NORTHWESTERN UNIVERSITY

Markov Partitions for Hyperbolic Automorphisms of the Two-Dimensional Torus

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Mathematics

By

Mark Richard Snavely

EVANSTON, ILLINOIS

June 1990

ERRATA

On page 15 in the last paragraph, the sentence beginning "We can then see..." should read:
We can then see that $W^s(x)=\{y:d(\sigma_A^n x, {}_A^n y)\rightarrow 0 \text{ as } n\rightarrow\infty\}=\{y:y_i=x_i \text{ for } i\geq n\}$ and $W^u(x)=\{y:y_i=x_i$ for $i\leq n\}$.

On page 15, 6 lines from the bottom, in the definition of the metric, $2^{|i|}$ should read $2^{-|i|}$.

On page 16, fourth line from top should have "$M_{i,j}=1$ if $\sigma(R_i)\cap R_j\neq\emptyset$."

On page 19, second line should read "$\phi^{-1}\mathcal{P}$ is a Markov partition."

On page 25 in Example 2.2.1, the Markov matrix M is incorrect. R_1 should be crossed by rectangles numbered 1,3,4,7, and 8 but not 5, and R_2 should be crossed by rectangles numbered 1, 5, and 7 but not 6.

On page 27, Figure 2.2.6, the point D should be on the upper boundary of R2, not the lower.

On page 34 in Example 2.2.18, the sentence beginning "We label..." should read:
We label $S_1=\{$all uncovered components in $R_1-\mathcal{A}R_1\}$ with A_1, $S_2=\{$all uncovered components in $R_1-S_1-\mathcal{A}^2R_1\}$ with $A_2,\ldots,S_n=\{$all uncovered components in $R_1-S_1-S_{n-1}-\mathcal{A}^nR_1\}$ with A_n and $S_{mn+k}=\{$all uncovered components in $R_1-S_1-S_{n-1}-\mathcal{A}^{nm+k}R_1$ for $m\in\mathbb{Z}_+$ and $0\leq k\leq n+1\}$ with A_k.

On page 36, Example 2.2 19 should refer to Figure 2.2.21 while Example 2.2.20 should refer to Figures 2.2.22-2.2.24.

On page 46, the proof of Theorem 3.2.1 only deals with the case in which all rectangles are connected. However, as the transition matrix for an FCC partition is the transition matrix for a partition with connected rectangles on which a row or column amalgamation has been performed, the theorem follows directly.

On page 77, Proposition 5.1.6, the assumptions should read that $a\geq 2$.

ABSTRACT

Markov Partitions for Hyperbolic Automorphisms of the Two-Dimensional Torus

Mark R. Snavely

Hyperbolic automorphisms of the two-dimensional torus are determined by hyperbolic, two-by-two, integer matrices with determinant plus or minus one. These maps were found to have relatively simple Markov partitions in the late 1960's independently by K. Berg in his Ph.D. thesis and by R. Adler and B. Weiss.

THEOREM: Let $\mathcal{A}$ be a hyperbolic automorphism of $\mathbf{T}^2$ and let $\mathfrak{P}$ be a Markov partition for $\mathcal{A}$ with a finite number of connected rectangles and suppose the transition matrix for this system is M. If the trace of $\mathcal{A}$ is positive, then the eigenvalues of M are the eigenvalues of $\mathcal{A}$ together with zeros and roots of unity. If the trace of $\mathcal{A}$ is negative, then the eigenvalues of $-$M are the eigenvalues of $\mathcal{A}$ together with zeros and roots of unity.

In the two rectangle case, we prove an even more powerful result.

THEOREM: Let $\mathcal{A}$ be a hyperbolic toral automorphism with the trace of $\mathcal{A}$ positive. Let M be a two-by-two, non-negative, integer matrix. There is a Markov partition for $\mathcal{A}$ with two rectangles and Markov matrix M if and only if $\mathcal{A}$ and M are similar over the integers. If the trace of $\mathcal{A}$ is negative, then such a partition exists if and only if $\mathcal{A}$ and $-$M are similar over the integers.

DEFINITION: We define the <u>unstable core of $\mathfrak{P}$</u>, denoted $\mathfrak{U}$, to be the intersection over backward time of the unstable boundary of $\mathfrak{P}$. We define the <u>stable core of $\mathfrak{P}$</u>, denoted $\mathfrak{S}$, to be the intersection over forward time of the stable boundary of $\mathfrak{P}$. We define the <u>core of $\mathfrak{P}$</u>, denoted $\mathfrak{C}$, to be $\mathfrak{C}=\mathfrak{U}\cup\mathfrak{S}$.

Ashley, Kitchens, and Stafford have shown that the core as defined above can include any sofic system with low enough entropy and few enough periodic points of each period. We construct examples which demonstrate the composition of the core and its relationship to M.

Finally, we return to the two rectangle case and prove the following theorem.

THEOREM: Suppose $\mathcal{A}$ is a hyperbolic toral automorphism. There is a finite set $\mathfrak{G}$ of two rectangle partitions such that any other two rectangle partition is the image of a partition in $\mathfrak{G}$ under an element of $GL(2,\mathbb{Z})$ which commutes with $\mathcal{A}$.

To my family – without their love and support none of this would have been possible.

"Though I have all knowledge, if I have not love, I am nothing."

from I Corinthians 13:2

ACKNOWLEDGEMENTS

I would like to thank my advisor, John Franks, for getting me started with Theorem 3.2.1 and for all his suggestions and encouragement. Among the graduate students here at Northwestern, two must be thanked personally: Charlie Jacobson for many hours of talking and listening throughout my tenure at N. U. and Kevin Woerner for his suggestions, particularly on Sections 4.1 and 5.1 and the computer system. Finally, a special thanks to Dr. Kenneth Rietz who lit the fires of inspiration back at Grove City College.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 0 - INTRODUCTION

SECTION 0.1: A BRIEF HISTORY OF THE PROBLEM

The dynamics of automorphisms of the torus have been studied for some time and have proven to be a rich mathematical construct. Applications of such dynamics for the two-dimensional torus can be found in a variety of places ranging from beginning texts in dynamical systems to texts in ergodic theory and measure theory. These maps exhibit interesting geometric, topological, algebraic, measure theoretic, and ergodic properties in a rather tractable setting. The beginning student in dynamical systems is often introduced to the notion of a Markov partition by looking at the torus, yet there is still much we do not know about these maps. Some questions about the behavior of these systems will be answered in this paper.

It was noticed independently by Berg [4] and by Adler and Weiss [2] in the late 1960's that the two-torus had relatively simple Markov partitions. This discovery allowed the use of symbolic dynamics in analyzing automorphisms of the torus – namely a relationship between Markov partitions for a smooth dynamical system and a symbol space associated with the partitions which was easier to study in certain instances. For a more detailed history, the reader is strongly encouraged to read Adler [1].

This work was inspired not only by the work of Adler and Weiss, but also by the work of Stafford on the doubling map $x \mapsto 2x$ for S^1 [12]. Indeed, Theorem 3.2.1 is an extension of his

Proposition C from $\mathbf{T}^1 = S^1$ to $\mathbf{T}^2$.

In this thesis, we begin by introducing the theory of Axiom A diffeomorphisms on which the ideas of Markov partitions are based. We then give a number of examples of Markov partitions and in the process describe the invariant set contained in the boundary of the partition. In Chapter 3 we give necessary conditions on the eigenvalues of the Markov matrix when the aforementioned invariant set is finite. In Chapter 4 we give necessary and sufficient conditions on the Markov matrix for a partition with two connected rectangles. Finally, in Chapter 5 we show that if we focus our attention on partitions with two connected rectangles, there is a finite set of partitions which generate all other partitions with two rectangles. In Chapter 6 we summarize and hypothesize as to where these results can lead.

CHAPTER 1 - AXIOM A AND THE TORUS

SECTION 1.1: AXIOM A DIFFEOMORPHISMS

The purpose of this thesis is to examine Markov partitions for hyperbolic automorphisms of the two-dimensional torus. To do so, we must first review the theory of Axiom A diffeomorphisms from which the idea of Markov partitions developed. Any proofs in this section not referenced or given can be found in Bowen, Chapter 3 [5].

We begin by defining terms which will be used throughout the paper. Suppose that $f:M \to M$ is a C^∞ diffeomorphism of a manifold M and let Λ be an invariant set for f; that is, $f(\Lambda) \subset \Lambda$. The derivative of f can be viewed as a map $Df:TM \to TM$ where $TM = \bigcup_{x \in M} T_xM$ is the tangent bundle of M, T_xM is the tangent space of M at $x \in M$, and $Df_x:T_xM \to T_{f(x)}M$.

DEFINITION 1.1.1: An invariant set Λ is said to be <u>hyperbolic</u> if for every $x \in \Lambda$, the tangent space T_xM has a direct sum decomposition into subspaces E^u_x and E^s_x such that

a) $T_xM = E^u_x \oplus E^s_x$

b) $Df(E^s_x) = E^s_{f(x)}$, $Df(E^u_x) = E^u_{f(x)}$

c) there exist $c > 0$ and $\lambda \in (0,1)$ such that

$$\|Df^n(v)\| \leq c\lambda^n \|v\| \text{ for } v \in E^s_x \text{ and } n \geq 0$$

and $$\|Df^{-n}(v)\| \leq c\lambda^{-n} \|v\| \text{ for } v \in E^u_x \text{ and } n \geq 0.$$

d) E^u_x and E^s_x vary continuously with x.

DEFINITION 1.1.2: A point $x \in M$ is called <u>nonwandering</u> if $U \cap \left(\bigcup_{n>0} f^n U\right) \neq \emptyset$ for any neighborhood U of x. We denote the set of all nonwandering points of f by <u>$\Omega(f)$</u>. $\Omega(f)$ is closed and invariant under f.

Another way of stating the definition of nonwandering is that every neighborhood of x returns and intersects itself under enough iterates of f.

DEFINITION 1.1.3: A point x is <u>periodic</u> if $f^n(x)=x$ for some $n>0$. We call n the <u>period</u> of f and denote the set of all periodic points of f by <u>Per(f)</u>. Clearly $\mathrm{Per}(f) \subset \Omega(f)$.

DEFINITION 1.1.4: A point x is called a <u>fixed</u> <u>point</u> if $f(x)=x$. We denote the set of all fixed points of f by <u>Fix(f)</u>. Clearly, $\mathrm{Fix}(f) \subset \mathrm{Per}(f)$

DEFINITION 1.1.5: f is said to satisfy <u>Axiom</u> <u>A</u> if $\Omega(f)$ is hyperbolic and $\Omega(f)=\overline{\mathrm{Per}(f)}$.

