

Homology of the Gray-Scott Model of Reaction-Diffusion

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Joel E. Hawkins

May 2015

Approved for the Division
(Physics)

Daniel Borrero

Table of Contents

| | |
|--|-----------|
| Introduction | 1 |
| Chapter 1: Reaction-Diffusion Systems | 3 |
| 1.1 The Gray-Scott Model | 3 |
| 1.2 Numerical Solution | 5 |
| Chapter 2: Computational Homology | 9 |
| 2.1 Image analysis | 9 |
| 2.2 Cubical Homology | 12 |
| Chapter 3: Methods and procedures | 21 |
| 3.1 Obtaining Betti numbers | 21 |
| 3.1.1 Thresholding | 21 |
| 3.1.2 Performing cubical homology | 23 |
| 3.1.3 Calculating entropy | 23 |
| Appendix A: The First Appendix | 27 |
| A.1 Extra definitions, theorems, etc. | 27 |
| Appendix B: Code | 29 |
| B.1 Gray-Scott simulation | 29 |
| References | 31 |

Introduction

The study of pattern formation is incredibly diverse and certainly one of the most compelling aspects of nonlinear phenomenology. Scientists from many disciplines have analyzed patterns on the scale of the entire universe all the way down to the microscopic¹. Just a cursory glance at the structure of a wind-swept sand dune, a snowflake, or even our own galaxy reveals something interesting. Observation of these patterns might lead a scientist to ask what causes the pattern and wonder why there are patterns at all. This question gets complicated quickly because whether it's God in the patterns or the result of a non-equilibrium universe, there is still the question of what it means to have "structure" or "complexity" or to be "interesting".

Fortunately, the theoretical model of pattern formation discussed here allows me to sidestep all of these questions while focusing primarily on the method of analysis. There are many mathematical tools available to help interpret image data but as the complexity of our information (i.e. size of our data) increases, it becomes increasingly difficult to parse relevant information. Take for example the Fourier transform, a powerful method for image analysis. We would expect the Fourier transform to provide some insight into the spatial frequency of the image². Examine the two distinct pattern types of the Gray-Scott system shown in Figure 1. The two pattern types α and κ (figures 1a and 1b respectively) are visually extremely distinct yet the Fourier transforms (figures 1c and 1d) are disappointingly similar. Although this method could certainly extract useful information, we would like some way to supplement our findings. The need for new methods arises and many times that means starting at the lowest level. The homology theory described later on considers the fundamental geometric structures of an image while providing insight into global properties.

¹See the introduction of Cross & Greenside for an overview of the study of pattern formation.

²In this case, the Fourier transform converts from the *spatial domain*, the image we see, to the *frequency domain*. The Fourier transform is often used to remove noise or apply filters to images. It has also been used for intricate pattern recognition, see [1].

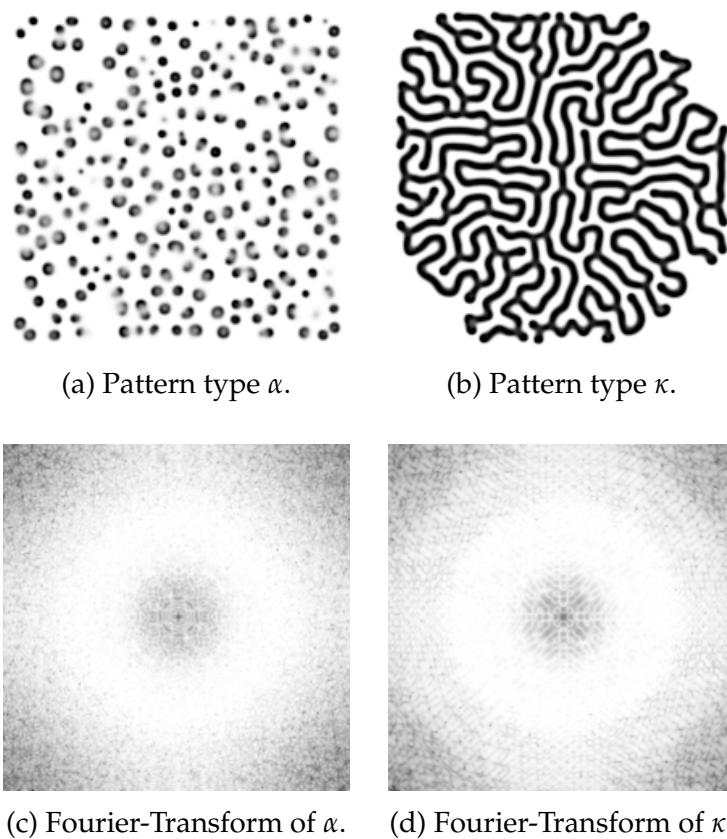


Figure 1: For two very distinct pattern types α and κ , the Fourier-Transform is visually very similar and extracting meaningful information is difficult.

Chapter 1

Reaction-Diffusion Systems

Reaction-diffusion (RD) systems are models that determine how concentrations of chemical species change in space. These systems are driven by two processes: chemical reaction and spatial diffusion. RD systems are governed by partial differential equations, the most basic of which might look something like

$$\frac{\partial u}{\partial t} = d\nabla^2 u + r(u). \quad (1.1)$$

This is sometimes called the Kolmogorov-Petrovsky-Piskounov equation in which u is a generic chemical species, d is a diffusion coefficient, $\nabla^2 u$ is the Laplace operator, and $r(u)$ is a general reaction term.

RD systems are interesting because their solutions can show wide variety of patterns, many of which resemble patterns of nature such as spirals, stripes, and spots. In 1952, Alan Turing suggested that RD systems of morphogens, chemicals that govern the pattern of tissue development, may be able to explain the presence of spots or stripes on an organism [2]. Although the science behind animal patterns is more complicated, Turing laid the framework by which patterns form from minor perturbations of otherwise homogenous systems. Since then, many others have noted the similarity between RD patterns and patterns in nature [3, 4, 5, 6, 7, 8, 9].

1.1 The Gray-Scott Model

One important model in the study of pattern formation is the Gray-Scott system which models the reaction of two generic chemical species, U and V [10]. The

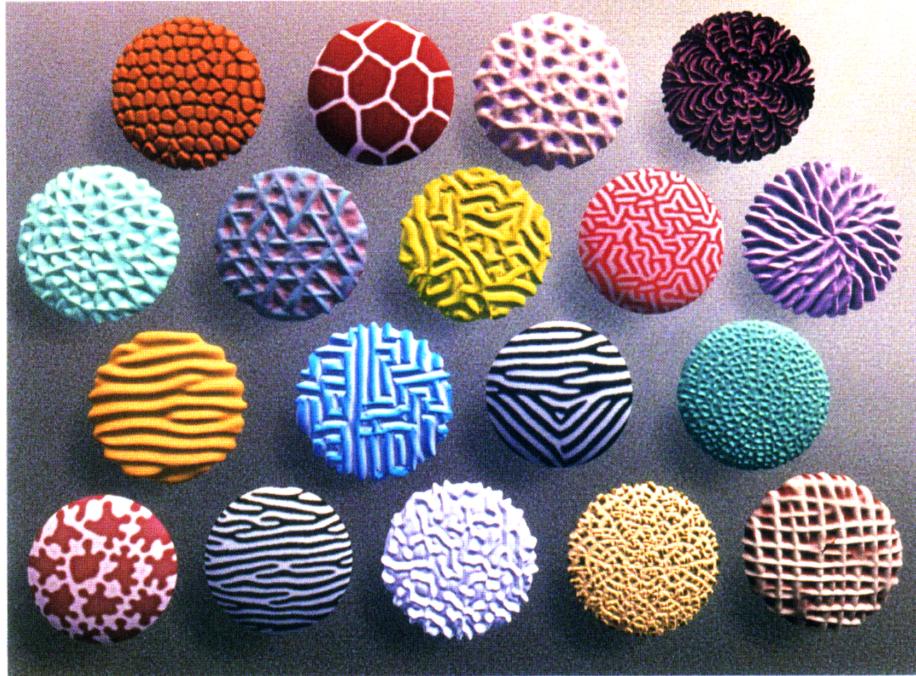


Figure 1.1: Patterns generated by reaction diffusion systems. Witkin compares these patterns to those of nature [9]; “Row 1: reptile, giraffe, coral, scalloped. Row 2: spiral, triweave, twisty maze, replication, purple thing. Row 3: sand, maze, zebra haunch, radial. Row 4: space giraffe, zebra, stucco, beats us.” Adapted from A. Witkin and M. Kass, “Reaction-diffusion textures,” *SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 299–308, 1991.

model is based on the chemical reaction in (1.2).



