be brought to coincide with another face $F^n_{k_2}$ of the same order by $F^n_{k_2} = K^m_2(F^n_{k_1})$. These are true for faces and G_2 in the standard scene as well.

Let t be a tangent line on one of the points of the set T. We say that a point x is an *inner point* with respect to t if x and the center of E_m are on the same side of t. A point x is an *outer point* with respect to t if x and the center of E_m lie on opposite sides of t.

The notion of *inner* and *outer* can be applied to faces as well. Given a tangent line t, if a face is on the same side of t as the center of E_m, then it is an *inner face* with respect to t. If it lies on the opposite side of t, then it is an *outer face* with respect to t.

Proposition 3.3.7 *Let t be a tangent line on one of the points of the set T. Suppose F is an outer face of order $n \geq 2$ having an edge along t, and F' is an inner face sharing the same edge. Then the order of F' is $n - 1$.*

Near the tangent point on which t lies, the outer faces along t make acute angles with t which become larger as order increases. However, the inner faces along t near the tangent point make obtuse angles with t and these angles become smaller as order increases.

Proof It is sufficient to consider the example in Figure 52. Take the tangent line t to be the orange partition line, which is on T_0. The faces $abed$ and $bcfe$ are outer faces along t, while $ehgd$ and $fihe$ are inner faces along t. Supppose the outer face $abed$ is order n. This face shares edge ed with inner face $chgd$. The edge ed is along

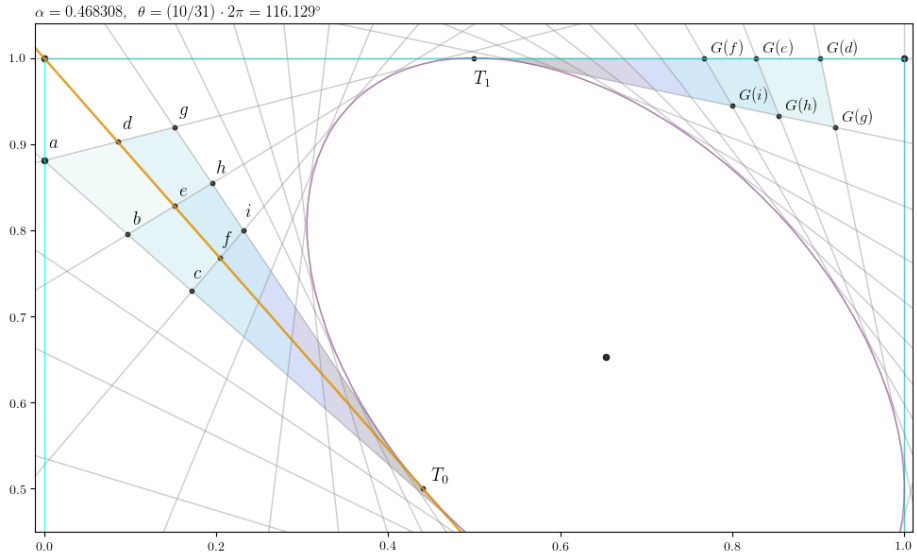

Figure 52: Inner and outer faces along tangent lines at T_0 and T_1.

t. Therefore the order of $ehgd$ is $n-1$. This is evident if you consider that these two faces are constucted by the tangent lines ag and bh. Likewise, outer face $bcfe$ has order $n-1$ while the inner face sharing an edge with it, $fihe$, has order $n-2$.

The outer faces along t make angles $\angle cfT_0$, $\angle beT_0$, $\angle adT_0$, etc. The outer face of order 1 on T_0 is an equilateral triangle, so here the angle is acute. Therefore the angles that the outer faces make with t begin as acute and grow larger as order increases.

Similarly, inner faces along t make angles $\angle gdT_0$, $\angle heT_0$, $\angle ifT_0$, etc. with line t. These angles grow smaller as order increases. The inner face closest to T_0 (having T_0 as one of its vertices) is an order 0 face making an obtuse angle with t. Therefore the angles made by inner faces begin as obtuse and become smaller as order increases.

Proposition 3.3.8 *G_2 maps inner faces along the partition line to inner faces of the tangent line on T_1, preserving order. The vertices and edges which lie precisely on the partition line are also mapped by G_1 to the vertices and edges of the inner faces lying along the tangent line on T_1.*

Proof See Figure 52 again. The inner faces along the partition tangent line at T_0 are in region A_2. G_2 in the standard scene is conjugate to K_2 in the isometric scene which is a clockwise rotation through angle θ. Therefore G_2 maps inner faces of order n along the T_0 tangent line to inner faces of order n along the T_1 tangent line. For example, face $ehgd$ of order m is mapped to face $G(e)G(h)G(g)G(d)$, also of order m,

while face $fihe$ of order $m-1$ is mapped to face $G(f)G(i)G(h)G(e)$ of order $m-1$.

If (x,y) is a point on the partition line then $(x,y) = \left(u, 1-\frac{1-\alpha}{\alpha}u\right)$ and

$$G_1(x,y) = \left(\frac{(\alpha-1)u+\alpha}{\alpha}, 1\right) = G_2(x,y).$$

Thus, under G_1, the points d, e, f are mapped as $G_1(d) = G_2(d) = G(d)$ and so on, and likewise for the edges de, ef etc., they are mapped by G_1 to corresponding edges along the T_1 tangent line. So, for example, if de is the edge of an inner face of order m along the T_0 tangent, then $G_1(de) = G(d)G(e)$ will be the face of an inner edge of order m along the T_1 tangent.

Now that we know what happens to the inner faces along the partition line under the action of G_2, It remains to explain what happens, under the action of G_1, to the *outer* faces along the partition line. We will develop this theme and then prove a proposition in this regard.

Because of Proposition 3.3.8, the edges of these outer faces must be mapped to corresponding edges along the T_1 tangent line, but where the rest goes needs to be determined by a different method.

The matrix of the G_1 map in (20) can be factored into elementary transformations of reflection, scaling and shear:

$$\begin{pmatrix} 0 & 1 \\ \frac{1-\alpha}{\alpha} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1-\alpha}{\alpha} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \Phi\Theta\Xi = \Upsilon\Xi$$

First comes Ξ, which is a reflection over the diagonal line $y = x$. Then there is Θ, a scaling along the y-axis, followed by Φ, a shear along the y-direction. Let us combine the y-scaling and y-shear, $\Phi\Theta$, into one matrix Υ. Note that Υ does not affect the x-coordinates of the points it acts on.

Figure 53 illustrates the action of the decomposition $G_1 = \Upsilon\Xi$. Attention is restricted to the outer faces along the partition line in the triangle $\triangle NSX$ of region Δ_3. The transformation Ξ reflects these outer faces over the diagonal line, resulting in outer faces with edges along the anti-partition line in $\triangle MRX$. Then Υ scales and shears these faces so that their edges are along the T_1 tangent line. Although they no longer faces according to the definition, the points in these images under Υ are all inner points.