DEFINITION 1.1.6: f is said to be <u>Anosov</u> if all of M is hyperbolic. It can be shown that an Anosov diffeomorphism always satisfies Axiom A [5].

DEFINITION 1.1.7: For $x \in M$, let us define the following:

$$W^s(x)=\{y \in M: d(f^n x, f^n y) \to 0 \text{ as } n \to \infty\}$$

$$W^s_\epsilon(x)=\{y \in M: d(f^n x, f^n y) \leq \epsilon \text{ for all } n \geq 0\}$$

$$W^u(x)=\{y \in M: d(f^{-n} x, f^{-n} y) \to 0 \text{ as } n \to \infty\}$$

$$W^u_\epsilon(x)=\{y \in M: d(f^{-n} x, f^{-n} y) \leq \epsilon \text{ for all } n \geq 0\}.$$

We call $W^s(x)$ the <u>stable</u> <u>manifold</u> <u>of</u> <u>x</u>. We call $W^u(x)$ the <u>unstable</u> <u>manifold</u> <u>of</u> <u>x</u>. $W^s_\epsilon(x)$ and $W^u_\epsilon(x)$ are called the <u>local</u> <u>stable</u> and <u>local</u> <u>unstable</u> <u>manifolds</u> respectively.

We now state a stable manifold theorem, which is proven in [6]. The statement is from [5].

THEOREM 1.1.8: Let Λ be a hyperbolic set for a C^r diffeomorphism f. For small $\epsilon>0$

a) $W^s_\epsilon(x)$ and $W^u_\epsilon(x)$ are C^r disks for $x\in\Lambda$ with $T_x W^s_\epsilon(x)=E^s_x$ and $T_x W^u_\epsilon(x)=E^u_x$.

b) $d(f^n x, f^n y)\leq\lambda^n d(x,y)$ for $y\in W^s_\epsilon(x)$ and $n\geq 0$ and $d(f^{-n}x, f^{-n}y)\leq\lambda^n d(x,y)$ for $y\in W^u_\epsilon(x)$ and $n\geq 0$.

c) $W^s_\epsilon(x)$ and $W^u_\epsilon(x)$ vary continuously with x.

From this we see that $W^s_\epsilon(x)\subset W^s(x)$ and $W^u_\epsilon(x)\subset W^u(x)$.

THEOREM 1.1.9 **Canonical Coordinates:** Suppose f satisfies Axiom A. Then for any small $\epsilon>0$ there is a $\delta>0$ such that $W^s_\epsilon(x)\cap W^u_\epsilon(y)$ consists of a single point, denoted $[x,y]$, for any $x,y\in\Omega(f)$ with $d(x,y)<\delta$. Further, $[x,y]\in\Omega(f)$.

DEFINITION 1.1.10: We say that f is <u>topologically transitive</u> if for every pair of open sets U and V there is an integer n such that $f^n(U)\cap V\neq\emptyset$. We say that f is <u>topologically mixing</u> if there is an integer N such that for every $n\geq N$, $f^n(U)\cap V\neq\emptyset$.

THEOREM 1.1.11 **Spectral Decomposition:** The nonwandering set $\Omega(f)$ can be decomposed into a finite number of pairwise disjoint closed sets Ω_i such that $\Omega(f)=\Omega_1\cup\Omega_2\cup\ldots\cup\Omega_s$ and

a) $f(\Omega_i)=\Omega_i$ and $f|_{\Omega_i}$ is topologically transitive

b) $\Omega_i=X_{1,i}\cup X_{2,i}\cup\ldots\cup X_{n_i,i}$ with the $X_{j,i}$ pairwise disjoint closed sets such that $f(X_{j,i})=X_{j+1,i}$ ($X_{n_i+1,i}=X_{1,i}$) and $f^{n_i}|_{X_{j,i}}$ is topologically mixing.

DEFINITION 1.1.12: The Ω_i in Theorem 1.1.11 are called basic sets.

We now define Markov partitions for Axiom A diffeomorphisms.

DEFINITION 1.1.13: Suppose Ω_s is a basic set for an Axiom A diffeomorphism. Then a subset $R \subset \Omega_s$ is called a rectangle if R has small diameter and $[x,y] \in R$ for any $x,y \in R$. R is called a proper rectangle if R is closed and $R = \overline{\text{int}(R)}$. By int(R), we mean the interior of R as a subset of Ω_s. Our rectangles will always be assumed to be proper.

DEFINITION 1.1.14: For $x \in R$, we define $W^s(x,R)$ and $W^u(x,R)$, the stable and unstable manifolds of x in R to be $W^s(x,R) = W^s_\epsilon(x) \cap R$ and $W^u(x,R) = W^u_\epsilon(x) \cap R$ respectively where the diameter of R is small in relation to ϵ.

LEMMA 1.1.15: Let R be a proper rectangle. As a subset of Ω_s, R has boundary

$$\partial R = \partial_s R \cup \partial_u R \text{ where}$$

$$\partial_s R = \{x \in R : x \notin \text{int}(W^u(x,R))\}$$

$$\partial_u R = \{x \in R : x \notin \text{int}(W^s(x,R))\}$$

and the interiors of $W^u(x,R)$ and $W^s(x,R)$ are as subsets of $W^u_\epsilon(x) \cap \Omega$ and $W^s_\epsilon(x) \cap \Omega$ respectively.

DEFINITION 1.1.16: A Markov partition for Ω_s is a finite covering $\mathcal{P} = \{R_1, R_2, \ldots, R_r\}$ of Ω_s where each R_i is a proper rectangle and

a) $\text{int}(R_i) \cap \text{int}(R_j) = \emptyset$ for $i \neq j$

b) $f(W^u(x,R_i)) \supset W^u(fx,R_j)$ and $f(W^s(x,R_i)) \subset W^s(fx,R_j)$

if $x \in \text{int}(R_i)$ and $fx \in \text{int}(R_j)$.

We conclude this section by showing how often Markov partitions occur in the Axiom A setting.

THEOREM 1.1.17: Let Ω_s be a basic set for an Axiom A diffeomorphism f. Then Ω_s has Markov partitions of arbitrarily small diameter.

SECTION 1.2: TORAL AUTOMORPHISMS

In this section we will apply the theory developed in the previous section to automorphisms of the two-dimensional torus. We begin with a discussion of toral automorphisms.

DEFINITION 1.2.1: We say that a matrix $\mathcal{A}$ is <u>hyperbolic</u> if none of the eigenvalues of $\mathcal{A}$ have modulus 1.

Suppose that a 2×2 matrix $\mathcal{A}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with integer entries is hyperbolic. Suppose further that $\det(\mathcal{A})=\pm 1$; that is, $\mathcal{A}\in GL(2,\mathbb{Z})$. Then $\mathcal{A}^{-1}$ is also an integer matrix with determinant ± 1 hence also in $GL(2,\mathbb{Z})$. The characteristic polynomial of $\mathcal{A}$ is $\chi(\lambda)=\lambda^2-\mathrm{Tr}(\mathcal{A})\lambda+\det(\mathcal{A})=\lambda^2-\mathrm{Tr}(\mathcal{A})\lambda\pm 1$. In order for $\mathcal{A}$ to be hyperbolic, $\mathrm{Tr}(\mathcal{A})\neq 0$. Since the product of the eigenvalues of $\mathcal{A}$ has to be ± 1, we must have real irrational eigenvalues λ_u and λ_s satisfying $|\lambda_u|>1>|\lambda_s|>0$; they are real and irrational by the Perron-Frobenius Theorem [10]. The eigenvectors corresponding to λ_u and λ_s are:

$$\begin{bmatrix} 1 \\ \frac{\lambda_u - a}{b} \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ \frac{\lambda_s - a}{b} \end{bmatrix}$$

respectively, denoted $\vec{v}_u$ and $\vec{v}_s$. The quantity $m_u = \frac{\lambda_u - a}{b}$ is the slope of $\vec{v}_u$ and similarly $m_s = \frac{\lambda_s - a}{b}$ is the slope of $\vec{v}_s$.

DEFINITION 1.2.2: We call λ_u the unstable eigenvalue of $\mathcal{A}$ and λ_s the stable eigenvalue of $\mathcal{A}$. The corresponding eigenvectors, $\vec{v}_u$ and $\vec{v}_s$, are called the unstable and stable eigenvectors respectively. We note at this time that if $\mathrm{Tr}(\mathcal{A})>0$, then $\lambda_u>0$ and if $\mathrm{Tr}(\mathcal{A})<0$, then $\lambda_u<0$.

We now consider the map that $\mathcal{A}$ induces on $\mathbf{R}^2$. ($\mathbf{R}$ is used to denote Euclidean space.) Since $\mathcal{A}$ is an integer matrix, $\mathcal{A}$ maps the integer lattice into itself. Further, the origin is always a fixed point of $\mathcal{A}$. We know that if (x,y) is on either of the lines $y=m_u x$ or $y=m_s x$ then $\mathcal{A}(x,y)=\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix}$ is also on that line. If (x,y) is on the line through (a,b) with slope m_u then $\mathcal{A}(x,y)$ is on the line through $\mathcal{A}(a,b)$ with slope m_u. To see this, suppose (x,y) is on such a line. Then (x,y) can be written as $(a,b)+k(1,m_u)$ for $k\in\mathbf{R}$. $\mathcal{A}(x,y)$ is then $\mathcal{A}((a,b)+k(1,m_u))$ $=\mathcal{A}(a,b)+k\lambda_u(1,m_u)$ which is on the line through $\mathcal{A}(a,b)$ with slope m_u. A similar result holds for line with slope m_s.

Let $\mathbf{T}^2$ be the space $\mathbf{R}^2/\mathbf{Z}^2$ where $\mathbf{Z}$ is the set of integers and $\mathbf{R}^2/\mathbf{Z}^2$ identifies (x,y) and $(x+\alpha,y+\beta)$ for $x,y\in\mathbf{R}$ and $\alpha,\beta\in\mathbf{Z}$. $\mathbf{T}^2$ can then be viewed as $[0,1]\times[0,1]$. Let π denote the map $\mathbf{R}^2 \xrightarrow{\pi} \mathbf{R}^2/\mathbf{Z}^2$. $\mathcal{A}$ induces an automorphism of $\mathbf{T}^2$ which we will also denote $\mathcal{A}$. Since $\mathcal{A}$ preserves the integer lattice, the following diagram commutes:

$$\begin{array}{ccc} \mathbf{R}^2 & \xrightarrow{\mathcal{A}} & \mathbf{R}^2 \\ \downarrow\pi & & \downarrow\pi \\ \mathbf{T}^2 & \xrightarrow{\mathcal{A}} & \mathbf{T}^2 \end{array}$$

DEFINITION 1.2.3: We call the automorphism induced by a 2×2 hyperbolic integer matrix $\mathcal{A}\in GL(2,\mathbb{Z})$ a hyperbolic toral automorphism or a toral Anosov map.