V is converted to inert product, P , which doesn’t interfere with the reaction of the system. V appears on both sides of the chemical reaction and thus catalyzes its own production. Gray and Scott developed the following non-dimensional PDE in which u and v represent the concentrations of chemicals U and V respectively.

$$\frac{\partial u}{\partial t} = d_u \nabla^2 u - uv^2 + F(1 - u) \tag{1.3}$$

$$\frac{\partial v}{\partial t} = d_v \nabla^2 v + uv^2 - (F + k)v \tag{1.4}$$

We see that both equations take the form of (1.1) only they are coupled. For simplicity, d_u , d_v , F , and k are taken to be constants. The first term in each equation,

$d_u \nabla^2$ and $d_v \nabla^2$, are the diffusion terms. The Laplace operator, ∇^2 , is responsible for the diffusion of each chemical in space (like the diffusion of heat in the more familiar heat equation) while the *diffusion coefficients*, d_u and d_v , govern the diffusion rate. The $\pm uv^2$ terms are the *reaction terms* which convert U into V : an increase in v is equal to the decrease in u , hence $+uv^2$ in the second equation. Since U will eventually get used up to generate V , the term $F(1 - u)$ is the *replenishment term* which reintroduces chemical U into the system (u has a maximum value of 1). Similarly, chemical V would increase without limit except for the *diminishment term*, $(F + k)v$, which serves to remove chemical V from the system. F is referred to the *feed rate* and determines the rate of replenishment while k is the difference between this rate and that of chemical V .

For some biological intuition, one can imagine the chemical reactions that occur in the development of an embryo as Turing theorized. In this case, the supply of chemicals might be the bloodstream where the replenishment and diminishment rates of the reaction are determined by the permeability of cell membranes.

The Gray-Scott system is particularly notable for the wide range of irregular patterns it produces. Previous analysis of the system by Pearson [11] identified at least 12 different pattern types, all of which occur at different F, k with $d_u = 2d_v$. Figure 1.2 shows the 12 quantifiably different patterns observed in this system determined by standard methods of nonlinear analysis [12]. Figure 1.3 shows provides a legend for the patterns, mapping each of them to their locations in k, F parameter space. Each subfigure of Figure 1.2, which plots the concentration of chemical U over a 256 by 256 computational domain, reveal the extremely variable behavior of this system.¹ One of the most compelling qualities of these patterns is their resemblance to patterns of nature, e.g. κ (Figure 1.2j) looks like coral and λ (Figure 1.2k) resembles the growth of bacteria. Other patterns, like β (Figure 1.2b), exhibit turbulence.

1.2 Numerical Solution

The system in (1.3) is solved by forward Euler integration of the discrete Laplacian obtained by the finite difference method shown in (1.5).

$$\nabla^2 u(x, y) \approx u(x - 1, y) + u(x + 1, y) + u(x, y - 1) + u(x, y + 1) - 4u(x, y) \quad (1.5)$$

¹A concentration plot of V would resemble the inverse of U so only one is shown.

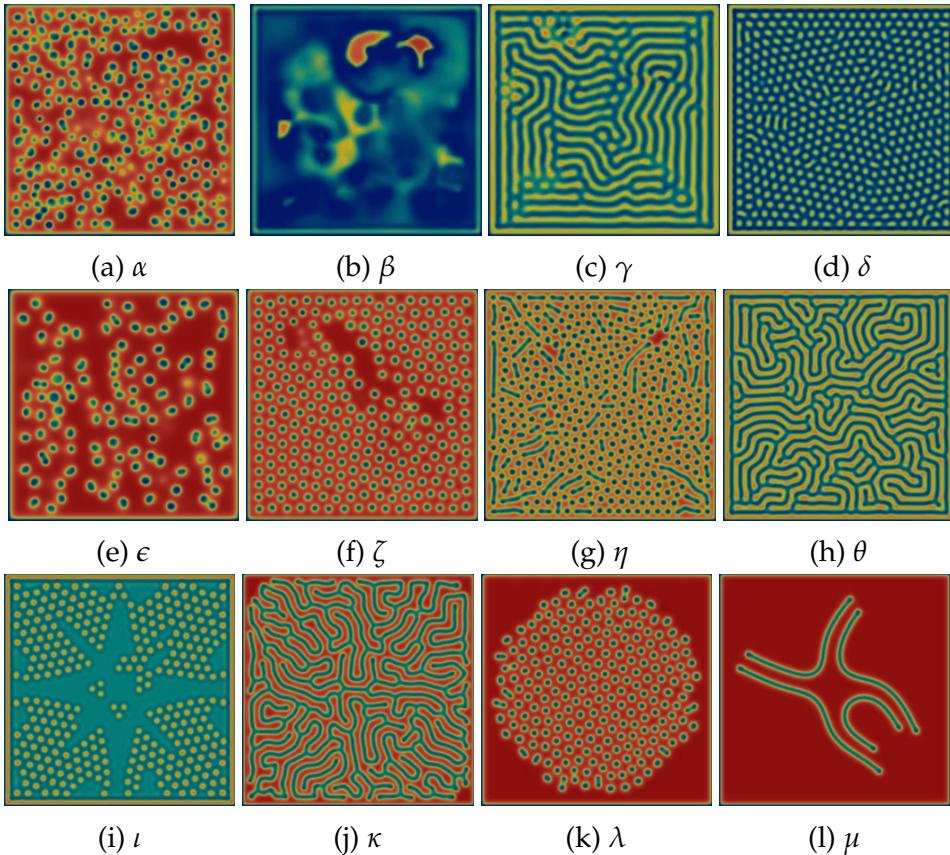


Figure 1.2: Patterns of chemical concentration U identified in [11]. Each pattern, Figure 1.2a—Figure 1.2l, is designated by a Greek letter which corresponds to the plot in Figure 1.3. Red and blue indicate $U = 1$ and $U \approx 0.2$ respectively. Note that a concentration plot of chemical V would appear as the inverse of U with red and blue swapped.

Similarly for v . In the Python programming language, this can be easily implemented using the Numpy package as below (see Sect. B.1 for full code).

```

Lu = ( U[0:-2,1:-1] + U[2:,1:-1] +
       U[1:-1,0:-2] + U[1:-1,2:] - 4*U[1:-1,1:-1] )
Lv = ( V[0:-2,1:-1] + V[2:,1:-1] +
       V[1:-1,0:-2] + V[1:-1,2:] - 4*V[1:-1,1:-1] )

uvv = u*v*v
u += (Du*Lu - uvv + F    *(1-u))
v += (Dv*Lv + uvv - (F+k)*v      )

```

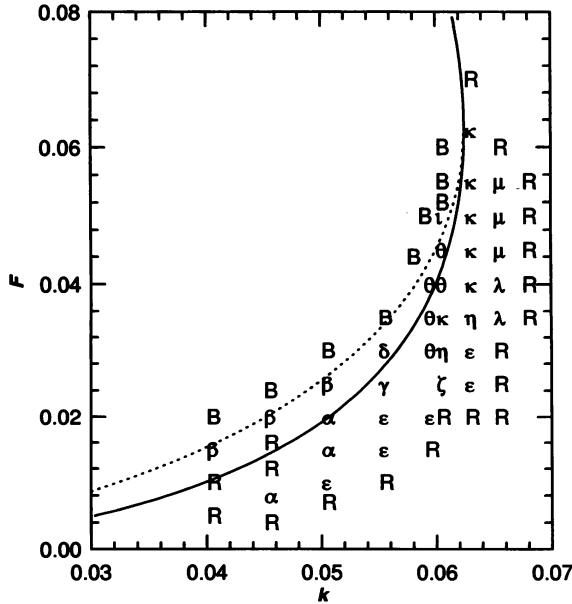


Figure 1.3: The mapping of Greek letters in Figure 1.2 to their location in k, F parameter space. R and B indicate that the system evolved to uniform red and blue states respectively. This figure also represents a phase diagram of the reaction kinetics. Between the solid and dotted line, the system is bistable for which there are two linearly stable steady states. As f passes below the dotted line, the non-trivial steady state becomes unstable through Hopf bifurcation giving stable periodic orbits for $k < 0.035$ and unstable for $k > 0.035$. The trivial state, $(U = 1, V = 0)$ exists for all (f, k) outside the solid line. Adapted from J. E. Pearson, “Complex patterns in a simple system,” *Science*, vol. 261, no. 5118, pp. 189–192, 1993.