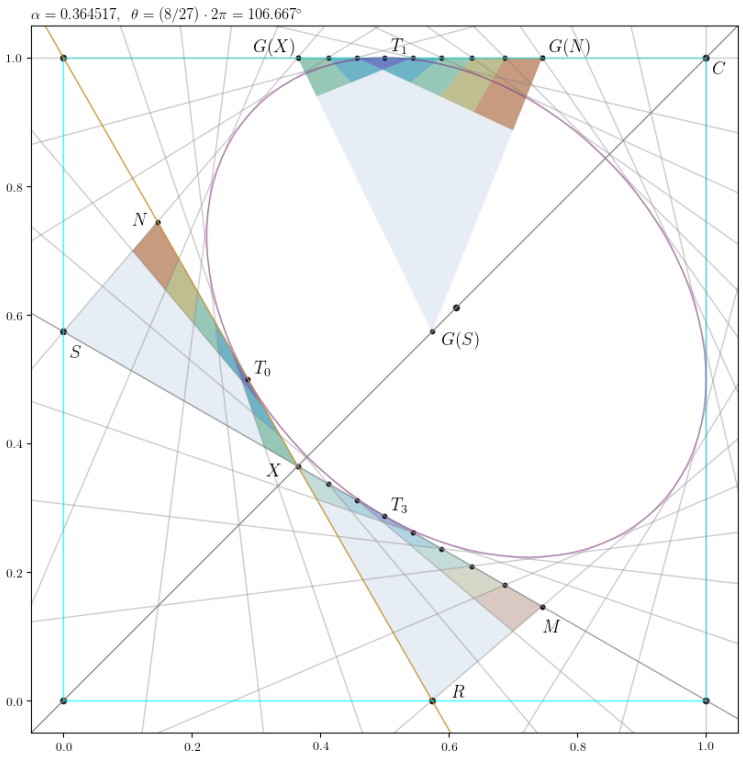

Figure 53: Decomposition of the G_1 matrix: the case where $G(S)$ is inside E_m.

By Proposition 3.3.8, Υ maps X to $G(X)$ and M to $G(N)$, and every edge of the outer faces along XM is mapped to an edge along $G(X)G(N)$ by projection in the y-direction. The endpoints of these edges and their images under Υ are denoted by black dots in Figure 53.

Proposition 3.3.9 *Suppose x is in an outer face of order m along the partition line in $\triangle NXS$. Then either G_1 maps x into a face of order $m-1$ along $G(X)G(N)$, or G_1 maps x elsewhere to a face of order $m-1$ or lower, or G_1 maps x into the maximal polygon Π_m.*

Proof Referring to Figure 53, the x-coordinate of R must always lie in between the x-coordinates of X and M, i.e., $X_x < R_x < M_x$. This is evident because of the way $\triangle NSX$ is constructed by tangent lines: $X_y < S_y < N_y$. Therefore, since Υ does not affect x-coordinates, it will map face angles $\angle T_3XR$ and $\angle T_3MR$ to acute angles $\angle T_1G(X)G(S)$ and $\angle T_1G(N)G(S)$. The images of the outer faces along XM under action of Υ are no longer faces, but the relationship between the angles of these images are preserved by Υ.

From Proposition 3.3.7, outer face angles are smaller as faces approach T_3, the same will be true for the angles made by their images under Υ with the T_1 tangent line. Then, since $\angle T_1G(X)G(S)$ and $\angle T_1G(N)G(S)$ are acute, all angles made by the other boundary lines of the images under Υ will also be acute.

Next we establish that the angles of the inner faces along $G(X)G(N)$ are always obtuse. An example can be seen in Figure 53, though one must ignore the images under Υ and look at the inner faces proper, along the BC tangent line. These angles are evidently obtuse in the figure, but we must explain why it is always so.

The leftmost of these inner faces makes an angle $\angle T_1G(X)N$ with the tangent line at T_1. Note that points S, N, $G(X)$ are collinear because they are all along a tangent line. This angle is smallest when N_x approaches S_x, (Figure 54, left) which happens as $\theta \to \frac{\pi}{4}$, even so it is bounded by $\frac{\pi}{2}$. Therefore, from Proposition 3.3.7, all inner faces along $G(X)T_1$ must make obtuse angles with the tangent line at T_1. Now for the inner faces to the right of T_1, along $T_1G(N)$. It is clear that, along the T_1 tangent line, the smallest angle that an inner face makes with this line is the one at C, which is $\frac{\pi}{2}$. Therefore again by Proposition 3.3.7, all the inner faces along $T_1G(N)$ must make obtuse angles with the tangent line at T_1. Thus all inner faces along $G(X)G(N)$ make obtuse angles.

Suppose $G(S)$ is inside E_m. Refer to Figure 53 for this case. Let F be an outer face along XM of order m, and let $x \in F$. Let $F' = \Upsilon F$ be the image of F under Υ. From Proposition. 3.3.8 we know that the edge of F which is along XM and its endpoints (black dots in the figure), are mapped by Υ precisely to the edge and endpoints of an inner face along $G(X)G(N)$. Therefore the image F' shares the edge and endpoints along $G(X)G(N)$ with an inner face of order $m-1$. However, we argued in the previous paragraphs that the inner faces along $G(X)G(N)$ are angled obtusely with respect to the T_1 tangent line, while the image F' is angled acutely. Thus there is no way for point x to reach a face of order m or greater. Therefore x must be sent by Υ either into a face of order $m-1$ or lower, or into the maximal polygon Π_m.

Finally we must rule out any possibility for the point x to escape to a face of order $\geq m$ by some other way. One such way is when $G(X)$ is outside E_m, along the diagonal $y = x$ between X and E_m (Figure 54, left). The outer face along XM with largest order, say m, is the one having R as a vertex. But $\Upsilon R = G(S)$ which falls in

a region $\triangle T_0 X T_3$ where no face is higher order than $m-1$. Furthermore, points of faces of order lower than m cannot move into a face of order m in this region because the images are angled acutely.

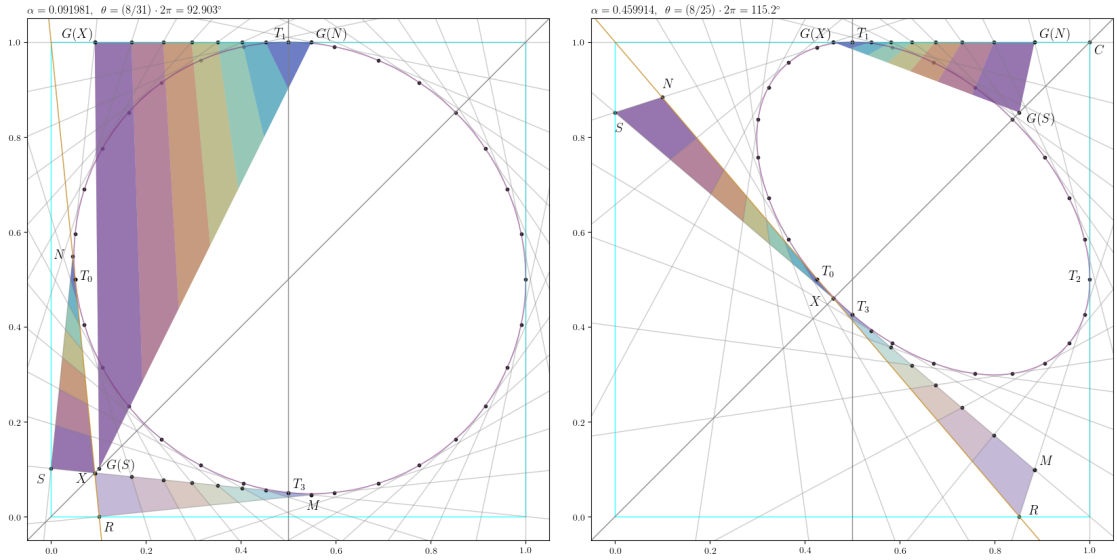

Figure 54: Left: $G(S)$ is along diagonal $y = x$ between X and E_m. Right: $G(S)$ is along diagonal $y = x$ between E_m and bounding box point C.