We note at this time that the following generalizes to $\mathbb{T}^n=\mathbb{R}^n/\mathbb{Z}^n$ and that all continuous automorphisms of $\mathbb{T}^n$ are determined by integer matrices in $GL(n,\mathbb{Z})$; the hyperbolic ones determined by the hyperbolic elements. There are discontinuous automorphisms of $\mathbb{T}^n$ but these turn out to be nonmeasurable [2]. This paper will deal only with continuous automorphisms of $\mathbb{T}^2$.

PROPOSITION 1.2.4: $W^u(x)$ for $x\in\mathbb{T}^2$ is the projection of a line L_u through $\pi^{-1}x$ parallel to $\vec{v}_u$ and $W^s(x)$ is the projection of a line L_s through $\pi^{-1}x$ parallel to $\vec{v}_s$.

Proof: We give the proof for $W^s(x)$ with the proof for $W^u(x)$ being similar. $W^s(x)=\{y\in\mathbb{T}^2: d(\mathcal{A}^n x,\mathcal{A}^n y)\to 0 \text{ as } n\to\infty\}$. Let $\pi^{-1}y$ be a point on L_s and let $I\subset L_s$ be the line segment joining $\pi^{-1}x$ and $\pi^{-1}y$ with length δ. The length of $\mathcal{A}^n I$ will then be $\lambda_s^n\delta\to 0$ as $n\to\infty$ since $\lambda_s<1$; therefore, $d(\mathcal{A}^n(\pi^{-1}x),\mathcal{A}^n(\pi^{-1}y))\to 0$ as $n\to\infty$ hence $d(\mathcal{A}^n x,\mathcal{A}^n y)\to 0$ as $n\to\infty$ and $y\in W^s(x)$. If y is not on L_s, then $y=c_1\vec{v}_u+c_2\vec{v}_s$ and $x=c_1'\vec{v}_u+c_2'\vec{v}_s$ with $c_1\neq c_1'$; for if $c_1=c_1'$, y is on L_s. Hence, $d(\mathcal{A}^n x,\mathcal{A}^n y)\geq\lambda_u^n|c_1-c_1'|$. Therefore as $n\to\infty$ the distance between $\mathcal{A}^n x$ and $\mathcal{A}^n y$ cannot approach zero and $y\notin W^s(x)$. □

PROPOSITION 1.2.5: For a hyperbolic toral automorphism $\mathcal{A}$, all of $\mathbb{T}^2$ is hyperbolic, every point $x\in\mathbb{T}^2$ is nonwandering, and $\mathcal{A}$ is topologically mixing on $\mathbb{T}^2$.

Proof: We can see that $T_x\mathbb{T}^2=W^u(x)\oplus W^s(x)$ and conditions b), c), and d) of Definition 1.1.1 are satisfied because $D\mathcal{A}=\mathcal{A}$. Every point $x\in\mathbb{T}^2$ is nonwandering because every neighborhood of x contains a line segment $I \subset W^u(x)$ and $\{\text{length of } \mathcal{A}^n(I)\} \to \infty$ as $n \to \infty$. $\mathcal{A}$ is mixing by a similar argument. □

This proposition tells us that $\mathcal{A}$ has one basic set, namely all of $\mathbf{T}^2$. Therefore by Theorem 1.1.17, there exist Markov partitions of $\mathbf{T}^2$ with arbitrarily small diameter. Before constructing a Markov partition for a hyperbolic automorphism $\mathcal{A}$ of $\mathbf{T}^2$, we make the following observations and definitions.

DEFINITION 1.2.6: We define the <u>boundary</u> of a Markov partition $\mathcal{P}=\{R_i\}_{i=1}^r$, denoted $\partial\mathcal{P}$, to be $\partial\mathcal{P}=\bigcup_{i=1}^r \partial R_i$. By Lemma 1.1.15, we know that $\partial R_i=\partial_u R_i \cup \partial_s R_i$. We then define the <u>unstable boundary of $\mathcal{P}$</u> as $\partial_u\mathcal{P}=\bigcup_{i=1}^r \partial_u R_i$ and the <u>stable boundary of $\mathcal{P}$</u> as $\partial_s\mathcal{P}=\bigcup_{i=1}^r \partial_s R_i$.

The R_i's for automorphisms of $\mathbf{T}^2$ can be unions of actual rectangles in the geometric sense. This will be true in this paper and for now we will assume that all rectangles are connected. Disconnected rectangles are covered in Chapter 2. Connected rectangles (when projected into $\mathbf{R}^2$) have two sides parallel to a line through the origin with slope m_u and two sides parallel to a line through the origin with slope m_s. From the definition of Markov partition (1.1.16), we see that *i)* $\mathcal{A}^{-1}(\partial_u\mathcal{P}) \subset \partial_u\mathcal{P}$ and *ii)* $\mathcal{A}(\partial_s\mathcal{P}) \subset \partial_s\mathcal{P}$. This allows us to define the following:

DEFINITION 1.2.7: The <u>unstable core of $\mathcal{P}$</u> , denoted $\mathcal{U}$, is defined as $\mathcal{U}=\bigcap_{i=1}^{\infty} \mathcal{A}^{-i}(\partial_u\mathcal{P})$. Similarly, define the <u>stable core of $\mathcal{P}$</u>, denoted $\mathcal{S}$, as $\mathcal{S}=\bigcap_{i=1}^{\infty} \mathcal{A}^{i}(\partial_s\mathcal{P})$. We then define the <u>core of $\mathcal{P}$</u>, denoted $\mathcal{C}$, as $\mathcal{C}=\mathcal{U}\cup\mathcal{S}$.

Chapter 2 will discuss the composition of the core. From b) in the definition of Markov partition (1.1.16), we know that if $\mathcal{A}(R_i)$ intersects R_j, $\mathcal{A}(R_i)$ crosses R_j from one end to the other in the unstable direction. Also, if $\mathcal{A}(W^s(x,R_i))\cap\text{int}(R_j)\neq\emptyset$, then $\mathcal{A}(W^s(x,R_i))\subset R_j$. Hence, we can define a transition matrix as follows:

DEFINITION 1.2.8: We define the Markov matrix for a Markov partition $\mathcal{P}$ with r rectangles for $\mathcal{A}$ by $m_{i,j}=$ {the number of times $\mathcal{A}(R_i)$ crosses the interior of R_j} for $1\leq i,j\leq r$.

EXAMPLE 1.2.9: Let $\mathcal{A}=\begin{bmatrix}2&1\\1&1\end{bmatrix}$. Then the eigenvalues of $\mathcal{A}$ are $\lambda_u=\frac{3+\sqrt{5}}{2}$ and

$\lambda_s=\frac{3-\sqrt{5}}{2}$ with corresponding eigenvectors $\begin{bmatrix}1\\ \frac{\sqrt{5}-1}{2}\end{bmatrix}$ and $\begin{bmatrix}1\\ \frac{-\sqrt{5}-1}{2}\end{bmatrix}$ respectively.

Figures 1.2.10—1.2.12 are a Markov partition $\mathcal{P}$ for $\mathcal{A}$, $\mathcal{A}(\mathcal{P})$, and $\mathcal{A}^{-1}(\mathcal{P})$. We draw $\mathbb{T}^2$ as $[0,1]\times[0,1]$ with opposite sides identified. We see that $\mathcal{A}(R_1)$ crosses R_1 twice and R_2 once and $\mathcal{A}(R_2)$ crosses R_1 and R_2 each once giving us $M=\begin{bmatrix}2&1\\1&1\end{bmatrix}$. Notice that if we reverse the labeling of R_1 and R_2 then $M=\begin{bmatrix}1&1\\1&2\end{bmatrix}$. The core of $\mathcal{P}$, as seen from the diagrams, is simply a single fixed point drawn at the origin.

SECTION 1.3: SYMBOLIC DYNAMICS

The importance of Markov partitions is that they allow us to use the tools of symbolic dynamics in the study of toral automorphisms. We begin by introducing the basics of symbolic dynamics.

Let A be an $r\times r$, non-negative, integral matrix. Let G(A) be the directed graph with states $\{1,2,\ldots,r\}$ and for $0\leq i,j\leq r$, G(A) has $A_{i,j}$ edges from state i to state j. Let $\mathcal{E}$ denote the set of edges of G(A).