A spacial grid of 256×256 points constitutes the mesh with a time step of 1 .² The system was initialized with the state $U = 1, V = 0$ with a 40×40 area located symmetrically in the center perturbed with $U = 0.5, V = 0.25$. This square area is then further sprinkled with 1% random “noise” to catalyze the reaction. The patterns in Figure 1.2 were generated using this method and depict the concentration of chemical U . A plot of chemical V would appear as the inverse.

For each F, k , the simulation is run for 25,000 time steps. Every 10th image is saved, so a total of 2,500 PNG files are produced to describe the pattern. The pattern images that are analyzed by the methods described in Chapter 3 use a greyscale colormap like that of Figures 1a and 1b.

² [11] notes no qualitative differences on mesh sizes up to 1024×1024 and time steps as low as 0.01. Initial conditions also have little to no effect on the qualitative features of the resulting pattern after some time.

Chapter 2

Computational Homology

As large amounts of data become available, it becomes more difficult to determine what information is relevant. There are, of course, high and low-level approaches. A high-level approach like a fingerprint scanner or handwriting recognition might be the end-goal of one's analysis, but lower-level approaches like homology look at the geometric makeup of an object and are often a requisite step toward building higher-level processes. Homology is one way of analyzing *local* properties in order to extract information about *global* phenomena.

At this time, computational homology is a relatively new field and its application in physics has only recently been explored [13, 14, 15, 16]. Although homology is a field of algebraic topology, it combines the mathematics of several other fields including combinatorics and computation. The mathematical formalism behind homology is difficult to grasp so only the relevant information will be detailed in Sect. 2.1. Sect. 2.2 on cubical homology presents a more thorough discussion of the mathematical background of the topic.

2.1 Image analysis

To put it simply, homology is concerned with enclosed *holes* and connected *pieces* in topological spaces. Homology provides a mathematical description of such geometric structures. Although the mathematics of homology are difficult, the relevant concepts are easily illustrated through examples. It's best to think about this in one dimension first. Figure 2.1 shows two simple topological spaces, X and Y . Although X and Y are spaces with one and two line segments respectively, in terms of homology one would say X consists of a single *connected piece* while Y

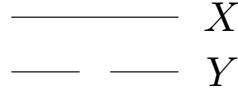


Figure 2.1: Topological spaces X and Y . X consists of one connected line segment and Y has two disconnected line segments.

has two distinct pieces. The fact that the line segments are straight or of different length is not important for the homology. In this one-dimensional example, the *zeroth homology groups* of each are

$$H_0(X) \cong \mathbf{Z}^1 \quad \text{and} \quad H_0(Y) \cong \mathbf{Z}^2 \quad (2.1)$$

where \mathbf{Z} is the group of integers. The homology pairs a topological space (e.g. X and Y) with an *abelian group*, a set of elements combined with operations that satisfy five axioms (closure, associativity, identity, invertibility, and commutativity). Notice, however, that the *zeroth homology group* of Y is \mathbf{Z}^2 ; the rank of the group, 2, is what accounts for the two distinct pieces, but more on this later.

Since there is a *zeroth homology group*, it makes sense that there would be a *first homology group*. Looking at the two-dimensional example in Figure 2.2, the homology of each space X_a , X_b , X_c , and X_d is

$$H_0(X_a) \cong \mathbf{Z} \quad H_0(X_b) \cong \mathbf{Z} \quad H_0(X_c) \cong \mathbf{Z} \quad H_0(X_d) \cong \mathbf{Z}^2 \quad (2.2)$$

$$H_1(X_a) \cong \mathbf{Z} \quad H_1(X_b) \cong 0 \quad H_1(X_c) \cong \mathbf{Z} \quad H_1(X_d) \cong \mathbf{Z} \quad (2.3)$$

Spaces X_a , X_b , and X_c have a *zeroth homology group* of \mathbf{Z} since there is a single connected component while X_d has \mathbf{Z}^2 to account for the two disconnected lines. In each space, a connected component forms an enclosed area (i.e. the squares). The square in Figure 2.2a forms a hole, i.e. a region completely enclosed by the black line. Figures 2.2a, 2.2c, and 2.2d each contain one hole. The shading in Figures 2.2b, 2.2c, and 2.2d indicates that the hole is filled and is thus no longer counted. Just as the *zeroth homology group* is concerned with connected segments, the *first homology group* is concerned with holes.

The terms “piece” and “hole” are informal. Formally, we could say that the k^{th} homology group, $H_k(X)$, represents the group of k -dimensional *holes* of X where a $k = 0$ hole is merely the gap between two components (e.g. Y in Figure 2.1). As I

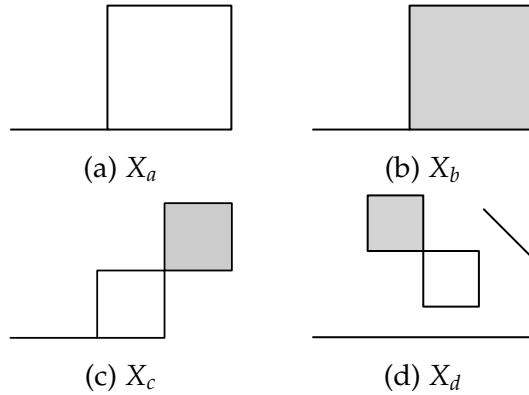


Figure 2.2: Topological spaces X_a , X_b , X_c , and X_d . Shading indicates that the enclosed area is filled.

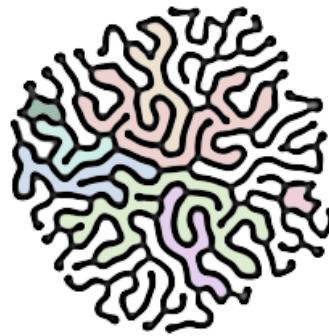


Figure 2.3: The pattern κ described in Sect. 1.1. The homology of κ gives Betti numbers $\beta_0 = 1$ and $\beta_1 = 9$. True to the homology, we can easily count a single black connected component and nine holes, each of which is filled with a different color to illustrate this fact.

alluded to earlier, the rank of the homology group (e.g. the rank 2 of \mathbf{Z}^2 in (2.1)) represents the *number* of k dimensional holes.¹ This is called the *Betti number* β_k . Indeed, Betti numbers are non-zero for all $k < d$ where d is the dimension of the topological space. Betti numbers are the most important feature of the homology in this thesis since it nicely provides a mathematical quantity to an otherwise visual characteristic of a topological space. Later on I will describe how this information has been used to elucidate information about the dynamics of a system.

¹The “dimension” of k is different from the dimension of the topological space. To describe the dimension of a space such as a cube in \mathbf{R}^3 , call it X , we write $\dim X = 3$. Structures of lower dimensions, such as the square faces that make up the cube, are *embedded* in the higher dimensional space X . For $k \geq \dim X$, $H_k(X) = 0$ since there are no structures embedded in a space with a higher dimension than that of the space itself. It is important to note that if we could place the topological spaces shown in Figure 2.2, which live in \mathbf{R}^2 on the page, into 3-dimensional space, \mathbf{R}^3 , this would not change the homology groups.