Similar considerations are made when $\Upsilon R = G(S)$ falls in the region outside E_m in between E_m and C (Figure 54, right). The outer face having R as a vertex is order m, but here $G(S)$ falls in a face of order $m-3$, and since all images are angled acutely, no point of an outer face along XM can ever reach a face of equal or higher order than it originally had.

We have completed the study of G with π-rational rotation angle, and we have proved the propositions we need for the main theorem. Now we turn to the π-irrational case. We will present some numerical studies which point to the right idea behind the behavior of orbits. Once this numerical phenomena is described and understood, it will point us toward the main theorem proof for the π-irrational case.

Figure 55 features limit sets for two π-irrational angles. Limit sets are analogous to Mandelbrot sets: a point x_0 is chosen and iterated under G. The number of iterations required for x_0 to fall into E_m is recorded, up to a specified limit, say 10^5. The data is binned into a 2D histogram, from which an image is made. Colors and brightness indicate how long it takes for points in those regions to fall into E_m. Black

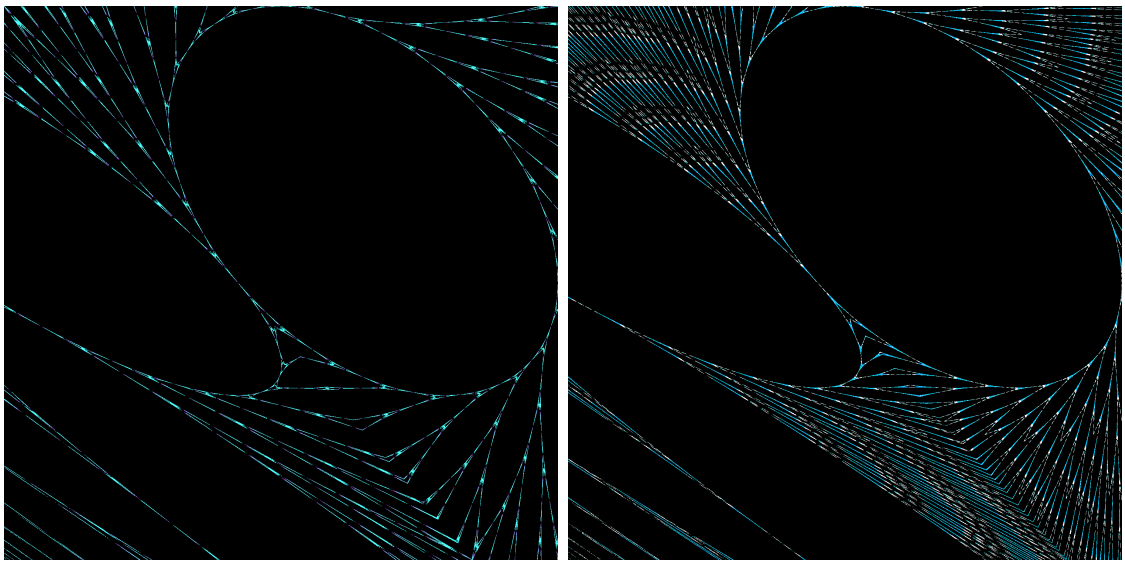

Figure 55: Limit sets. Left: $r = 1/\sqrt{9.8}$. Right: $r = 1/\sqrt{9.5}$.

areas take relatively few iterations. Bright areas may require thousands. Some close ups of details in limit sets are shown in Figure 56.

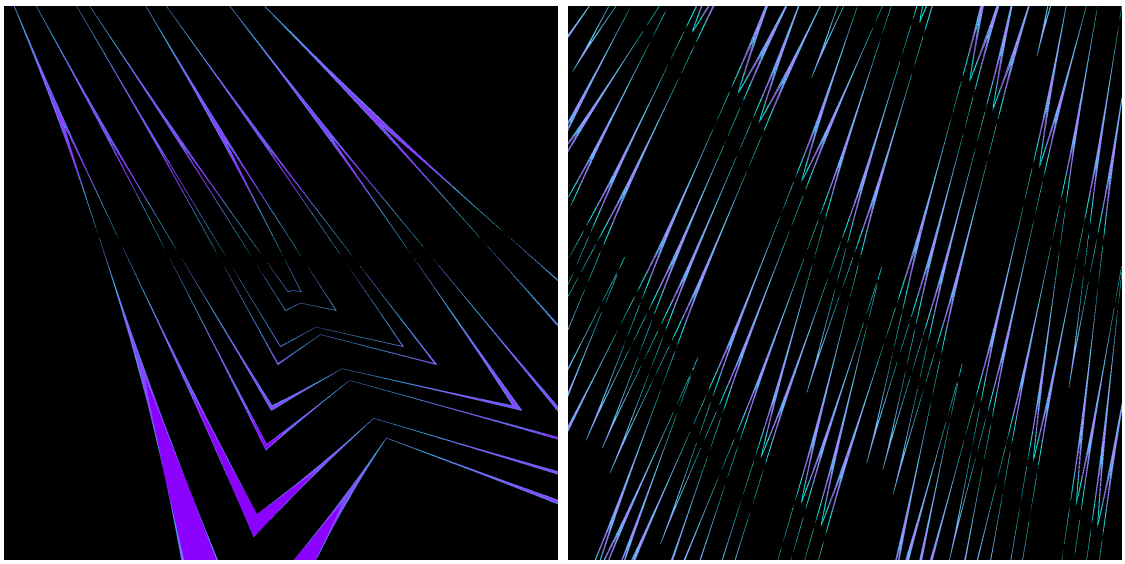

Figure 56: Zoom on limit sets. Left: $r = 1/\sqrt{9.8}$. Right: $r = 1/\sqrt{9.3}$.

The limit sets in Figures 55 and 56 were constructed from 400,000,000 random initial points. Iterations were cut off above 10^5 and below 50 to 100. The histograms had 1500×1500 bins. In the left image $\theta = r \cdot (2\pi) \approx 115°$ and on the right, a slightly larger angle $\theta = r \cdot (2\pi) \approx 116.8°$. We see branching fractal patterns. These patterns are thin and scarce when $\theta \to 90°$, but densely fill up the space in between

the preimages of E_m (the dark ellipse-like regions) as $\theta \to 120°$. The closer θ is to $120°$, the easier it is to find points that take a very long time to fall into E_m.

This leads us to make conjectures about the high iteration points. Here is a rough sketch. A particle (as it can be imagined) which takes a very long time to fall into E_m must spend most of its life in region A_2 above the partition line. If the particle is very close to E_m, it can rotate around A_2 under the action of G_2 and avoid entering A_1 for a long time. When it does enter A_1, it will do so near the point T_0, as this is the only "entry point" for a particle which is orbiting very close to E_m. By Proposition 3.3.3, the distance to E_m will shrink and by Proposition 3.3.4 the particle returns to A_2 where it circulates again for very many iterations of G_2. And so on.

This will be proved in the main theorem, Theorem 3.3.10, but it is illuminating to first examine some numerical evidence.