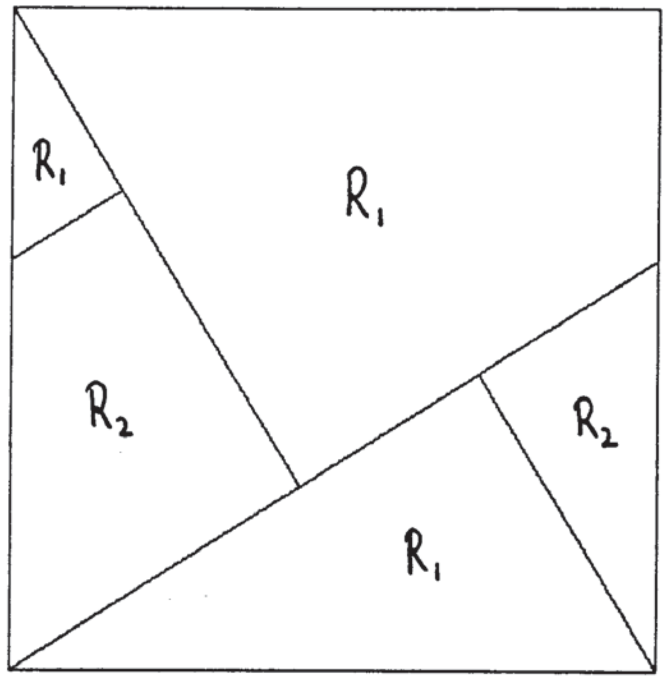

Figure 1.2.10: The Partition $\mathcal{P}$ of Example 1.2.9

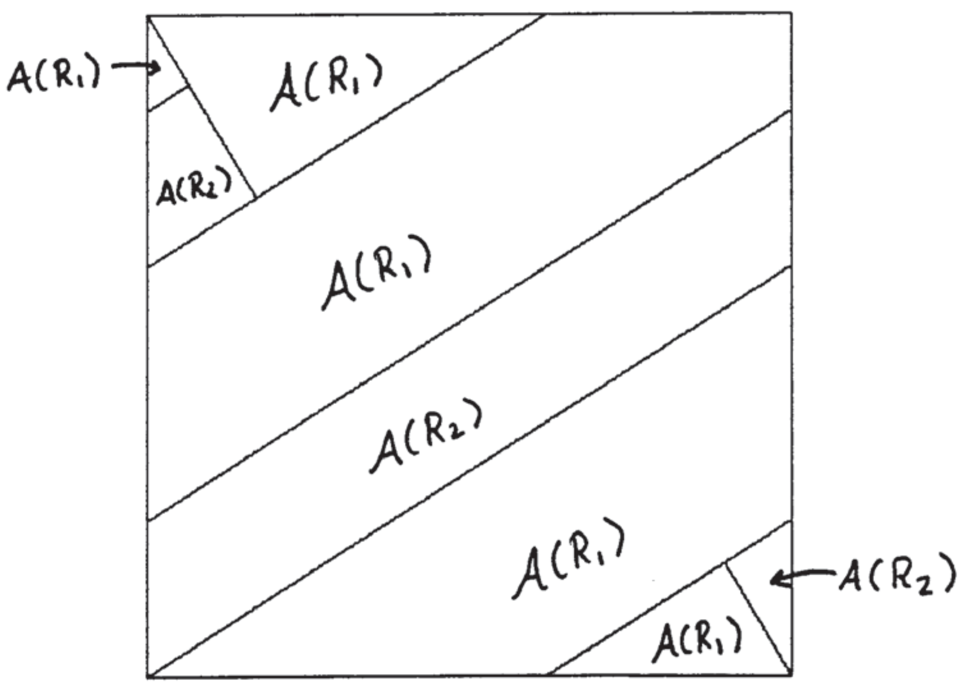

Figure 1.2.11: $\mathcal{A}(\mathcal{P})$

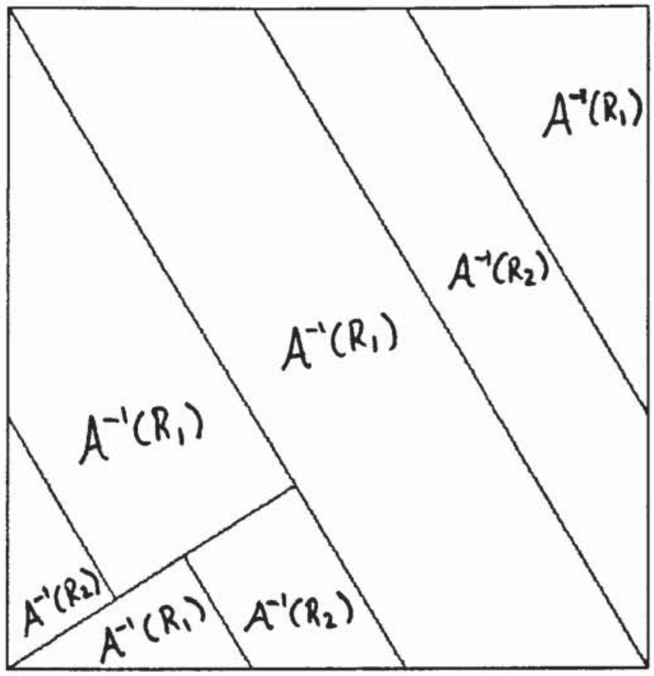

Figure 1.2.12: $\mathcal{A}^{-1}(\mathcal{P})$

DEFINITION 1.3.1: We say that edge f follows edge e if the initial vertex of f is the terminal vertex of e. We denote the set of all followers of a given state e by $\mathcal{F}(e)$.

DEFINITION 1.3.2: We define the shift of finite type, abbreviated SFT, also called a topological Markov shift,

$$\Sigma_A = \{x \in \mathcal{E}^{\mathbf{Z}} : x_{i+1} \text{ follows } x_i \text{ for all } i \in \mathbf{Z}\}$$

with the product topology.

We also define the shift map $\sigma_A : \Sigma_A \rightarrow \Sigma_A$ by

$$\sigma_A(x) = y \text{ where } y_i = x_{i+1} \text{ for all } i \in \mathbf{Z}.$$

We will often use Σ_A to denote the pair (Σ_A, σ_A). The elements of $\mathcal{E}$ will often be referred to as the symbols of Σ_A. We abbreviate σ_A by just σ. A shift defined in this way is called an edge shift for obvious reasons.

EXAMPLE 1.3.3: If A is the 1×1 matrix [k], then $\mathcal{S}=\{1,2,\ldots,k\}$ and $\Sigma_A=\{1,2,\ldots,k\}^{\mathbb{Z}}$; in other words, Σ_k is all bi-infinite sequences consisting of the elements of $\mathcal{S}$. Σ_k is called the full k shift or a Bernoulli shift. See Figure 1.3.4 for a diagram of the full 2 shift.

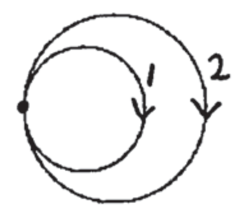

Figure 1.3.4: A Full 2 Shift - Edge Shift

DEFINITION 1.3.5: If A has only 0's and 1's as entries, we can define a SFT called a vertex shift as follows: Let G(A) be the directed graph with states $S=\{1,2,\ldots,r\}$ and for $0\leq i,j\leq r$, G(A) has an edge from state i to state j if and only if $A_{i,j}=1$. We then define

$$\Sigma_A=\{x\in S^{\mathbb{Z}}: A_{x_i,x_{i+1}}=1 \text{ for all } i\in\mathbb{Z}\}$$

and σ_A as before.

EXAMPLE 1.3.6: Figure 1.3.7 is the graph for a full 2 shift. The matrix for the shift is $A=\begin{bmatrix}1&1\\1&1\end{bmatrix}$.

Figure 1.3.7: A Full 2 Shift - Vertex Shift

We note at this time that an edge shift as defined above is topologically conjugate (defined below) to a vertex shift and we will use both constructions as we need them.

DEFINITION 1.3.8: A subshift is a closed, σ invariant subset of Σ_k. All SFT's are clearly subshifts.

DEFINITION 1.3.9: A non-negative integral matrix A is said to be irreducible if for each pair of states i,j, there is an integer $m \geq 0$ such that $A^m_{i,j} > 0$. If there is an integer $m \geq 0$ such that $A^m > 0$ then A is called aperiodic.

We recall the definitions of topologically transitive and topologically mixing from Section 1.1. If we restrict these definitions to positive integers, we have defined forward transitive and mixing for SFT's. In terms of the graph of Σ_A, topologically transitive means that we can find a path from any edge to any other edge including itself. Mixing means that for some m we can find a path of length m from any edge to any other edge including itself. It is well known that Σ_A is forward transitive $\Leftrightarrow$ A is irreducible and Σ_A is mixing $\Leftrightarrow$ A is aperiodic.

There is a metric d on Σ_A defined as follows: if $x,y \in \Sigma_A$, then $d(x,y) = \sum_{i=-\infty}^{\infty} 2^{|i|}\delta(x_i,y_i)$ where $\delta(x_i,y_i)=1$ if $x_i \neq y_i$ and 0 if $x_i = y_i$. We can then see $W^s(x) = \{y : d(\sigma^n_A x, \sigma^n_A y) \to 0 \text{ as } n \to \infty\}$ $= \{y : y_i = x_i \text{ for } i \leq n\}$ and $W^u(x) = \{y : y_i = x_i \text{ for } i \geq n\}$. If x and y are close enough, namely if $x_0 = y_0$, then $W^s(x) \cap W^u(y)$ is a single point $z = (\ldots, x_{-2}, x_{-1}, x_0 = y_0, y_1, y_2, \ldots)$. Therefore if $x_0 = y_0$, then $[x,y] = z$ in the sense of Definition 1.1.9. We can then construct a Markov partition for (Σ_A, σ_A).

EXAMPLE 1.3.10: Let $A=[k]$. We define rectangles $R_i=\{y:y_0=i\}$ for $1\leq i\leq k$. These are clearly closed (and open), cover all of Σ_A, and satisfy both the $[\cdot,\cdot]$ condition and the stable and unstable manifold condition in Definition 1.1.16. We define the transition matrix in a slightly different fashion: $M_{i,j}=1$ if $\sigma(R_j)\cap R_i\neq\emptyset$ and 0 otherwise.

DEFINITION 1.3.11: Let X and Y be two topological spaces and let $f:X\to X$ and $g:Y\to Y$ be two continuous maps. A continuous, surjective map $h:X\to Y$ is said to be a <u>semiconjugacy</u> or a <u>factor map</u> if $hf=gh$; that is, if the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \downarrow h & & \downarrow h \\ Y & \xrightarrow{g} & Y \end{array}$$

If in addition h is a homeomorphism, we call h a <u>topological conjugacy</u> and say that f and g are <u>topologically conjugate</u>.

It is known that every non-negative integer matrix is topologically conjugate to an integer matrix with only 0's and 1's as entries. Also, for a given Markov partition of $\mathbf{T}^2$, we may always divide our rectangles so that all entries in the Markov matrix are either 0's or 1's; in other words, $\mathcal{A}(R_i)$ crosses R_j at most once for all i and j. As a matter of fact, Adler and Weiss [2] used this form of the partition in their original paper. Hence, for the rest of this section, we will assume that both the matrix A for a SFT and the Markov matrix M for a Markov partition of $\mathbf{T}^2$ are 0-1 matrices and use larger integers only for simplicity in examples.

We now consider Σ_M where M is the Markov matrix for a Markov partition $\mathcal{P}$ for a toral automorphism $\mathcal{A}$. $M_{i,j}=1$ if $\mathcal{A}(R_i)$ crosses R_j once. Therefore, in Σ_M, there is an edge from state i to state j if and only if $\mathcal{A}(R_i)$ crosses R_j. We then define a map $h:\Sigma_M\to\mathbf{T}^2$ by $h(a)=x$

where $\mathcal{A}^n(x) \in R_{a_n}$ for $n \in \mathbb{Z}$.