2.2 Cubical Homology

In cubical homology, topological spaces are represented as a collection of cubes. This thesis is concerned with the interpretation of digital images as topological spaces. Digital images are quite literally a collection of two-dimensional cubes, *pixels*, thus a homology that examines these objects is a natural environment for examining the output of a computer simulation. In this section, I present a brief mathematical description of cubical homology that closely follows that of [17] and [18]. By skipping this section, one would miss some of the interesting subtleties of homology theory but a thorough understanding is by no means essential to this thesis. We'll start by defining elementary cubes, which make up the building blocks for the theory. It is important to keep in mind here that one of the fundamental ideas in homology theory is to connect topological objects (e.g. connected pieces and holes) to algebraic objects.

Definition 2.2.1. An *elementary interval* is an interval $I \subset \mathbb{R}$ of the form

$$I = [l, l+1] \quad \text{or} \quad I = [l, l]$$

for some $l \in \mathbb{R}$. To simplify notation, say

$$[l] = [l, l]$$

is an interval containing a single point, which we call a *degenerate* interval. Intervals of the form $[l, l+1]$ are called *nondegenerate*.

Definition 2.2.2. An *elementary cube* Q is a finite product of elementary intervals,

$$Q = I_1 \times I_2 \times \dots \times I_d \subset \mathbb{R}^d$$

where each I_i is an elementary interval. We denote the set of all elementary cubes in \mathbb{R}^d as \mathcal{K}^d .

The set of all elementary cubes, \mathcal{K} ,

$$\mathcal{K} := \bigcup_{d=1}^{\infty} \mathcal{K}^d.$$

Definition 2.2.3. Let $Q = I_1 \times I_2 \times \dots \times I_d \subset \mathbb{R}^d$ be an elementary cube. The *embedding number* of Q is defined to be d which we denote by $\text{emb}(Q)$. Interval I_i

is the i^{th} component of Q and is written $I_i(Q)$. The dimension of Q is defined as the number of nondegenerate components in Q and denoted $\dim Q$. We refer to an elementary cube Q with $\dim Q = k$ as a k -cube and denote

$$\mathcal{K}_k := \{Q \in \mathcal{K} \mid \dim Q = k\},$$

and

$$\mathcal{K}_k^d := \mathcal{K}_k \cap \mathcal{K}^d.$$

The relationship between the embedding number and dimension might be a little muddy since it seems that they would always be the same. Observe that for elementary cube Q , if $\text{emb}(Q) = d$, then $Q \in \mathcal{K}^d$. The only general relation between the embedding number and dimension of Q is that

$$0 \leq \dim Q \leq \text{emb}(Q).$$

To illustrate this, imagine a Rubik's cube on a desk. The Rubik's cube itself has both $\text{emb}, \dim = 3$ while any one square *face* has $\text{emb} = 3$ but $\dim = 2$ (also see Example (2.2.1)).

Example 2.2.1. Given elementary cube $Q := [1, 2] \times [1, 2] \times [-1] \subset \mathbb{R}^3$, we have $I_1(Q) = [1, 2]$, $I_2(Q) = [1, 2]$, and $I_3(Q) = [-1]$ (*degenerate*). Therefore, $\text{emb}(Q) = 3$ and $\dim Q = 2$, due to the degenerate I_3 .

Now we must define the class of topological spaces for which we define the homology.

Definition 2.2.4. A set $X \subset \mathbb{R}^d$ is *cubical complex* if X can be written as a finite union of elementary cubes.

Given cubical complex $X \subset \mathbb{R}^d$, we define

$$\mathcal{K}(X) := \{Q \in \mathcal{K} \mid Q \subset X\}$$

and

$$\mathcal{K}_k(X) := \{Q \in \mathcal{K}(X) \mid \dim Q = k\}.$$

We can write $\mathcal{K}^d(X)$ to remind us that $X \subset \mathbb{R}^d$ as well as $\mathcal{K}_k^d := \mathcal{K}^d(X) \cap \mathcal{K}_k(X)$. For example, elements of $\mathcal{K}_0(X)$ are *vertices* of X , elements of $\mathcal{K}_1(X)$ are *edges* and so forth. $\mathcal{K}_k(X)$ are the k -*cubes* of X .

As previously mentioned, our goal is to establish a relationship between algebraic objects and topological spaces. The first step, then, is to associate some algebraic object with the elementary cubes we just defined.

Definition 2.2.5. For each elementary k -cube $Q \in \mathcal{K}_k^d$, we associate an algebraic object \widehat{Q} which we call an *elementary k -chain* of \mathbb{R}^d where $\widehat{Q} : \mathcal{K}_k^d \rightarrow \mathbb{Z}$ is the function defined by

$$\widehat{Q}(P) = \begin{cases} 1 & \text{if } P = Q, \\ 0 & \text{otherwise.} \end{cases}$$

We also define $\widehat{0} : \mathcal{K}_k^d \rightarrow \mathbb{Z}$ to be the zero function, i.e. $\widehat{0}(Q) = 0$ for all $Q \in \mathcal{K}_k^d$. The set of all *elementary k -chains* of \mathbb{R}^d is given by

$$\widehat{\mathcal{K}}_k^d := \left\{ \widehat{Q} \mid Q \in \mathcal{K}_k^d \right\}$$

and the set of all *elementary chains* of \mathbb{R}^d is given by

$$\widehat{\mathcal{K}}^d := \bigcup_{k=0}^{\infty} \widehat{\mathcal{K}}_k^d.$$

For an elementary cube Q , we refer to \widehat{Q} as its *dual elementary chain*. Conversely, given elementary chain \widehat{Q} , we call Q its *dual elementary cube*. What we want is that there is a one-to-one relationship between the elementary k -cubes (topological objects) and *elementary k -chains* (algebraic objects). In other words, the map of k -cubes (\mathcal{K}_k^d) to k -chains ($\widehat{\mathcal{K}}_k^d$) is a *bijection*.

Proposition 2.2.1. *The map $\phi : \mathcal{K}_k^d \rightarrow \widehat{\mathcal{K}}_k^d$ given by $\phi(Q) = \widehat{Q}$ is a bijection.*

Proof. See [18]. □

Proposition 2.2.1 allows us invoke the inverse of ϕ to go from an algebraic object, the elementary chain \widehat{Q} , to a topological set, Q . The following definition uses the algebra we've built up to give the elementary k -chains algebraic structure.

Definition 2.2.6. The group C_k^d of k -dimensional chains (or k -chains) of \mathbb{R}^d is the free abelian group generated by the elementary chains of \mathcal{K}_k^d . Thus the elements of C_k^d are functions $c : \mathcal{K}_k^d \rightarrow \mathbb{Z}$ such that $c(Q) = 0$ for all but a finite number of $Q \in \mathcal{K}_k^d$. In particular, $\widehat{\mathcal{K}}_k^d$ is the basis for C_k^d . By the same notation as (A.1.2),

$$C_k^d := \mathbb{Z}(\mathcal{K}_k^d)$$

If $c \in C_k^d$, then $\dim c := k$.

Figure 2.4 illustrates how we can use information about the boundary to exact information about the k -cubes, but we are again using topological information (the existence of a boundary) to derive more topological information (the existence of loops). What we would really like to do is use algebra to get at this information, so we start by defining the algebraic boundary of a k -chain.

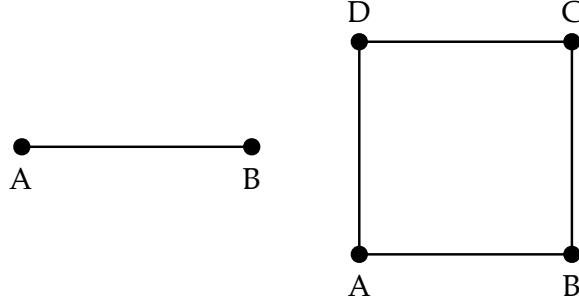


Figure 2.4: The left picture has a boundary given by $\{A\} \cup \{B\}$ and therefore does not form a loop (a 0-cube). The picture on the right, however, does not have a boundary and therefore forms a loop (a 1-cube).