By sifting though the data behind the graphics in Figures 55 and 56, points which take very many iterations to enter E_m can be found. Table 1 is a summary of the orbit of one such point. Data is only recorded when the point enters A_1, we ignore iterations in A_2, although we keep track of the total number of iterations in the right-most column. From the table, it is clear that most of the iterations are spent in A_2—the point only rarely enters A_1. Between the last two appearances in A_1, the point had spent more than 240,000 iterations in A_2. This is consistent with our conjectures. Each time the point enters A_1 we also record the distance to T_0. It can be seen in the table that this distance decreases with every visit to A_1: again consistent with our conjectures.

Another example is given in Table 2, This time θ is closer to $120°$. As mentioned previously, the $I \times I$ region fills up with high-iteration points when $\theta \to 120°$. We expect it to be easy to find points requiring a million or more iteration. Out of more than 2.7 million iterations, this particular point visited A_1 only 13 times. Each time, its distance to T_0 became smaller, as expected. Again, this is consistent with our conjectures about π-irrational dynamics.

Now that we have a good grasp of what seems to be going on for the π-irrational case, we move to the main theorem of this chapter.

Theorem 3.3.10 *When θ is π-rational, then every point $x \in I \times I \setminus \{(0,0)\}$ goes into Π_m. When θ is π-irrational, then for every point $x \in I \times I \setminus \{(0,0)\}$, either some iterate of x falls into E_m or the orbit $G^n(x)$ is accumulated on E_m.*

x	y	Distance from T_0	Iteration
0.71495708668	0.07676770558	0.51440070071	0
0.08696440497	0.78310688511	0.43907797182	4
0.41764247718	0.50582381296	$0.76372288813 \times 10^{-2}$	661
0.41916457850	0.50403506482	$0.52885282662 \times 10^{-2}$	805
0.41992607811	0.50314242717	$0.41152153853 \times 10^{-2}$	877
0.42067391038	0.50225092744	$0.29516021389 \times 10^{-2}$	949
0.42218447586	0.50047143696	$0.61742335748 \times 10^{-3}$	1093
0.42244562726	0.50016269733	$0.21304703524 \times 10^{-3}$	3372
0.42258458756	0.49999832526	$0.21926923008 \times 10^{-5}$	244561

Table 1: The approach to T_0 for $r = 1/\sqrt{9.8}$.

Proof Refer to Figures 47 and 48. Points in $Q \setminus \Delta_2$ either go into Π_m or they return to Δ_2. Points in Δ_2 are mapped into Δ_3. Therefore it is sufficient to consider only Δ_3. Faces in $\Delta_3 \cap A_2$ will eventually return by rotation to overlap a face of the same order in $\Delta_3 \cap A_1$. Therefore it is enough to consider faces in $\Delta_3 \cap A_1$, which is the triangle $\triangle NSX$.

Suppose that x is in a face in $\triangle NSX$ of higher order than the ones having edges along NX. Then by Proposition 3.3.3, x will either move strictly closer to E_m on the next iteration, or go into E_m. Therefore eventually x will either (a) go into a face of order m which can be rotated to coincide with a face of order m along NX or (b) go into Π_m. Suppose (a). When this face is brought to coincide with a corresponding face along NX, then by Proposition 3.3.9, x will, on the next iteration, end up in a face of order $m-1$ or less, or inside Π_m. This continues until x either goes in a face of order 1, or goes in Π_m. Suppose x is now in a face of order 1. Then by rotation, this face will eventually coincide with the unique order 1 face along NX, and in the next step, by Proposition 3.3.6, x will go into Π_m. Therefore all points go into Π_m.

We now consider the case where θ is π-irrational. The movement function D, (24), is strictly positive in A_1 except at the origin and on the partition line. Since D is continuous, this means that the orbit points $G^n(x)$ cannot accumulate in the interior of A_1. Thus, the orbit points $G^n(x)$ either eventually stay in A_2, or they switch between A_1 and A_2 infinitely often. If, starting from some iteration, they stay forever in A_2, then they have fallen in the closure of E_m. If they switch infinitely many times

x	y	Distance from T_0	Iteration
0.13449007107667	0.42243560817160	0.35415528931041	0
0.43979753159904	0.54025659355535	$0.56926528793746 \times 10^{-1}$	4
0.46134295108585	0.51882765202083	$0.26539209923934 \times 10^{-1}$	7
0.47856068143989	0.50154705979212	$0.21454956821362 \times 10^{-2}$	101
0.47888425518990	0.50121073972254	$0.16787933708968 \times 10^{-2}$	1748
0.47909954150914	0.50098664182530	$0.13680393426402 \times 10^{-2}$	2846
0.47931466394059	0.50076263641774	$0.10574655472880 \times 10^{-2}$	3944
0.47952977570801	0.50053872352893	$0.74696586389585 \times 10^{-3}$	5042
0.47974471722800	0.50031490305058	$0.43665076179508 \times 10^{-3}$	6140
0.47995965486046	0.50009117501756	$0.12640502515786 \times 10^{-3}$	7238
0.48004619512294	0.50000105375159	$0.14607991662494 \times 10^{-5}$	2538784
0.48004678503418	0.50000043932247	$0.60902547823245 \times 10^{-6}$	2638699
0.48004737494356	0.49999982489409	$0.24274638491963 \times 10^{-6}$	2738614

Table 2: The approach to T_0 for $r = 1/\sqrt{9.3}$.

between A_1 and A_2, then the iterates $G^n(x)$ which are in A_1, must accumulate at T_0, which is the only point both in E_m and on the boundary of A_1.

We have shown that, for the π-rational case, all points lying in faces of order $n \geq 1$ eventually visit A_1. Therefore by Lemma 3.3.5, the only periodic points are the points in faces of order 0, i.e., points on or in Π_m.

3.4 Appearance and growth of islands

We have seen that densities for G with π-irrational rotation angles are concentric ellipses supporting ACIMs on each ellipse. They are periodic in the case of π-rational rotation angles. The elliptical map with memory G is defined by $0 < \alpha = \beta < \frac{1}{2}$, but an interesting phenomenon occurs in the π-rational case when β is slightly perturbed.

If β is slightly increased, the periodic density does not remain confined to points. The support becomes two-dimensional islands, These islands begin very small, but grow in size as β increases.

For example, if $\theta = \frac{p}{q}2\pi$ then $\alpha = \frac{2\cos\theta}{2\cos\theta - 1}$ and we can try a perturbed β, $\beta = \alpha + \epsilon$, where ϵ is some small number. Making images of such perturbed densities will reveal

q islands which grow larger as ϵ is made larger. This phenomenon of island growth is illustrated in Figure 57.

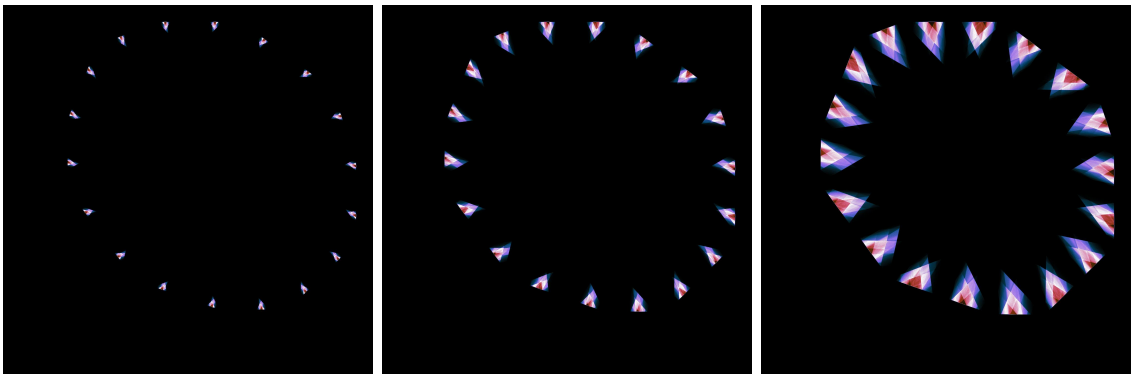

Figure 57: Growth of islands. Beginning with $\theta = \frac{5}{18}2\pi$, take $\beta > \alpha$ to be $\alpha + 0.003$, $\alpha + 0.006$, $\alpha + 0.01$.