THEOREM 1.3.12: For each $a \in \Sigma_M$, the set $\bigcap_{n \in \mathbb{Z}} \mathcal{A}^{-n} R_{a_n}$ consists of a single point denoted $h(a)$. The map h defined above is a continuous surjection, $h\sigma = \mathcal{A}h$, and h is one-to-one over the residual set $\mathcal{I} = \{y : \mathcal{A}^n(y) \notin \partial\mathcal{P} \text{ for all } n \in \mathbb{Z}\}$.

In other words, the following diagram commutes:

$$\begin{array}{ccc} \Sigma_M & \xrightarrow{\sigma} & \Sigma_M \\ \downarrow h & & \downarrow h \\ \mathbb{T}^2 & \xrightarrow{\mathcal{A}} & \mathbb{T}^2 \end{array}$$

Proof: The proof for Axiom A diffeomorphisms in general is in Bowen [5].

In assuming that M is a 0-1 we assumed that the rectangles in $\mathcal{P}$ were small enough such that the above intersection was a single point. See Chapter 2 for examples of partitions which are tolopogically generating and some which are not. We make that notion more formal.

DEFINITION 1.3.13: A partition $\mathcal{P}$ is said to be <u>topologically generating</u> if the above intersection is a single point for any $a \in \Sigma_M$.

We can now see the importance of the Markov matrix. The behavior of the SFT Σ_M is relatively easy to study and gives us information about the dynamics of $(\mathbb{T}^2, \mathcal{A})$. Chapters 3 and 4 of this paper are devoted to the study of M. Section 1.4 gives some elementary results about M.

SECTION 1.4: THE MARKOV MATRIX

In this section we present a few elementary propositions dealing with the Markov matrix.

PROPOSITION 1.4.1: Let M be the Markov matrix for a hyperbolic toral automorphism $\mathcal{A}$ and a Markov partition $\mathcal{P}$ with r rectangles. Then M^n is the Markov matrix for $\mathcal{P}$ under $\mathcal{A}^n$ for $n \in \mathbb{Z}_+$.

Proof: We prove by induction. $n=1$ is trivial. Therefore, let $n=2$. We want to compute how many times R_i crosses R_j under $\mathcal{A}^2$. If $\mathcal{A}R_k$ crosses R_j $m_{k,j}$ times and $\mathcal{A}R_i$ crosses R_k $m_{i,k}$ times, then we can see that $\mathcal{A}^2R_i$ crosses R_j exactly $m_{k,j}m_{i,k}$ times. Therefore, $\mathcal{A}^2R_i$ crosses R_j exactly $\sum_{k=1}^{r} m_{k,j}m_{i,k}$ which is exactly the i,j entry of M^2. We now assume our result is true for any $1 \leq i \leq n-1$ and a proof analogous to the one above will prove the result for M^n. □

PROPOSITION 1.4.2: Let $\mathcal{P}$ be a Markov partition for $\mathcal{A}$ with Markov matrix M. Then $\mathcal{P}$ is a Markov partition for $\mathcal{A}^{-1}$ with Markov matrix M^T.

Proof: Because an unstable (stable) eigenvector for $\mathcal{A}$ and eigenvalue λ is a stable (unstable) eigenvector for $\mathcal{A}^{-1}$ with eigenvalue $\frac{1}{\lambda}$, it is clear that $\mathcal{P}$ is a Markov partition for $\mathcal{A}^{-1}$. If $\mathcal{A}R_i$ crosses R_j k times, then under $\mathcal{A}^{-1}$, R_j will cross R_i k times under $\mathcal{A}^{-1}$ so that the Markov matrix for $\mathcal{A}^{-1}$ and $\mathcal{P}$ is M^T. □

PROPOSITION 1.4.3: Suppose there is a $\phi \in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi = \mathcal{B}$ where $\mathcal{A}$ and $\mathcal{B}$ are hyperbolic toral automorphisms. Then if $\mathcal{P}$ is a Markov partition for $\mathcal{A}$ with Markov matrix M, then $\phi^{-1}\mathcal{P}$ is a Markov partition for $\mathcal{B}$ with Markov matrix M.

Proof: By $\phi^{-1}\mathcal{P}$, we mean the image of $\mathcal{P}$ under ϕ^{-1} with $\phi^{-1}R_i = R_i' \in \phi^{-1}\mathcal{P}$. Since ϕ is a topological conjugacy between the two systems, $\mathcal{P}$ is a Markov partition. Another way to see this is that if $\vec{v}$ is an eigenvector for $\mathcal{A}$ with corresponding eigenvalue λ, then $\phi^{-1}\vec{v}$ is an eigenvector for $\mathcal{B}$ with the corresponding eigenvalue λ. If $\mathcal{A}R_i$ crosses R_j, then is clear that $\mathcal{B}R_i'$ crosses R_j'. □

CHAPTER 2 - THE CORE AND THE BOUNDARY

SECTION 2.1: DISCONNECTED RECTANGLES

In this section we begin our study of the core and its structure. As seen from Definition 1.2.7, the core is a closed, totally disconnected $\mathcal{A}$-invariant set which is finitely presented and hence a sofic system. Ashley, Kitchens, and Stafford [3] have shown the following about the structure of the core:

THEOREM 2.1.1: Let $\mathcal{A}$ be a hyperbolic toral automorphism. If Λ is any sofic system such that *i*) Λ has lower entropy than $\mathcal{A}$ and *ii*) Λ has fewer periodic points of every period than $\mathcal{A}$ then there is a Markov partition for $\mathcal{A}$ with $\mathfrak{C}=\Lambda$.

In order to see how such large cores can occur, we must first investigate the notion of disconnected rectangles. We must develop critera for when two rectangles can be considered as one rectangle and further define the stable and unstable manifolds in such a situation.

In Lemma 1.1.15, we saw that the boundary of any rectangle can actually be viewed as two sets: an unstable boundary and a stable boundary. In fact, in the case of the two-torus, the whole rectangle can be viewed as the product $\partial_u\hat{\mathcal{P}}\times\partial_s\hat{\mathcal{P}}$ where $\partial_u\hat{\mathcal{P}}$ is one of the two unstable sides of the rectangle and $\partial_s\hat{\mathcal{P}}$ is one of the two stable sides. It is this notion which will allow us to define disconnected rectangles.

Suppose the images of two rectangles cross the same rectangles under $\mathcal{A}$. This would mean that in the Markov matrix, these two rectangles have identical corresponding rows. They have the same "personality" and we would like to consider them as one rectangle. If they are adjacent, that is, they share an unstable boundary, then we can simply remove that line (if that would still give us a Markov partition) and we now have just one rectangle. Stable and unstable manifolds in this new rectangle cause us no problem. However, if the original rectangles are not adjacent, we need to be a bit more careful. For this reason, we now define $W^u(x,R_i)$ and $W^s(x,R_i)$ in a different way. We begin by defining them for connected rectangles and then expand that notion to include disconnected rectangles.

DEFINITION 2.1.2: Let R be a closed, connected region in $\mathbf{T}^2$ such that $R=\overline{\text{int}(R)}$. Let $\hat{R}$ be a closed, connected region in $\mathbf{R}^2$ such that

a) $\hat{R}=\overline{\text{int}(\hat{R})}$

b) $\pi{:}\mathbf{R}^2\rightarrow\mathbf{T}^2$ maps $\text{int}(\hat{R}) \mapsto \text{int}(R)$ bijectively

c) π maps $\partial\hat{R}\mapsto\partial R$ surjectively in a finite-to-one fashion.

We call $\hat{R}$ a <u>covering</u> <u>region</u> for R. The covering region is simply one of the preimages of R under π in $\mathbf{R}^2$.

We recall that $W^u(x)$ for $x\in\mathbf{T}^2$ is the projection of a line in $\mathbf{R}^2$ through $\pi^{-1}x$. We will call this line in $\mathbf{R}^2$ $W^u(\pi^{-1}x)$.

DEFINITION 2.1.3: Let R be a closed, connected region in $\mathbf{T}^2$ with covering region $\hat{R}$. For $x\in\text{int}R$ with $\hat{x}\in\text{int}\hat{R}$ such that $\pi(\hat{x})=x$, define $W^u(x,R)=\pi(W^u(\hat{x})\cap\hat{R})$. Similarly, define $W^s(x,R)=\pi(W^s(\hat{x})\cap\hat{R})$.

DEFINITION 2.1.4: For $x,y \in \mathbf{T}^2$ and $\hat{x},\hat{y} \in \mathbf{R}^2$ such that $\pi(\hat{x})=x$ and $\pi(\hat{y})=y$, we define $[x,y]'$ to be $\pi\left(W^s(\hat{x}) \cap W^u(\hat{y})\right)$. This intersection will always be a single point.

We replace the bracket condition in Definition 1.1.13 by $[\cdot,\cdot]'$. This condition also eliminates another problem with our other definition. In previous examples, we had ignored the stipulation that our rectangles be of "small diameter". In the case of the two-torus, we can see Markov-like behavior in partitions with large rectangles.

DEFINITION 2.1.5: In a Markov partition, the cardinality of $\mathcal{A}W^u(x,R_i) \cap W^s(y,R_j)$ must be independent of the choice of $x \in R_i$ and $y \in R_j$. We call this the <u>*i,j*-cardinality</u>.

We now must be sure that the *i,j*-cardinality remains independent of choice when we discuss disconnected rectangles. There are two types of rectangles which we would like to combine: those which have identical rows in the Markov matrix as mentioned above and those which have identical columns, in which case these rectangles are covered by exactly the same rectangles.

DEFINITION 2.1.6: Two rectangles R_i and R_j are said to be <u>forward equivalent</u> if rows i and j in the Markov matrix M are identical. They are said to be <u>inverse equivalent</u> if columns i and j in the Markov matrix are identical.

Suppose R_i and R_j are forward equivalent and suppose their images under $\mathcal{A}$ both cross R_k. Let $R_l = R_i \cup R_j$. Then the *l,k*-cardinality is independent of the choice of $z \in R_k$ and $x \in R_l$. Therefore we want $W^u(x,R_l)$ to be $W^u(x,R_i)$ for $x \in R_i$ and similarly for $y \in R_j$. We must, however, alter the definition of $W^s(x,R_l)$.