The set of elementary chains forms a basis for C_k^d , thus we would like to easily describe an arbitrary chain $c \in C_k^d$ in terms of the elements of $\widehat{\mathcal{K}}_k^d$. Definition (2.2.7) provides such a relation which is analogous to the dot product in vector space.

Definition 2.2.7. Consider $c_1, c_2 \in C_k^d$, where $c_1 = \sum_{i=1}^m \alpha_i \widehat{Q}_i$ and $c_2 = \sum_{i=1}^n \beta_i \widehat{Q}_i$ and α_i and β_i scalars. The *scalar product* of the chains c_1 and c_2 is defined as

$$\langle c_1, c_2 \rangle := \sum_{i=1}^m \alpha_i \beta_i.$$

The astute reader will notice that this definition restricts us to describing a k -chain only in terms of k -dimensional cubes. We know, however, that cubes may be decomposed into lower dimensional faces. For example, the right picture in Figure 2.4 may be constructed from the four edges (or 1D faces) $[A, B]$, $[B, C]$, $[C, D]$, and $[D, A]$, and we would like to be able to write all k -chains in terms of lower-dimensional faces. This will be essential when the boundary operator is defined in 2.2.9 and provides the motivation for the following definition and proposition.

Definition 2.2.8. Given two elementary cubes $P \in \mathcal{K}_{k_1}^{d_1}$ and $Q \in \mathcal{K}_{k_2}^{d_2}$, let

$$\widehat{P} \diamond \widehat{Q} := \widehat{P \times Q}$$

This extends to arbitrary chains $c_1 \in C_{k_1}^{d_1}$ and $c_2 \in C_{k_2}^{d_2}$ by

$$c_1 \diamond c_2 := \sum_{P \in \mathcal{K}_{k_1}, Q \in \mathcal{K}_{k_2}} \langle c_1, \widehat{P} \rangle \langle c_2, \widehat{Q} \rangle \widehat{P \times Q}$$

The chain $c_1 \diamond c_2 \in C_{k_1+k_2}^{d_1+d_2}$ is called the *cubical product* of c_1 and c_2 .

Proposition 2.2.2. *Let \widehat{Q} be an elementary cubical chain of \mathbb{R}^d with $d > 1$. Then there exist unique elementary cubical chains \widehat{I} and \widehat{P} with $\text{emb}(I) = 1$ and $\text{emb}(P) = d - 1$ such that*

$$\widehat{Q} = \widehat{I} \diamond \widehat{P}$$

Proof. See [18]. □

Definition 2.2.9. Given $k \in \mathbb{Z}$, the *cubical boundary operator* or *cubical boundary map* given by

$$\partial_k : C_k^d \rightarrow C_{k-1}^d$$

is a homomorphism of free abelian groups, which is defined for an elementary chain $\widehat{Q} \in \widehat{\mathcal{K}_k^d}$ by induction on the embedding number d as follows. Consider first the case $d = 1$. Then Q is an elementary interval and hence $Q = [l] \in \mathcal{K}_0^1$ or $Q = [l, l+1] \in \mathcal{K}_1^1$ for some $l \in \mathbb{Z}$. Define

$$\partial_k \widehat{Q} := \begin{cases} 0 & \text{if } Q = [l], \\ [\widehat{l+1}] - [\widehat{l}] & \text{if } Q = [l, l+1] \end{cases}$$

Now assume that $d > 1$. Let $I = I_1(Q)$ and $P = I_2(Q) \times \dots \times I_d(Q)$. Then by 2.2.2,

$$\widehat{Q} = \widehat{I} \diamond \widehat{P}.$$

Define,

$$\partial_k \widehat{Q} := \partial_{k_1} \widehat{I} \diamond \widehat{P} + (-1)^{k_1} \widehat{I} \diamond \partial_{k_2} \widehat{P},$$

where $k_1 = \dim I$ and $k_2 = \dim P$. Finally, we extend the definition to all chains by linearity; that is, if $c = \alpha_1 \widehat{Q}_1 + \alpha_2 \widehat{Q}_2 + \dots + \alpha_m \widehat{Q}_m$, then

$$\partial_k c = \alpha_1 \partial_k \widehat{Q}_1 + \alpha_2 \partial_k \widehat{Q}_2 + \dots + \alpha_m \partial_k \widehat{Q}_m.$$

The domain of ∂_k is the k -chains, so if we know that $c \in C_k^d$, it is redundant and labor intensive to write the subscript k so we simplify to ∂ . Geometrically speaking,

the boundary of a k -chain is simply the alternating sum of its $(k - 1)$ -dimensional faces. As Figure 2.5 demonstrates, however, merely having a boundary sum equal to zero is not enough to constitute a loop. The right picture does not characterize a hole since it is filled in by the 2-chain Q .² This is represented algebraically by $\partial(\widehat{Q}) = [\widehat{A}, \widehat{B}] + [\widehat{B}, \widehat{C}] - [\widehat{C}, \widehat{D}] - [\widehat{D}, \widehat{A}]$.

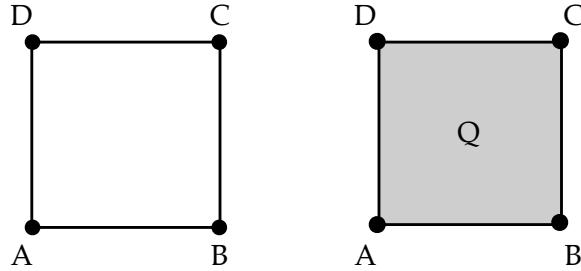


Figure 2.5: The boundary of the chain $[\widehat{A}, \widehat{B}] + [\widehat{B}, \widehat{C}] - [\widehat{C}, \widehat{D}] - [\widehat{D}, \widehat{A}]$ is zero in both pictures. The left picture, however, characterizes a loop (or hole) while the one on the right does not since it is “filled in” by 2-chain Q .

We can now say that holes are characterized by chains that have a boundary equal to zero, but are not themselves the boundary of other chains. In order to count the holes, then, we must count the chains which have zero boundary, but are not boundaries. The following definitions help us achieve this.

Definition 2.2.10. Let $X \subset \mathbb{R}^d$ be a cubical complex. Let $\widehat{\mathcal{K}}_k(X) := \{\widehat{Q} \mid Q \in \mathcal{K}_k(X)\}$. We define the set of k -chains of X as the subgroup $C_k(X)$ of C_k^d generated by the elements of $\widehat{\mathcal{K}}_k(X)$.

Proposition 2.2.3. Let $X \subset \mathbb{R}^d$ be a cubical complex. Then

$$\partial_k(C_k(X)) \subset C_{k-1}(X)$$

Proof. See [18]. □

This leads to the following definition.

Definition 2.2.11. The boundary operator of the cubical complex X is defined to be

$$\partial_k^X : C_k(X) \rightarrow C_{k-1}(X)$$

obtained by restricting $\partial_k : C_k^d \rightarrow C_{k-1}^d$ to $C_k(X)$.

²Not to be confused with 2 Chainz

An extremely important property of the boundary operator is defined in the following proposition.

Proposition 2.2.4.

$$\partial \circ \partial = 0$$

Proof. See [18]. □

It should make sense that if we are to take the boundary of a topological object, the boundary itself should have a lower embedding number (the boundary of a square, $k_1 = 2$, is made up of lines, $k_2 = 1$). It should also seem intuitive that the boundary of a boundary is zero. At this point we are tantalizingly close to defining homology—but wait! The following descriptions will be helpful in a moment.