Rather than the usual $\beta = \alpha$ for the elliptical map, we have taken $\beta = \alpha + 0.003$, etc., and this has resulted in the 18 islands seen in the images.

Using the combinatorial method of [9] and as described in Section 2.3 and in Theorem 2.3.2, it can be shown by the same method that the three island densities of Figure 57 support absolutely continuous invariant measures. Many other (but not all) island densities grown this way can be proved to support ACIMs by Theorem 2.3.2. Some close-ups of islands grown from point-periodic densities by perturbing β are shown in Figure 58. All of these islands support absolutely continuous invariant measures.

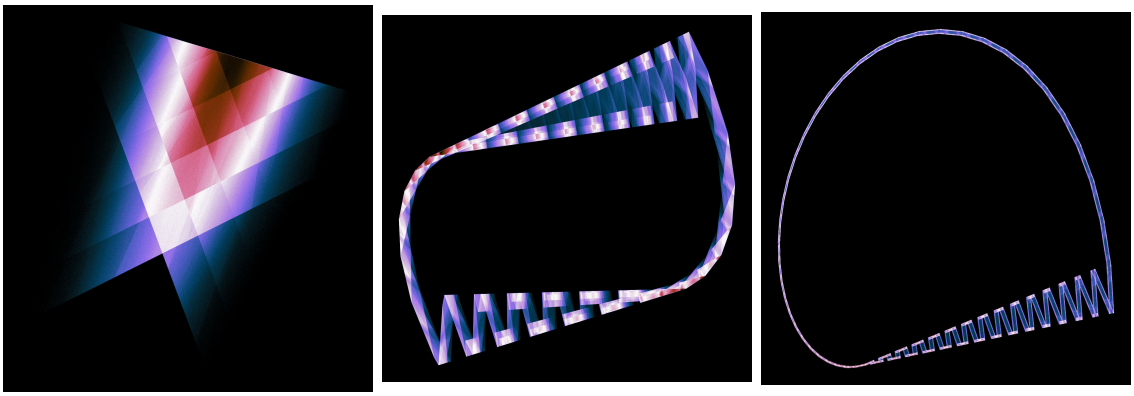

Figure 58: Islands. Left: $\theta = \frac{11}{42}2\pi$, $\beta = \alpha + 700 \times 10^{-6}$. Middle: $\theta = \frac{21}{65}2\pi$, $\beta = \alpha + 300 \times 10^{-6}$. Right: $\theta = \frac{36}{109}2\pi$, $\beta = \alpha + 20 \times 10^{-6}$.

The Nakamura-Mackey map S of [19] and Section 2.5 also exhibits the phenomenon of island growth. Figure 59 is an example.

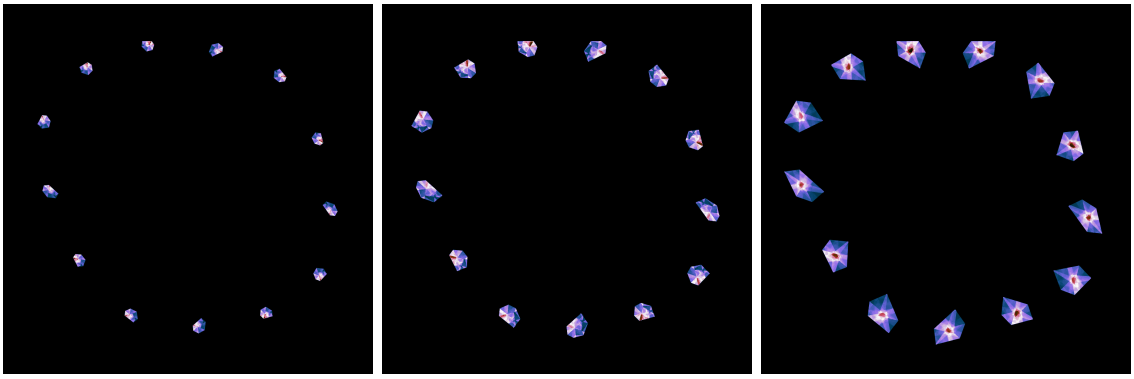

Figure 59: Island growth for Nakamura map. $a = 0.245$, $b = 1.025$ (left), $b = 1.03$ (middle), $b = 1.04$ (right).

Densities for the Nakamura map S in (18) were examined for $a = 0.245$ and $1.0 \leq b \leq 1.05$. For this range of b, a system of 13 islands grows from points or possibly tiny point-like islands. Using the combinatorial methods of Theorem 2.3.2, all of the island systems of Figure 59 can be shown to support ACIMs.

Nakamura-Mackey island growth is also associated with ellipses, but this topic is deferred to the epilogue in Chapter 4.

Chapter 4

Further Directions

The system of 448 islands at $\alpha = 0.4973$ was one of several island systems mentioned in [10]. As pointed out in [9], this system has 49 islands of region index 3. To prove the existence of an ACIM, it would be necessary to evaluate $2^{49} = 562{,}949{,}953{,}421{,}312$ possible region index trajectories and the corresponding singular values σ_2 of $D(G^{448})$, This is far beyond our present computing abilities. How to prove the existence of an ACIM for this case is an open question. Perhaps some refinement can be made to the combinatorial method of Theorem 2.3.2, possibly by grouping the islands in some way. Images of some islands in this system are shown in Figure 60.

We saw that entropy plots were a practical tool for discovering interesting phenomena in memory map dynamics. Another tool with good potential is the bifurcation plot, which is described in [17, prob.1.3]. To make a bifurcation plot, we return to a one dimensional picture of the memory map as described in Section 2.3. The horizontal axis of a 2D histogram is associated with values α_k, and a one dimensional orbit is computed for each α_k, filling the vertical bins of the histogram above α_k. This can be done by taking a 2D memory map G and recording only the y coordinates of the orbit points in the histogram bins above α_k. Figure 61 is a bifurcation map for (11) over a narrow interval of α centered exactly on $\alpha = 0.493$. The indication of island phenomena as described in Section 2.3 near $\alpha = 0.493$ is clearly visible as a heavy dark vertical band through the center of the image. There are other bands distributed in what appears to be regular patterns. Some bands are faint. Lying to the right, there is another vertical band of comparable prominence to the central one. This band lies near $\alpha = 0.49312328082020507$. A study of the entropy well for

this value of α reveals a system with 178 islands which support an ACIM. Images of some islands for this system are shown in Figure 62. Why there is regularity in the distribution of these bands is unknown.

Some observations on the appearances of islands. There are repetitions of themes for systems of islands that vary widely in α and island count. Note the similarity between the 175, 448 and 178 systems in Figures 19, 60 and 62. Then note the common theme shared between the systems in Figures 21, 26 and 30. There are systems for which all islands seem to be affine transformations of a single island. Both the Nakamura 22 and 31 systems studied in Section 2.5 are like this. For other systems, it's hard to avoid the impression that the islands are 3D objects being projected down into 2D, complete with shadows in some cases. For example, some (or all) of the islands in the 203 β-tent system of Figure 38 look like 2D projections of a single object which is being rotated in three dimensions.