DEFINITION 2.1.7: Suppose R_i and R_j are forward equivalent and let $R_l = R_i \cup R_j$. Define

$$W^s(x,R_l) = R_l \cap \mathcal{A}^{-1}(W^s(\mathcal{A}x,R_k)) \text{ if } \mathcal{A}W^s(x,R_i) \cap \text{int}R_k \neq \emptyset \text{ for } x \in R_i \text{ and}$$

similarly for $x \in R_j$.

The stable manifold for each point in R_l will have two components using the above definition: one in what was formerly R_i and one in what was formerly R_j. Now, if $\mathcal{A}R_k$ crosses R_l, the k,l-cardinality will be independent of the choice of $x \in R_l$ and the new partition with $R_l = R_i \cup R_j$ will be Markov.

Suppose now that R_i and R_j are inverse equivalent and let $R_l = R_i \cup R_j$. By an argument similar to the one above, we see that $W^s(x,R_l)$ is simply $W^s(x,R_i)$ for $x \in R_i$ and similarly for $y \in R_j$.

DEFINITION 2.1.8: Suppose R_i and R_j are inverse equivalent and let $R_l = R_i \cup R_j$. Define

$$W^u(x,R_l) = R_l \cap \mathcal{A}(W^u(\mathcal{A}^{-1}x,R_k)) \text{ if } \mathcal{A}^{-1}(W^s(x,R_i)) \cap \text{int}R_k \neq \emptyset \text{ for } x \in R_i \text{ and}$$

similarly for $x \in R_j$.

In this case, the unstable manifold of each point in R_l will have two components.

In the above discussion, we never used the fact that the R_i were connected, only that they had well defined stable and unstable manifolds for each point in R_i. Using the above process, we can have rectangles with a countable number of components.

Having completed the introductory material needed to study disconnected rectangles, we now classify partitions with simple cores using the following definitions.

DEFINITION 2.1.9: A Markov partition $\mathcal{P}$ is said to be core-connected if every rectangle in $\mathcal{P}$ is connected and

i) $W^u(x) \cap \partial_u \mathcal{P}$ is connected for all $x \in \mathcal{U}$ and

ii) $W^s(x) \cap \partial_s \mathcal{P}$ is connected for all $x \in \mathcal{S}$.

DEFINITION 2.1.10: A Markov partition $\mathcal{P}$ is said to satisfy the finite core condition, or FCC, if every rectangle has a finite number of components and

i) $W^u(x) \cap \partial_u \mathcal{P}$ has a finite number of components for all $x \in \mathcal{U}$ and

ii) $W^s(x) \cap \partial_s \mathcal{P}$ has a finite number of components for all $x \in \mathcal{S}$.

Clearly every FCC partition can be made into a partition which is core-connected by "filling in" any gaps in the boundary and giving each component of each rectangle a unique labeling. FCC partitions have a finite core and are often what we want to work with.

SECTION 2.2: EXAMPLES OF THE CORE

The purpose of this section is to give a number of examples of Markov partitions for various automorphisms and discuss the composition of the core and the structure of the boundary.

Example 1.2.9 gives an example where $\mathcal{U} = \mathcal{S} = \mathcal{C} = \{\text{one fixed point}\}$. We will often assume without loss of generality that if $\mathcal{C}$ consists of a single fixed point, it is the origin, which we will denote **0**; for if such a fixed point was not at the origin, we could translate the entire partition so that it was. Also, in diagrams we will often abbreviate R_i, $\mathcal{A}(R_i)$, and $\mathcal{A}^{-1}(R_i)$ with just an *i* when the context is clear.

EXAMPLE 2.2.1: See Figures 2.2.3-2.2.5. This partition for $\mathcal{A}=\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ has $\mathcal{U}=\{$the periodic orbit $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0)$, $(0,\frac{1}{2})\}$ and $\mathcal{S}=\{$the periodic orbit $(\frac{3}{5},\frac{1}{5})$, $(\frac{2}{5},\frac{4}{5})\}$. We see then that $\mathcal{U}$ is not necessarily equal to $\mathcal{S}$. The Markov matrix is:

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

EXAMPLE 2.2.2: See Figures 2.2.6-2.2.8. $\mathcal{A}=\begin{bmatrix} 3 & -1 \\ 1 & 0 \end{bmatrix}$. This example has $\mathfrak{C}=0$ but $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ are not connected. Notice that we could draw in $\overline{AB}$ and $\overline{CD}$ and we would then have $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ connected. R_4 also has two components. This partition satisfies FCC but is not core-connected. The Markov matrix is