Definition 2.2.12. Let $X \subset \mathbb{R}^d$ be a cubical set. A k -chain $z \in C_k(X)$ is called a *k -cycle* in X if $\partial z = 0$. Thus the set of all k -cycles of X is the kernel of ∂_k^X and so it is a subgroup of $C_k(X)$. We denote the set of all k -cycles by $Z_k(X)$. In short,

$$Z_k(X) := \ker \partial_k^X = C_k(X) \cap \ker \partial_k \subset C_k(X). \quad (2.4)$$

A k -chain $z \in C_k(X)$ is called a *boundary* in X if there exists $c \in C_{k+1}(X)$ such that $\partial c = z$. Thus the set of all boundary elements in $C_k(X)$ is the image of ∂_{k+1}^X , and so it is also a subgroup of $C_k(X)$. We denote the set of all boundary elements in $C_k(X)$ by $B_k(X)$. Once again,

$$B_k(X) := \text{im } \partial_{k+1}^X = \partial_{k+1}(C_{k+1}(X)) \subset C_k(X). \quad (2.5)$$

We can now say more precisely what we wanted before; we want to count the elements of $Z_k(X)$ that are not in $B_k(X)$. This is easily done by taking the quotient group of $Z_k(X)$ by $B_k(X)$ but this requires that $B_k(X)$ be a subgroup of $Z_k(X)$. By Proposition 2.2.4, $\partial c = z$ implies $\partial z = \partial^2 c = 0$, therefore every boundary is a cycle and $B_k(X)$ is a subgroup of $Z_k(X)$ so we may rightfully proceed to the most important definition of this section.

Definition 2.2.13. The k^{th} *cubical homology group* of X is the quotient group

$$H_k(X) := Z_k(X) / B_k(X).$$

The homology of X is the collection of all homology groups of X . The shorthand

for this is

$$H_*(X) := \{H_k(X)\}_{k \in \mathbb{Z}}.$$

For a cubical set $X \subset \mathbb{R}^d$ we can show that, for $i = 0, \dots, d - 1$, $H_i(X) = \mathbb{Z}^{\beta_i} \oplus \mathbb{Z}_{b_1} \oplus \mathbb{Z}_{b_2} \oplus \dots \oplus \mathbb{Z}_{b_k}$, where β_i is a nonnegative integer, \mathbb{Z}_b is the group of integers modulo b , $b_i > 1$ provided $k > 0$, and b_i divides b_{i+1} for $i \in \{1, 2, \dots, k-1\}$ provided $k > 1$. For $i \geq d$ we have $H_i(X) = 0$. Integer β_i is known as the i^{th} *Betti number* of X and b_1, b_2, \dots, b_k are the *torsion coefficients* of $H_i(X)$. In general, $\beta_i := \text{rank}(H_i(X))$. Spaces with dimension $d \leq 3$ do not have torsion coefficients, just $H_i(X) = \mathbb{Z}^{\beta_i}$, so we need not worry about them for our purposes [17].

This is certainly a great abstraction from what we started with before, connected pieces and holes in a topological space, but here is some geometrical intuition. As previously indicated, the Betti numbers encode some geometrical information. β_0 is equal to the number of connected pieces of X , β_1 is the number of holes (or loops) if $d = 2$ or the number of tunnels if $d = 3$. β_2 equals the number of cavities if $d = 3$.

Example 2.2.2. An ordinary bike tube (a torus) has $\beta_0 = 1$, the single connected piece of rubber; $\beta_1 = 1$, one hole in the center; and $\beta_2 = 1$, the hollow cavity inside the tube.

The mathematical tour-de-force that is homology theory might seem like major overkill. After all, here we are merely concerned with counting geometric structures that could theoretically be eyeballed (tedious as that may be). I argue, however, that the homology maintains some nice features for us. For one, it provides a mathematically rigorous definition of the structures in question. Furthermore, the homology of any structure is unchanged in any higher-dimensional space (e.g. the homology groups of an empty square are the same in \mathbf{R}^2 as in \mathbf{R}^3 or \mathbf{R}^4 for that matter). Homology is not concerned with size or shape of any object, either; the homology of a coffee mug is identical to that of Figure 2.2a. The theory reduces the amount of information required to describe an object to a few topological quantities which may be difficult to grasp visually. Analysis of higher dimensional data, such as a 4D construction of medical imaging data, which would require a great amount of thought to assess visually, is completely feasible through cubical homology [18]. Due to its dimension-independent formulation, the applications of cubical homology are limited only by the ability to construct sensible topological information.

Chapter 3

Methods and procedures

In this section I outline the methods for actually extracting homology information, i.e. Betti numbers, from images generated by simulating the Gray-Scott system. Sect. ?? describes the process of preparing the images for computation and the considerations and problems that arise. Sect. 3.1.3 is concerned with calculating the entropy of the system given the homological data. Section 1.2 details the conditions for generating the pattern images.

3.1 Obtaining Betti numbers

Given the complicated overview of cubical homology theory given in Section 2.2, one might expect extracting the Betti numbers from an image to be a difficult undertaking. Fortunately, homology software developed at Rutgers, with contributions from Georgia Tech and Konstantin Mischaikow (Mathematics '79), titled CHOMP, (Computational Homology Project), facilitates this process [19]. Furthermore, the way it works is extremely intuitive, essentially counting clusters of adjacent pixels. This requires a 2-bit binary image as input.¹ Since the images output by the Gray-Scott simulation are greyscale, they must first be converted to a 2-bit image

3.1.1 Thresholding

The output of the Gray-Scott simulation is a series of 8-bit greyscale images, i.e. there are 256 possible shades of grey (0 is black, 255 is white). The color map is such that

¹That is, an image with *only* black and white pixels.

$V_{min} = 0.0 \rightarrow 255$ and $V_{max} = 0.4 \rightarrow 0$. In other words, white and black indicate low and high concentration of V respectively². Each image is thresholded at some value $T \in [0, 255]$. That is, all pixels with intensity $< T$ are now black and those with intensity $> T$ are now white.