Island systems also arise if we use other τ, such as the sine map $\tau(u) = \sin \pi u$ or the logistic map $\tau(u) = 4u(1-u)$. From these we can construct 2D memory maps G. Typical islands for these G look totally different from anything we have seen so far. An island from a sine map system of 7 islands is shown in Figure 63. Attempts were made to adapt singular value methods to prove that sine and logistic islands support ACIMs, with no success yet. It's an open question.

The sine map with memory exhibits an extaordinary phenomenon of island fission, i.e., one island which splits into three. An island fission sequence is shown in Figure 64. It happens over a range of α between $\alpha_1 = 0.126934$ to $\alpha_2 = 0.1269395$ in a tiny window centered on $(0.5308, 0.2829)$ with width $\Delta x = 0.007$ and height $\Delta y = 0.0021$. The image frames were chosen from this range. Other baffling phenomena can be seen in this interval of α.

Figure 65 illustrates logistic map islands from systems with 5 and 6 islands respectively, at $\alpha = 0.1278$ and $\alpha = 0.10907$. The islands in these systems are of the affine type: each island is an affine transformation of some reference island. The sine map system of Figure 63 is also affine. This shows that the phenomenon of affine islands seems to be universal across many types of τ. Why this is so, and what it means, is unknown.

The Nakamura-Mackey map (18) produces ellipses for certain values of a and b. This is seen in Figure 66 (left), which is a density made by choosing $a = 0.245$ and

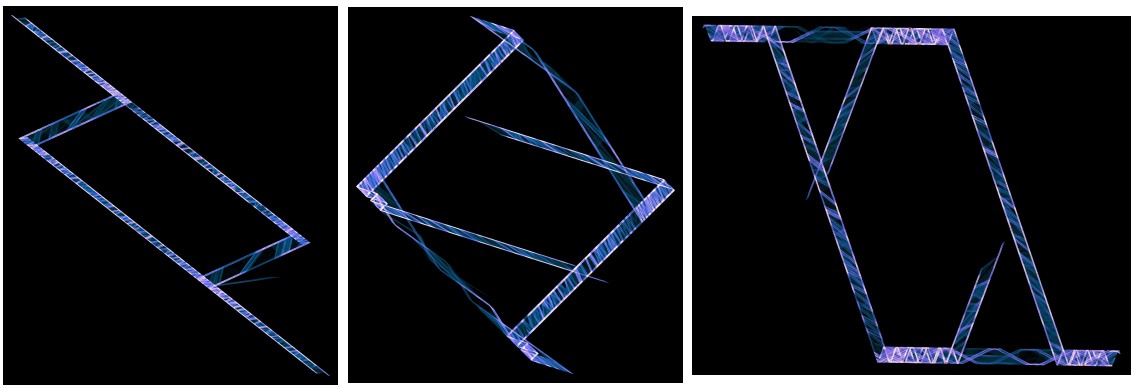

Figure 60: Islands 161, 291 and 346 in the 448 island system at $\alpha = 0.4973$.

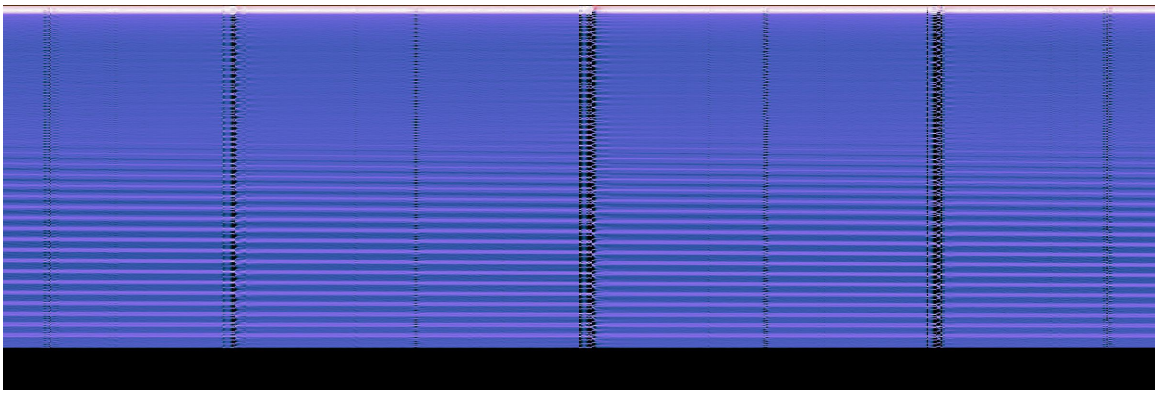

Figure 61: Bifurcation plot for tent memory map. Narrow α-interval centered on $\alpha = 0.493$.

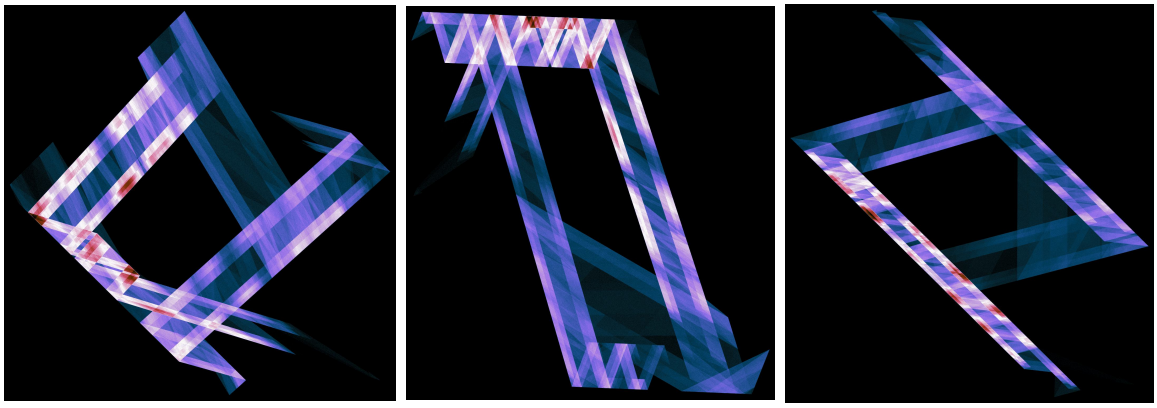

Figure 62: Islands 20, 39 and 92 from the 178 island system lying to the right of center in Figure 61, $\alpha = 0.49312328082020507$