$$M = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

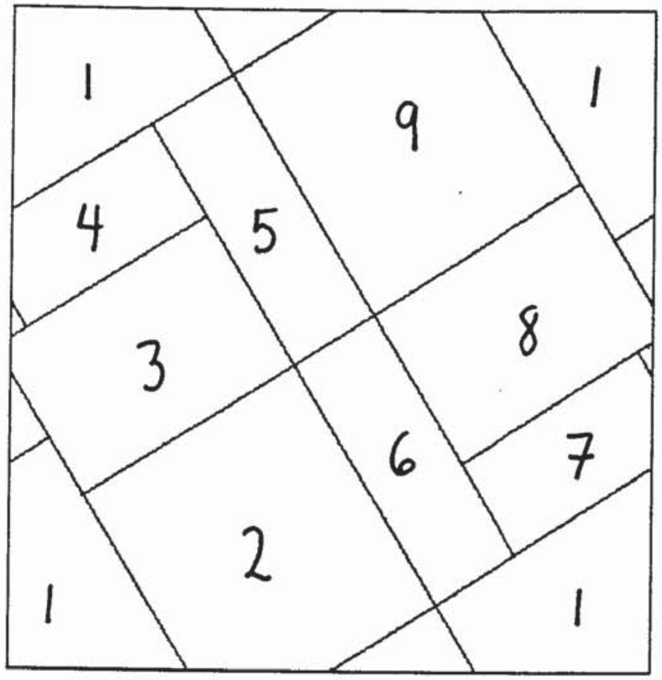

Figure 2.2.3: The Partition $\mathcal{P}$ of Example 2.2.1

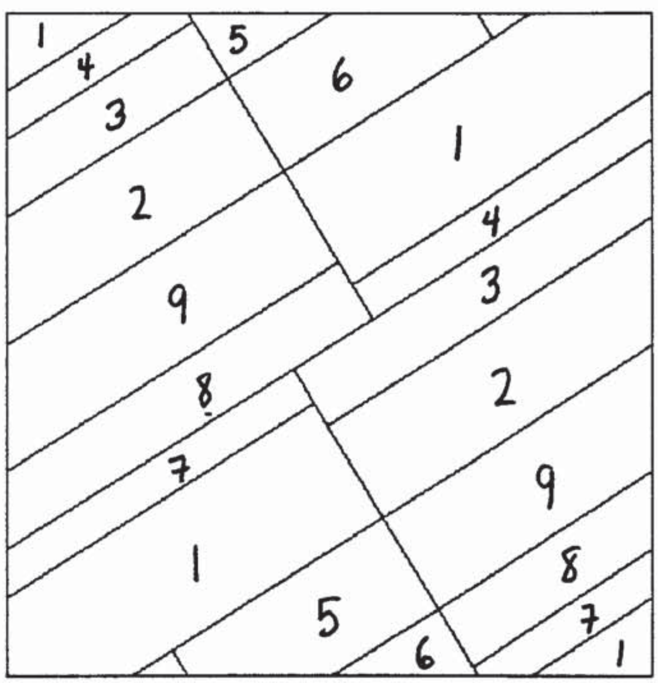

Figure 2.2.4: $\mathcal{A}(\mathcal{P})$

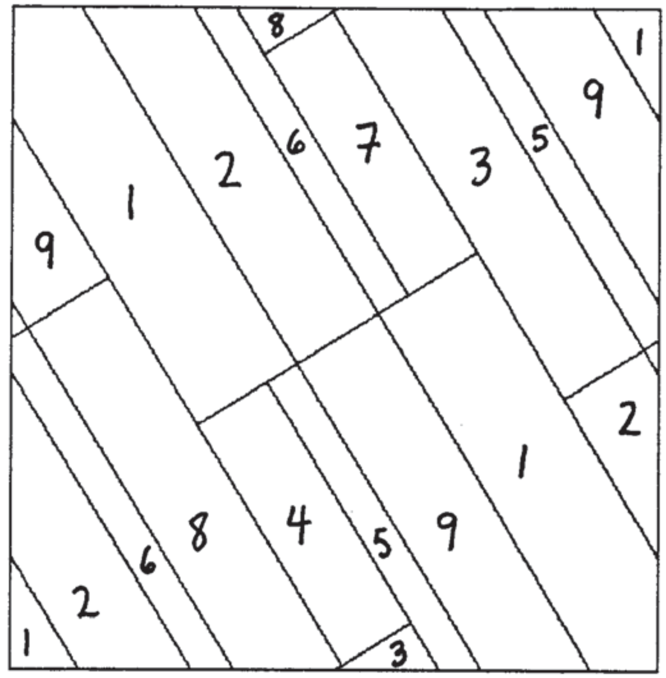

Figure 2.2.5: $\mathcal{A}^{-1}(\mathcal{P})$

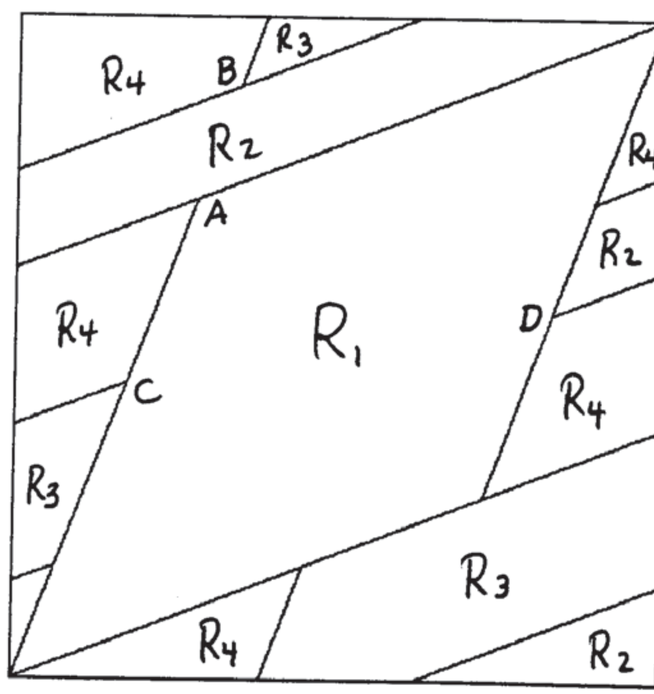

Figure 2.2.6: The Partition $\mathcal{P}$ of Example 2.2.2

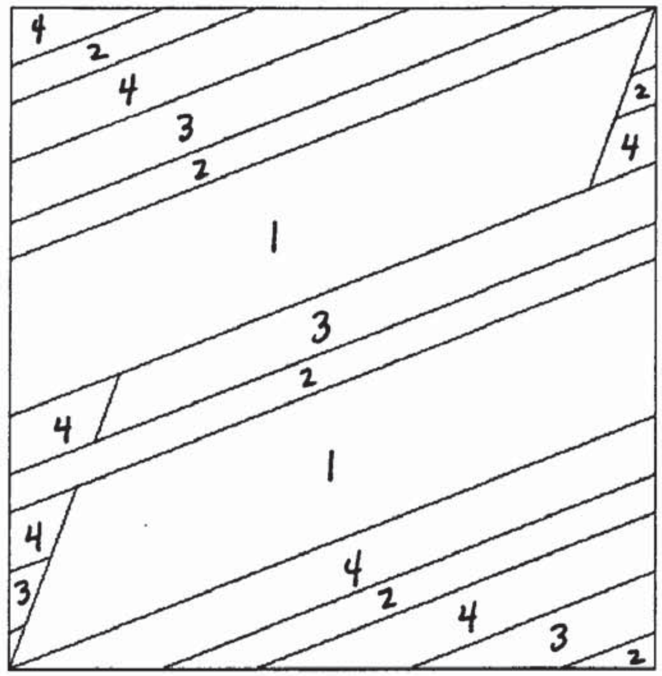

Figure 2.2.7: $\mathcal{A}(\mathcal{P})$

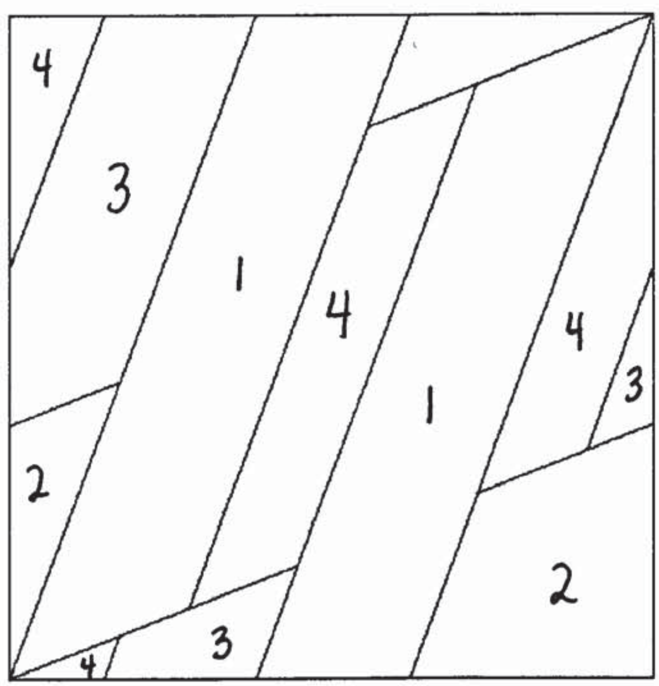

Figure 2.2.8: $\mathcal{A}^{-1}(\mathcal{P})$

EXAMPLE 2.2.9: See Figures 2.2.11-2.2.13. $\mathcal{A}=\begin{bmatrix}5&3\\3&2\end{bmatrix}=\begin{bmatrix}2&1\\1&1\end{bmatrix}^2$. This partition has $\mathcal{S}=\{$the two fixed points $(\frac{1}{5},\frac{2}{5})$, $(\frac{4}{5},\frac{3}{5})\}$ while $\mathcal{U}=\{$two fixed points $(\frac{3}{5},\frac{1}{5})$, $(\frac{2}{5},\frac{4}{5})\}$. The Markov matrix is

$$M=\begin{bmatrix} 4 & 2 & 0 & 1 \\ 2 & 5 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \end{bmatrix}.$$

This partition is not topologically generating. Notice that R_1 crosses itself four times. By Proposition 1.4.2, we know that $\mathcal{A}^{-1}R_1$ also crosses R_1 four times. Figure 2.2.14 is an enlargement of $R_1 \cap \mathcal{A}R_1$. Figure 2.2.15 shows $R_1 \cap \mathcal{A}R_1 \cap \mathcal{A}^2R_1$. As we continue this process, we see that there are points which stay in R_1 for all forward time. Using a standard construction, we see that we have a Smale horseshoe with a full four shift contained in R_1. The points which stay in R_1 for all forward time form a Cantor set of lines in the stable direction. The points which stay in R_1 for all backward time form a Cantor set of lines in the unstable direction. The points which stay in R_1 for all time are the intersection of these two Cantor sets of lines which is a full four shift. We now construct a partition with this set as its core.

EXAMPLE 2.2.10: See Figure 2.2.16. This partition is for $\mathcal{A}=\begin{bmatrix}5&3\\3&2\end{bmatrix}=\begin{bmatrix}2&1\\1&1\end{bmatrix}^2$ and is the same as in the previous example except we have broken R_1 up into three rectangles labeled A, B, and C in the following way. When we consider $R_1 \cap \mathcal{A}R_1$, we see three regions in R_1 which are not covered by $\mathcal{A}R_1$ which we have labeled A, B, and C. When we consider $R_1 \cap \mathcal{A}R_1 \cap \mathcal{A}^2R_1$, we see a similar pattern in each component of $R_1 \cap \mathcal{A}R_1$. In each component, we label these regions similarly and continue this process indefinitely. Notice that each component labeled A is forward equivalent to every other component labeled A and similarly for B and C. Therefore, each point in A has stable manifold in every component of A. This partition has