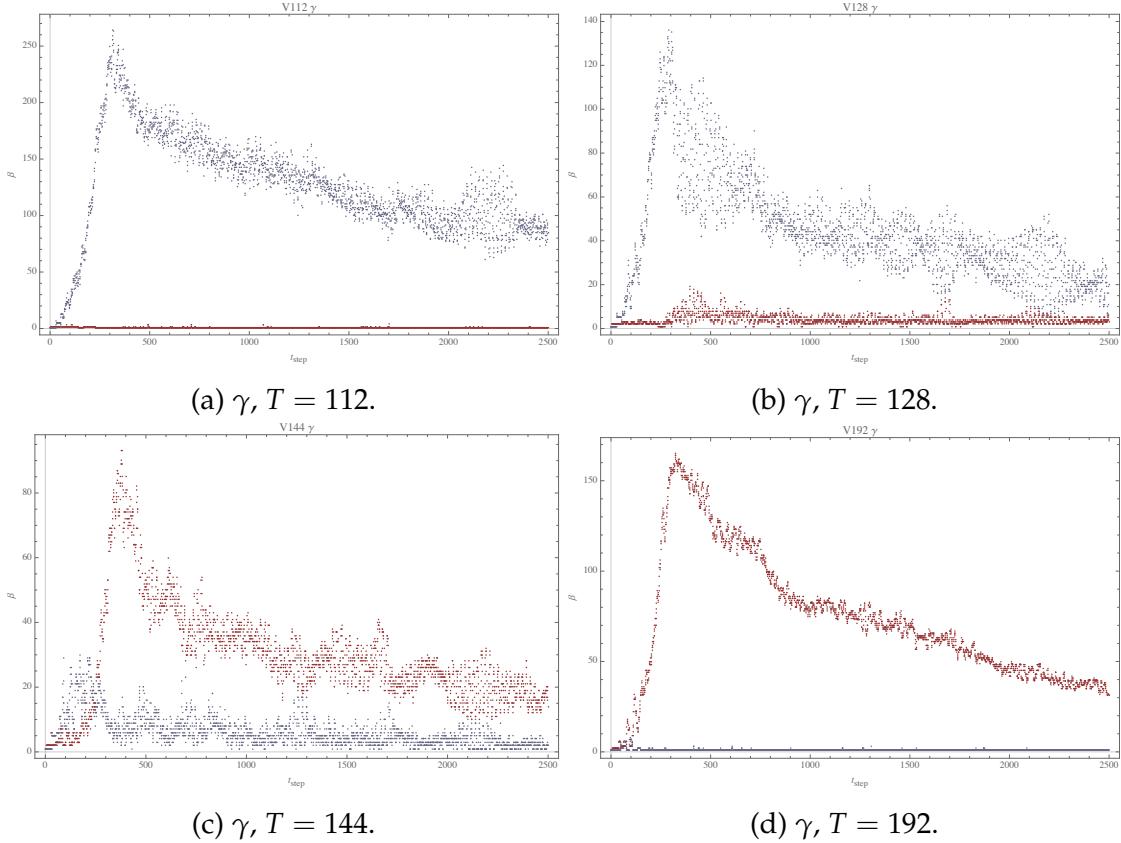


Figure 3.1: A plot of the time series of Betti numbers for pattern γ . The zeroth Betti number β_0 is shown in red and the first Betti number β_1 shown in blue. Different thresholds $T = 112, 128, 144$, and 192 demonstrate the dramatic effect of thresholding for some patterns. For very high and low T , the image loses any resemblance to the original image. Slightly varying the threshold, however, can reduce the amount of “noise” in the data.

A logical choice for T is the median pixel intensity of the image [13]. But, in the case of sparse patterns like α and ϵ (Figures 1.2a and 1.2e respectively), this results in a completely black image since the median is very high. Although other adaptive methods of thresholding exist [?], the definitive answer to thresholding problems is *persistent homology*, which is a large leap in terms of complexity [20]. This approach is concerned with the “birth” and “death” of homology components as the threshold is varied in a way that circumvents the need for a threshold

² V_{max} was chosen based on the average maximum of $V \approx 0.4$ for all the patterns examined.

altogether. Persistent homology has seen great success in analyzing large sets of nonlinear data [21].

³Short of persistent homology, another reasonable choice would be to split right down the middle, $T = 128$, but as Figures 3.1 and 3.2 illustrate, some information is lost in the process. Thus the optimal choice of threshold depends entirely on the characteristics of that image. The patterns produced by the Gray-Scott system are varied and no single value of T is ideal for all (F, k) , but experimentally, slightly higher T better agrees with the characteristics of the original pattern. For this reason, the value $T = 144$ was chosen to perform the all the calculations.

Provided in the CHOMP software package is a method for simply thresholding images, `chomp-greyscale-to-cubical` (which takes a single image input, at threshold, and an output filename). The output is a text file which contains the coordinates of white pixels. This is what is then analyzed to produce Betti numbers. A script automates this process for all 2,500 png files generated by the simulation (see Sect. 1.2).

3.1.2 Performing cubical homology

Once the images are thresholded at some value, the CHOMP method `chomp-cubical` analyzes and returns Betti numbers β_0 , β_1 , and β_2 . A single calculation of Betti numbers takes about 1s on a 4.2 GHz Intel i7 processor. This process is performed for each of 2,500 text files produced by `chomp-greyscale-to-cubical` to generate a single CSV file of the Betti numbers β_0 , β_1 , and β_2 for each time step.⁴

3.1.3 Calculating entropy

In physics, entropy usually denotes the amount of “disorder” of a system. Shannon’s entropy, $H(X)$, indicates the average amount of information that an observer gains *after* receiving a realized outcome x of the random variable X [22]. We wish to use homology information to provide a sense of how predictable (how complex) the Gray-Scott system is for a given choice of parameters (namely F, k). In general, the Shannon entropy H of some variable X with possible values $\{x_1, \dots, x_N\}$ and

³JH Include example of GS model pic w/ thresholded image and betti numbers:

⁴The Betti number β_2 is included only for posterity; $\beta_2 = 0$ for all time steps.

probability distribution $P(x_i) = P_i$ is defined by

$$H(X) = - \sum_{i=1}^N P_i \log P_i. \quad (3.1)$$

In our case, the Shannon entropy gives a picture of the average minimum number of states required to describe the system based on the frequency of the states. A *state* in this case is taken to be a unique pair of Betti numbers $s_i = \{\beta_0, \beta_1\}_i$ at the i th time step. Although s_i in general doesn't describe a *unique* pattern (any two states such that $s_i = s_j$ could look very different), it captures the fundamental topology of the system at that moment. Furthermore, the set of states within a given set of parameters (any F, k) is more meaningful. For example, if we observe the topology of state s_i for pattern α , then we can informed prediction as to what the state s_j actually looks like for that system. In general, we would expect higher entropy for more dynamic, complex patterns.

For N total (non-unique) states equal to the number of time steps, the probability P_i of state s_i given N_i , the number of times state s_i occurs, is simply

$$P_i = \frac{N_i}{N}. \quad (3.2)$$

In order to elucidate information about the dynamics of the Gray-Scott system in particular, (3.1) is calculated for all $\{F, k \mid F \in [0.004, 0.08], k \in [0.03, 0.07]\}$ ⁵ in an evenly spaced 20×20 grid according to Pearson's map of F, k parameter space (see Figure 1.3). There are then 400 total values of F, k for which the entropy H is calculated. With 2,500 time steps for each F, k , producing a map of the entropy over F, k space requires 1 million calls to chomp-cubical⁶. Computed with eight parallel processes, this takes 1-2 days of computation time.

⁵JH 5 apr: Not sure if this is the correct notation

⁶Computing the Betti numbers of the system is by far the slowest operation and significantly bottlenecks the processing time.