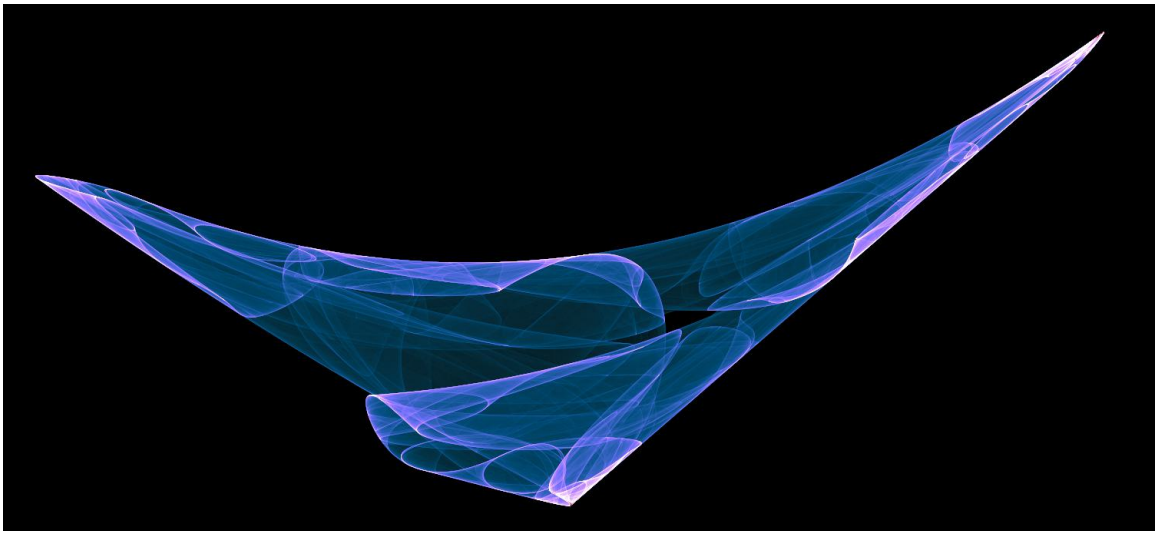

Figure 63: Sine map with memory island, $\alpha = 0.1269$.

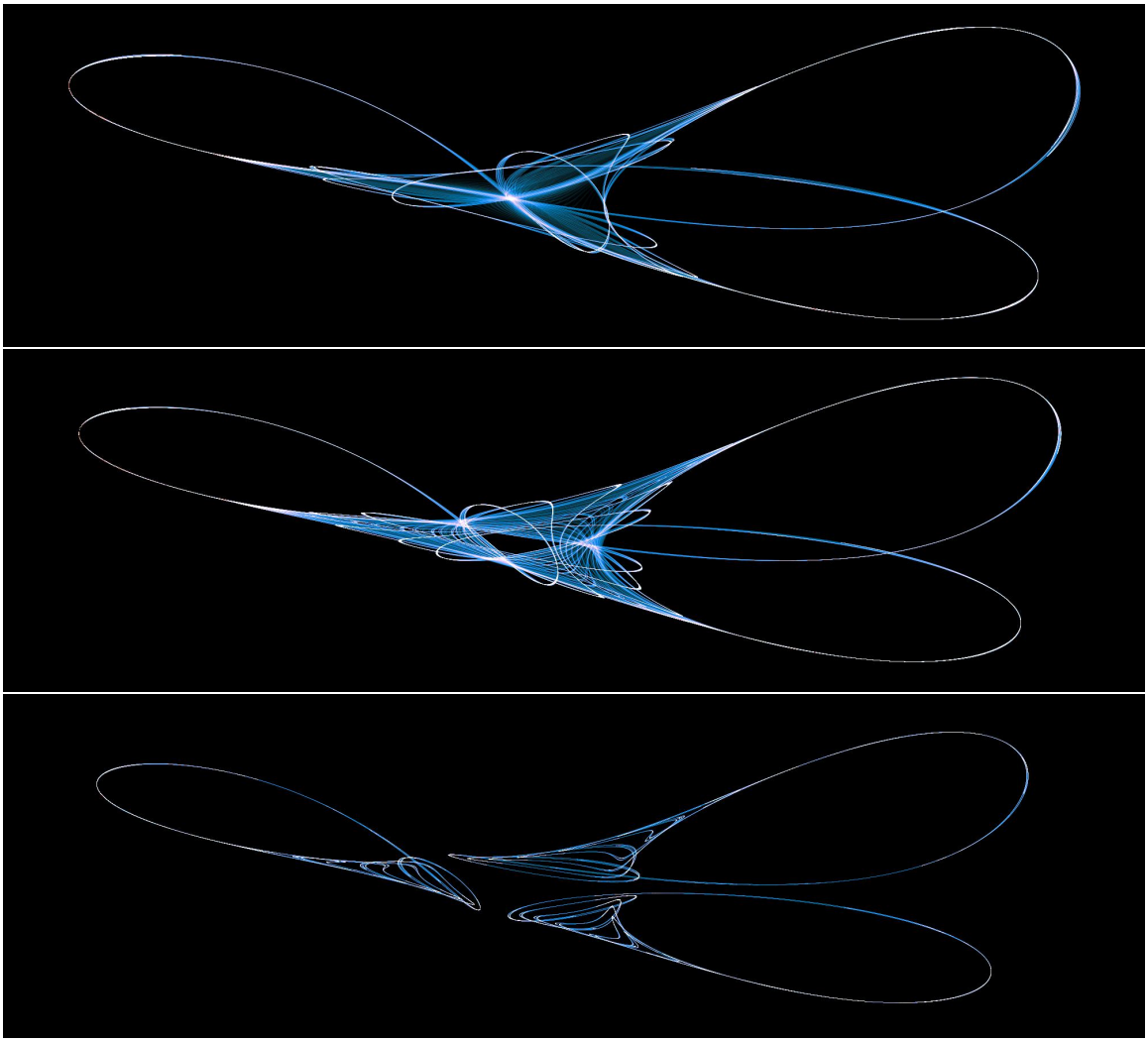

Figure 64: Island fission for sine map with memory, $0.126934 < \alpha < 0.1269395$.