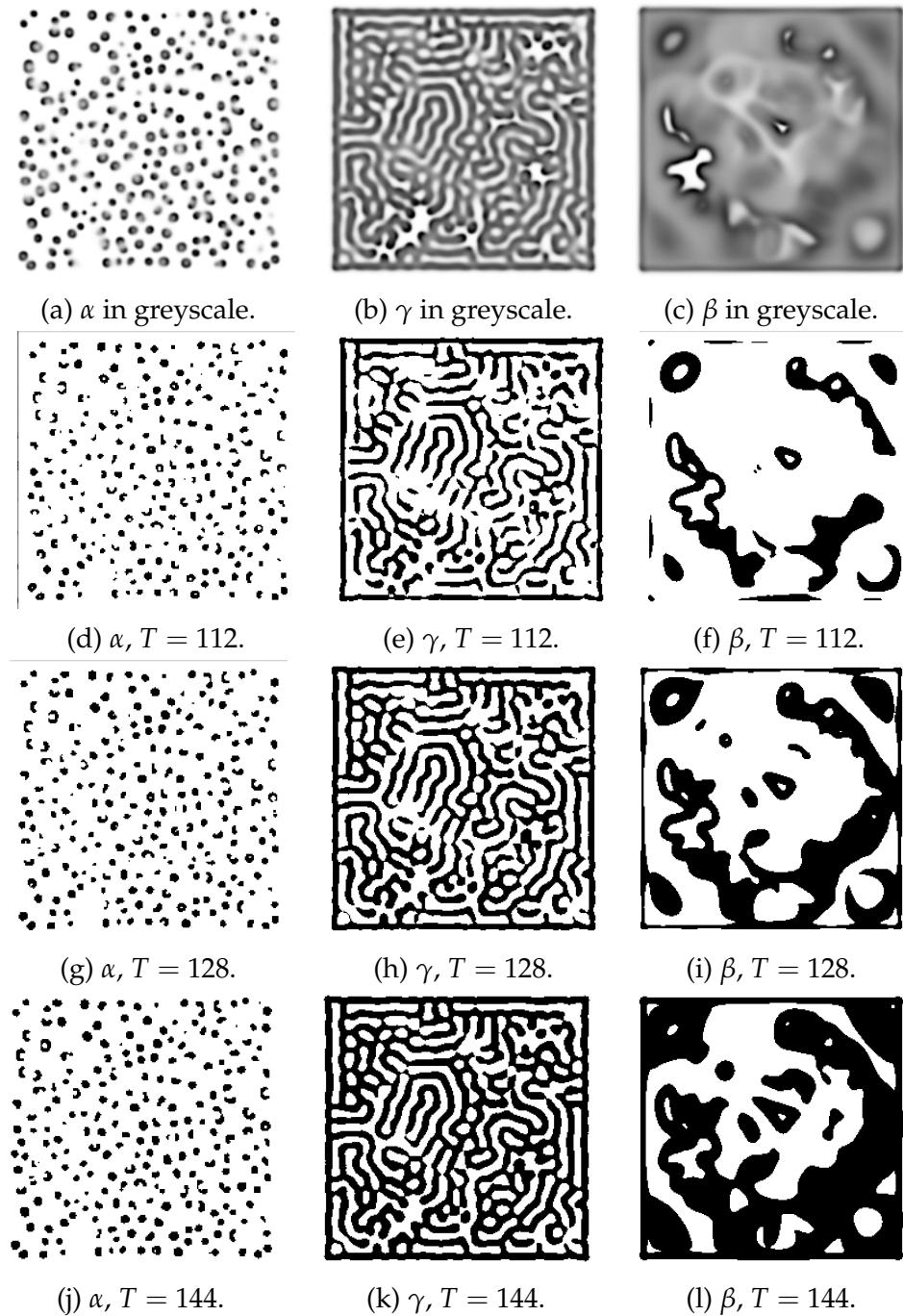


Figure 3.2: Patterns α , γ , and β at various thresholds. Some features of less stable patterns such as γ and β are lost when thresholded. Slightly higher thresholds (> 128) tend to capture the topology of the original pattern better.

Appendix A

The First Appendix

A.1 Extra definitions, theorems, etc.

Definition A.1.1. The *free abelian group generated by a finite set*

$$S = \{s_1, s_2, \dots, s_n\}$$

is the set of all functions $f : S \rightarrow \mathbb{Z}$, with the pointwise addition

$$(f + g)(s_i) := f(s_i) + g(s_i), \quad i = 1, 2, \dots, n.$$

Definition A.1.2. The *free abelian group generated by a possibly infinite set S* is the subgroup of \mathbb{Z}^S , consisting of all functions $f : S \rightarrow \mathbb{Z}$ satisfying

$$f(s) = 0 \quad \text{for all but finitely many } s \in S.$$

Appendix B

Code

B.1 Gray-Scott simulation

```
def runGS(Du, Dv, F, k, name):
    n = 256

    Z = np.zeros((n+2,n+2), [(‘U’, np.double), (‘V’, np.double)])
    U,V = Z[‘U’], Z[‘V’]
    u,v = U[1:-1,1:-1], V[1:-1,1:-1]
    r = 20
    u[...] = 1.0
    U[n/2-r:n/2+r,n/2-r:n/2+r] = 0.50
    V[n/2-r:n/2+r,n/2-r:n/2+r] = 0.25
    u += 0.05*np.random.random((n,n))
    v += 0.05*np.random.random((n,n))

    plt.ion()

    size = np.array(Z.shape)
    dpi = 120.0
    figsize= size[1]/float(dpi),size[0]/float(dpi)
    fig = plt.figure(figsize=figsize, dpi=dpi, facecolor="white")
    fig.add_axes([0.0, 0.0, 1.0, 1.0], frameon=False)
    cmap = plt.cm.binary
    im = plt.imshow(V, interpolation='bicubic', cmap=cmap)
```

```
plt.xticks([]), plt.yticks([])

for i in xrange(25000):
    Lu = ( U[0:-2,1:-1] + U[2:,1:-1] +
           U[1:-1,0:-2] + U[1:-1,2:] - 4*U[1:-1,1:-1] )
    Lv = ( V[0:-2,1:-1] + V[2:,1:-1] +
           V[1:-1,0:-2] + V[1:-1,2:] - 4*V[1:-1,1:-1] )

    uvv = u*v*v
    u += (Du*Lu - uvv + F*(1-u))
    v += (Dv*Lv + uvv - (F+k)*v)

if i % 10 == 0:
    im.set_data(V)
    im.set_clim(vmin=0.0, vmax=0.4)
    plt.draw()

plt.ioff()
plt.close()
```

References

- [1] S. Hui and S. ak, "Discrete fourier transform based pattern classifiers," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 62, no. 1, pp. 15–22, 2014.
- [2] A. M. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 237, no. 641, pp. 37–72, 1952.
- [3] J. Bard and I. Lauder, "How well does turing's theory of morphogenesis work?," *Journal of Theoretical Biology*, vol. 45, no. 2, pp. 501–531, 1974.
- [4] J. B. L. Bard, "A model for generating aspects of zebra and other mammalian coat patterns," *Journal of Theoretical Biology*, vol. 93, no. 2, pp. 363–385, 1981.
- [5] J. Murray, "On pattern formation mechanisms for lepidopteran wing patterns and mammalian coat markings," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 295, no. 1078, pp. 473–496, 1981.
- [6] H. Meinhardt and H. Meinhardt, *Models of biological pattern formation*, vol. 6. Academic Press London, 1982.
- [7] E. Dabelsteen, K. Buschard, S.-I. Hakomori, and W. W. Young, "Pattern of distribution of blood group antigens on human epidermal cells during maturation," *Journal of Investigative Dermatology*, vol. 82, no. 1, pp. 13–17, 1984.
- [8] G. Turk, "Generating textures on arbitrary surfaces using reaction-diffusion," *SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 289–298, 1991.
- [9] A. Witkin and M. Kass, "Reaction-diffusion textures," *SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 299–308, 1991.

- [10] P. Gray and S. Scott, "Autocatalytic reactions in the isothermal, continuous stirred tank reactor: Oscillations and instabilities in the system $a + 2b \rightarrow b$; $b \rightarrow c$," *Chemical Engineering Science*, vol. 39, no. 6, pp. 1087–1097, 1984.
- [11] J. E. Pearson, "Complex patterns in a simple system," *Science*, vol. 261, no. 5118, pp. 189–192, 1993.
- [12] S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*, vol. 272. Westview Press, 2001.
- [13] K. Krishan, H. Kurtuldu, M. F. Schatz, M. Gameiro, K. Mischaikow, and S. Madruga, "Homology and symmetry breaking in rayleigh-bnard convection: Experiments and simulations," *Physics of Fluids (1994-present)*, vol. 19, no. 11, p. 117105, 2007.
- [14] M. Gameiro, K. Mischaikow, and W. Kalies, "Topological characterization of spatial-temporal chaos," *Physical Review E*, vol. 70, no. 3, p. 035203, 2004.
- [15] H. Kurtuldu, K. Mischaikow, and M. Schatz, "Measuring the departures from the boussinesq approximation in rayleighbnard convection experiments," *Journal of Fluid Mechanics*, vol. 682, pp. 543–557, 2011.
- [16] A. Szymczak, A. Stillman, A. Tannenbaum, and K. Mischaikow, "Coronary vessel trees from 3d imagery: a topological approach," *Medical image analysis*, vol. 10, no. 4, pp. 548–559, 2006.
- [17] M. F. Gameiro, "Topological analysis of patterns," 2005.
- [18] T. Kaczynski, K. M. Mischaikow, and M. Mrozek, *Computational homology*, vol. 157. Springer Science & Business Media, 2004.
- [19] S. Harker, "Chomp," 2015. <http://chomp.rutgers.edu/>.
- [20] H. Edelsbrunner and J. Harer, "Persistent homology-a survey," *Contemporary mathematics*, vol. 453, pp. 257–282, 2008.
- [21] S. Weinberger, "What is... persistent homology?," *Notices AMS*, vol. 58, no. 01, pp. 36–39, 2011.
- [22] P. Grunwald and P. Vitnyi, "Shannon information and kolmogorov complexity," *arXiv preprint cs/0410002*, 2004.