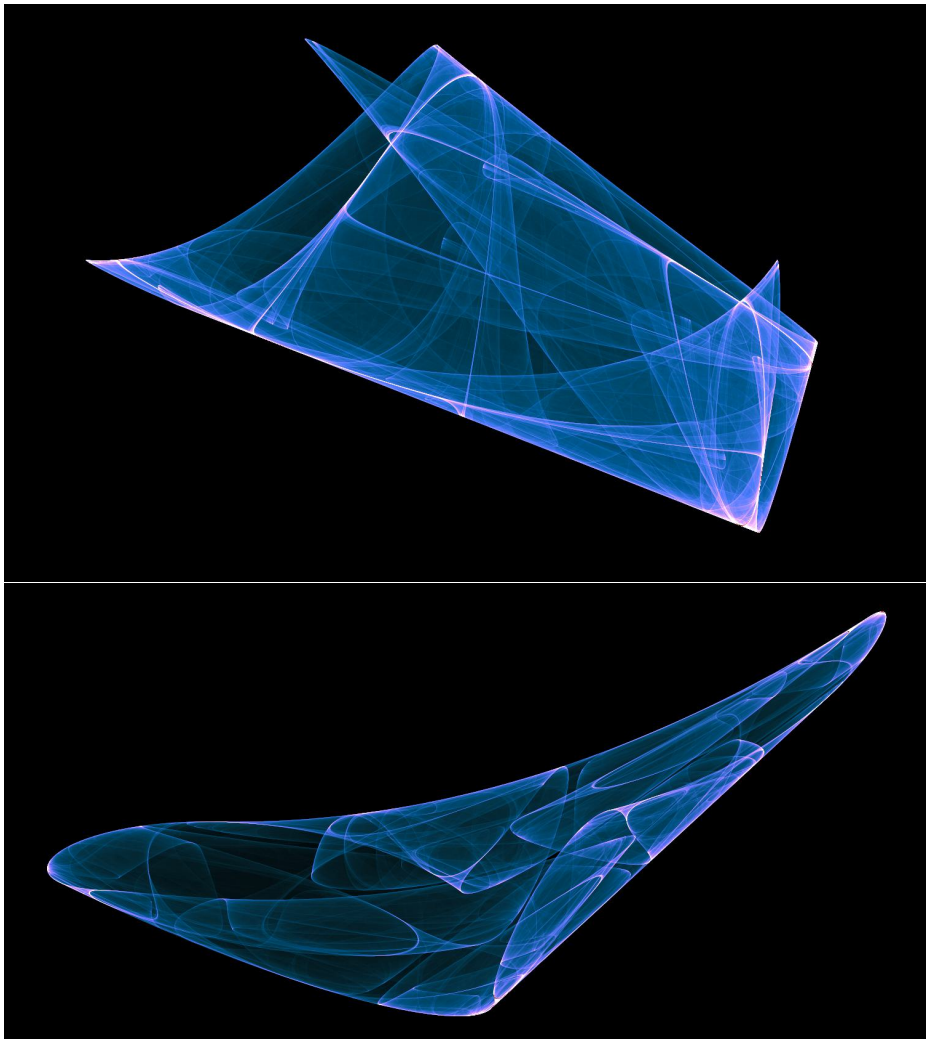

Figure 65: Logistic islands. Top: $\alpha = 0.1278$. Bottom: $\alpha = 0.10907$.

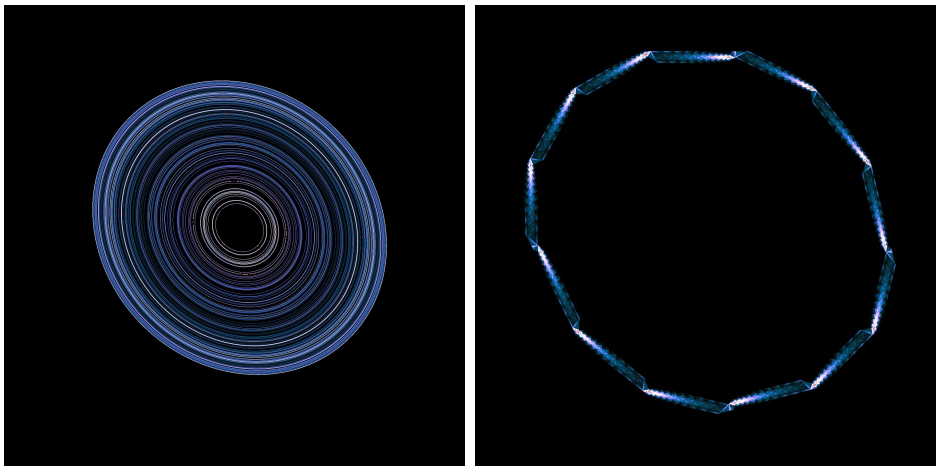

Figure 66: Ellipse phenomenon for Nakamura map, $b = 1.0$, $b = 1.01$.

$b = 1.0$ with 200 random points and allowing their orbits to settle. They settle into concentric ellipses reminiscent of the elliptical memory maps of Chapter 3.

When b is increased slightly, the orbits quickly move into "maximal" position, expanding into the domain up to the parition line, which is visible in Figures 39 and 41. The orbits go through some remarkable changes before becoming, at $b = 1.01$, polygonal with 13 sides. These sides reveal incredibly complex and beautiful details if imaged at very high resolution. When b is increased a little more, we have a regime of island growth, as pictured in Figure 59. This leads to some final questions. Are ellipses universal for certain types of piecewise linear 2D maps? Can the geometrical analysis of Chapter 3 be extented to the Nakamura map?

Bibliography

[1] Kathleen T. Alligood, Tim D. Sauer, and James A. Yorke. *CHAOS: An Introduction to Dynamical Systems*. Springer, 1996.

[2] Abraham Boyarsky and Pawel Góra. *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension*. Probability and Its Applications. Birkhauser, 1997.

[3] Jerome Buzzi. Absolutely continuous invariant measures for arbitrary explanding piecewise $\mathbb{R}$-analytic mappings of the plane. *Ergodic Theory and Dynamical Systems*, 30(3):697–708, 2000.

[4] Geon Ho Choe. *Computational Ergodic Theory*. Algorithms and Computation in Mathematics. Springer, 2005.

[5] Freeman Dyson. Birds and frogs. *Notices of the American Mathematical Society*, 56(2):212–223, 2009.

[6] P. Góra and A. Boyarsky. Why computers like lebesgue measure. *Computers & Mathematics with Applications*, 16(4):321–329, 1988.

[7] P. Góra and A. Boyarsky. Absolutely continuous invariant measures for piecewise expanding c^2 transformations in $\mathbb{R}^n$. *Israel Journal of Mathematics*, 67(3):272–286, 1989.

[8] P. Góra, A. Boyarsky, and Z. Li. Singular srb measures for a non 1–1 map of the unit square. *Journal of Statistical Physics*, 165(2):409–433, 2016.

[9] P. Góra and T. Szylowiec. Islands supporting acim for two-dimensional map. *International Journal of Bifurcation and Chaos*, 32(8), 2022.

[10] Pawel Gora, Abraham Boyarsky, Zhenyang Li, and Harald Proppe. Statistical and deterministic dynamics of maps with memory. *Discrete and Continuous Dynamical Systems*, 37(8):4347–4378, August 2017.

[11] Dan Kalman. A singularly valuable decomposition: The svd of a matrix. *The College Mathematics Journal*, 27(1):2–23, 1996.

[12] J. Komornik and A. Lasota. Asymptotic decomposition of markov operators. *Bull. Polish Acad. Sci. Math.*, 35:321–327, 1987.

[13] Jozef Komornik. Asymptotic periodicity of the iterates of weakly constrictive markov operators. *Tohoku Mathematical Journal*, 38(2), 1986.

[14] Jozef Komornik. Asymptotic periodicity of markov and related operators. 2:31–68, 1993.

[15] A. Lasota, T. Y. Li, and J. A. Yorke. Asymptotic periodicity of the iterates of markov operators. *Transactions of the American Mathematical Society*, 286(2), 1984.

[16] A. Lasota and J. A. Yorke. On the existence of invariant measure for piecewise monotonic transformations. *Transactions of the American Mathematica Society*, 186:481–488, 1973.

[17] Andrzej Lasota and Michael C. Mackey. *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics*. Applied Mathematical Sciences. Springer, 2nd edition, 1994.

[18] Tien-Yien Li and James A. Yorke. Period three implies chaos. *American Mathematical Monthly*, 82(10):985–992, 1975.

[19] Fumihiko Nakamura and Michael C. Mackey. Asymptotic (statistical) periodicity in two-dimensional maps. *Discrete and Continuous Dynamical Systems B*, 27(8):4285–4303, 2021.

[20] I. P. Natanson. *Theory of Functions of a Real Variable*, volume 1. Frederick Ungar, revised edition, 1961.

[21] H.L. Royden and P.M. Fitzpatrick. *Real Analysis*. Prentice Hall, 4th edition, 2010.

[22] Michael Spivak. *Calculus on Manifolds: a Modern Approach to Classical Theorems of Advanced Calculus*. Addison-Wesley, 1965.

[23] Gilbert Strang. The fundamental theorem of linear algebra. *American Mathematical Monthly*, 100(9):848–855, Nov. 1993.

[24] M. Tsujii. Absolutely continuous invariant measures for piecewise real-analytic expanding maps on the plane. *Commun. Math Phys.*, 208:605–622, 2000.