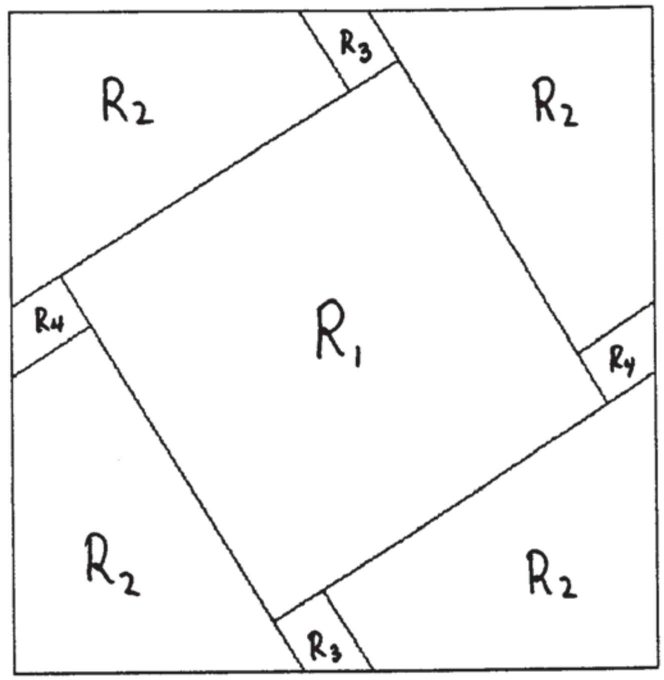

Figure 2.2.11: The Partition $\mathcal{P}$ of Example 2.2.9

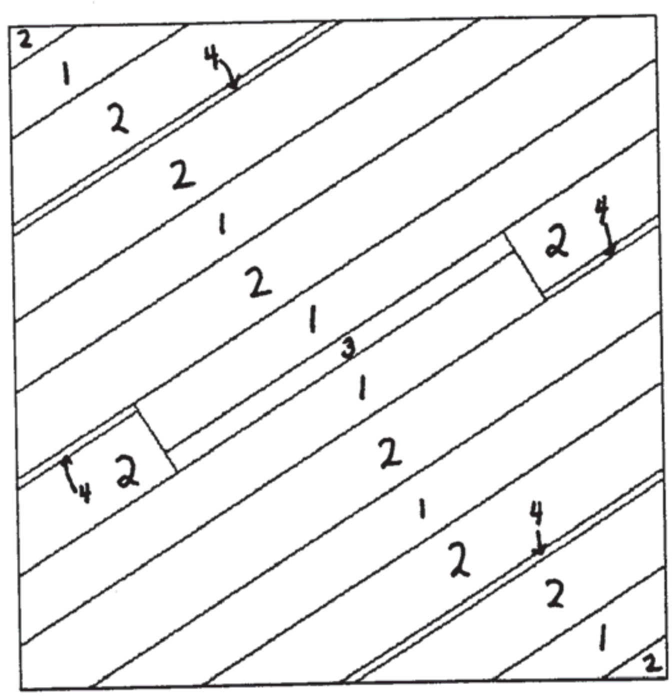

Figure 2.2.12: $\mathcal{A}(\mathcal{P})$

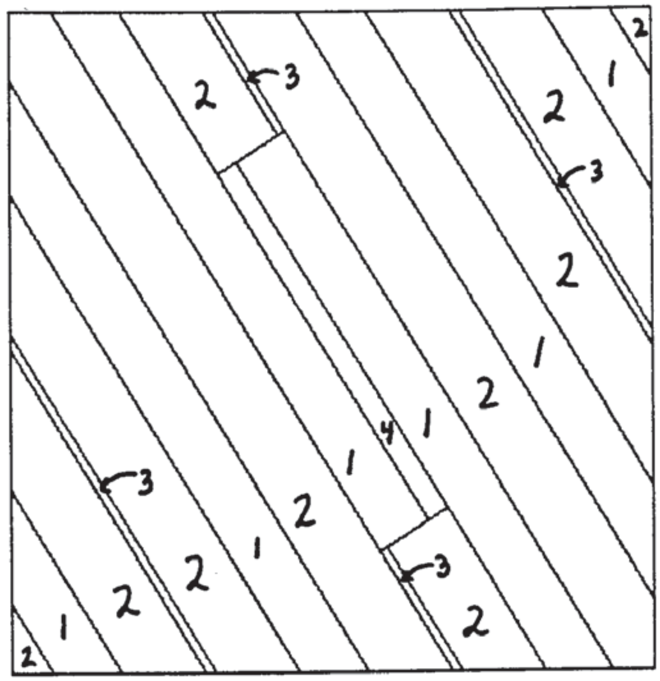

Figure 2.2.13: $\mathcal{A}^{-1}(\mathcal{P})$

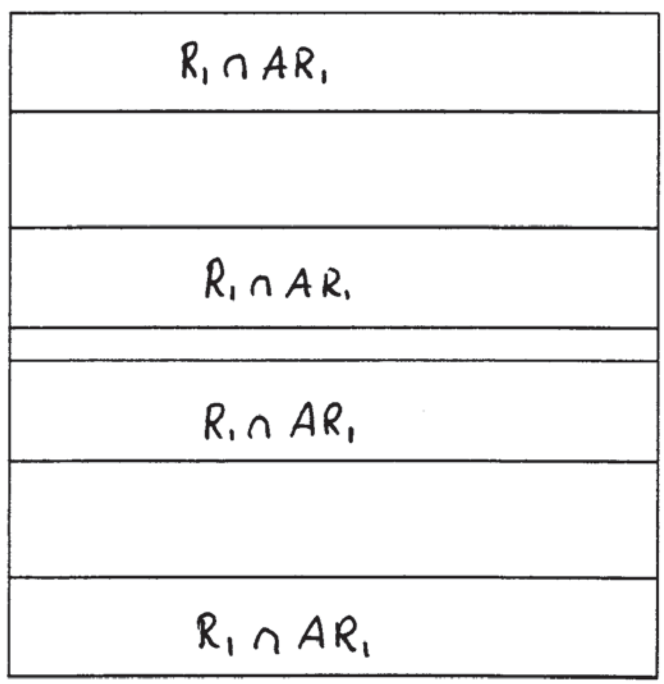

Figure 2.2.14: $R_1 \cap \mathcal{A}R_1$

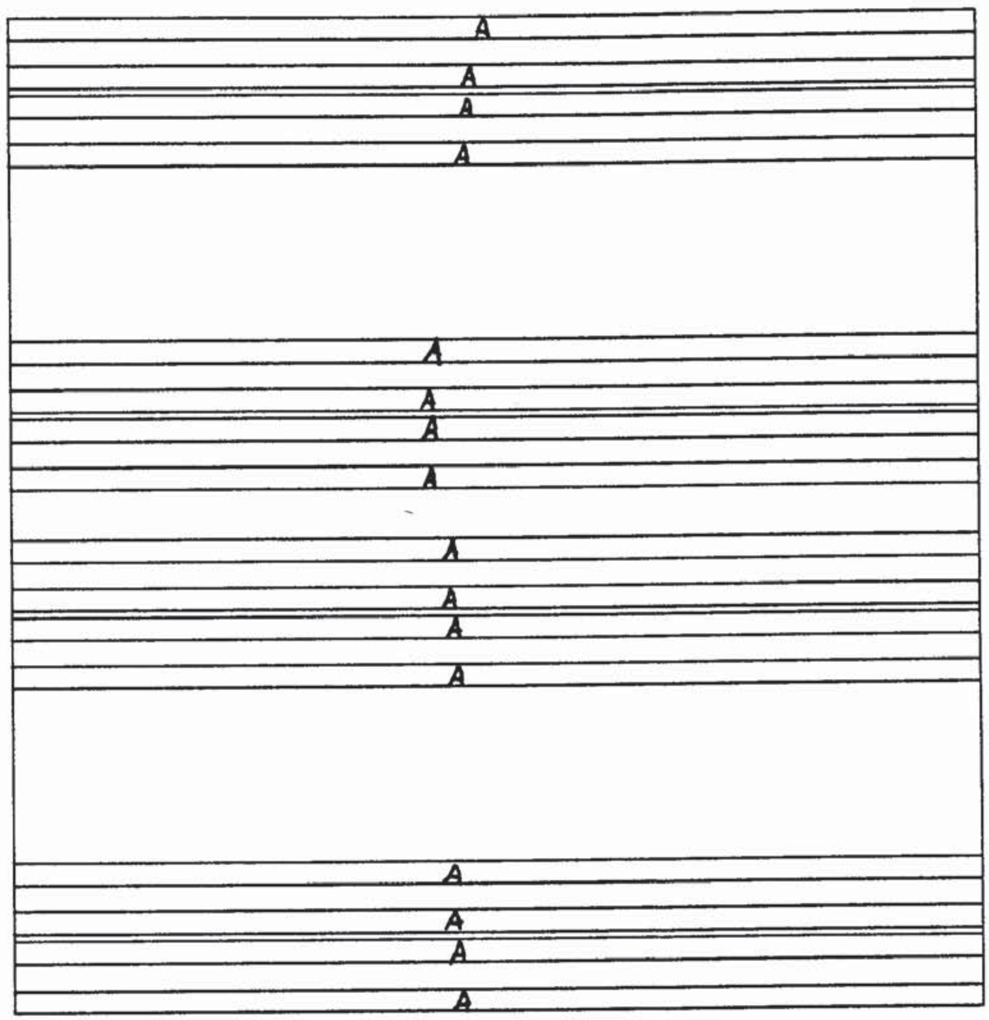

Figure 2.2.15: $A = R_1 \cap \mathcal{A}R_1 \cap \mathcal{A}^2R_1$

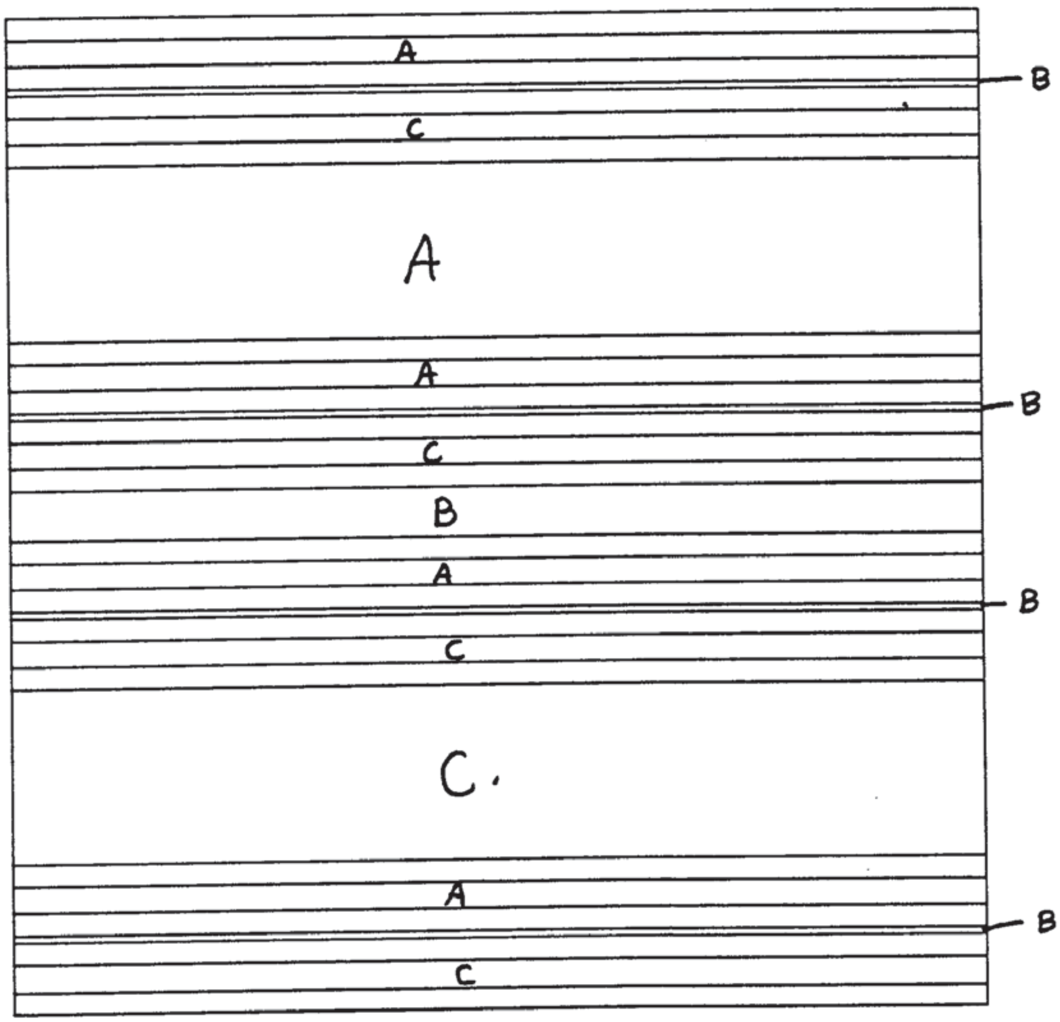

Figure 2.2.16: The Labeling of R_J in Example 2.2.10

$\mathcal{S}$={the two fixed points $(\frac{1}{5},\frac{2}{5})$, $(\frac{4}{5},\frac{3}{5})$} while $\mathcal{U}$={full four shift}. The Markov matrix, with rows 1, 2, and 3 corresponding to A, B, and C is

$$M=\begin{bmatrix} 4 & 0 & 0 & 2 & 0 & 1 \\ 0 & 4 & 0 & 2 & 0 & 1 \\ 0 & 0 & 4 & 2 & 0 & 1 \\ 1 & 0 & 1 & 5 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \end{bmatrix}.$$

EXAMPLE 2.2.17: See Figure 2.2.18. $\mathcal{A}=\begin{bmatrix}5&3\\3&2\end{bmatrix}=\begin{bmatrix}2&1\\1&1\end{bmatrix}^2$. This partition has $\mathcal{S}$={the two fixed points} while $\mathcal{U}$={full four shift}. The only difference between this partition and the previous one is that we have labeled the components of R_1 differently. We label {all uncovered components in $R_1-\mathcal{A}R_1$ A_1}, {all uncovered components in $R_1-\mathcal{A}R_1-\mathcal{A}^2R_1$ with A_2},..., {all uncovered components in $R_1-\mathcal{A}R_1-\mathcal{A}^2R_1-\ldots-\mathcal{A}^{n-1}R_1$ with A_{n-1}}, and {all uncovered components in $R_1-\mathcal{A}R_1-\mathcal{A}^2R_1-\ldots-\mathcal{A}^{mn+k}R_1$ with A_k for $m\in\mathbb{Z}_+$ and $0\leq k\leq n-1$}. The Markov matrix for $n=3$ is

$$M=\begin{bmatrix} 0 & 4 & 0 & 2 & 0 & 1 \\ 0 & 0 & 4 & 2 & 0 & 1 \\ 4 & 0 & 0 & 2 & 0 & 1 \\ 2 & 0 & 0 & 5 & 2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \end{bmatrix}$$

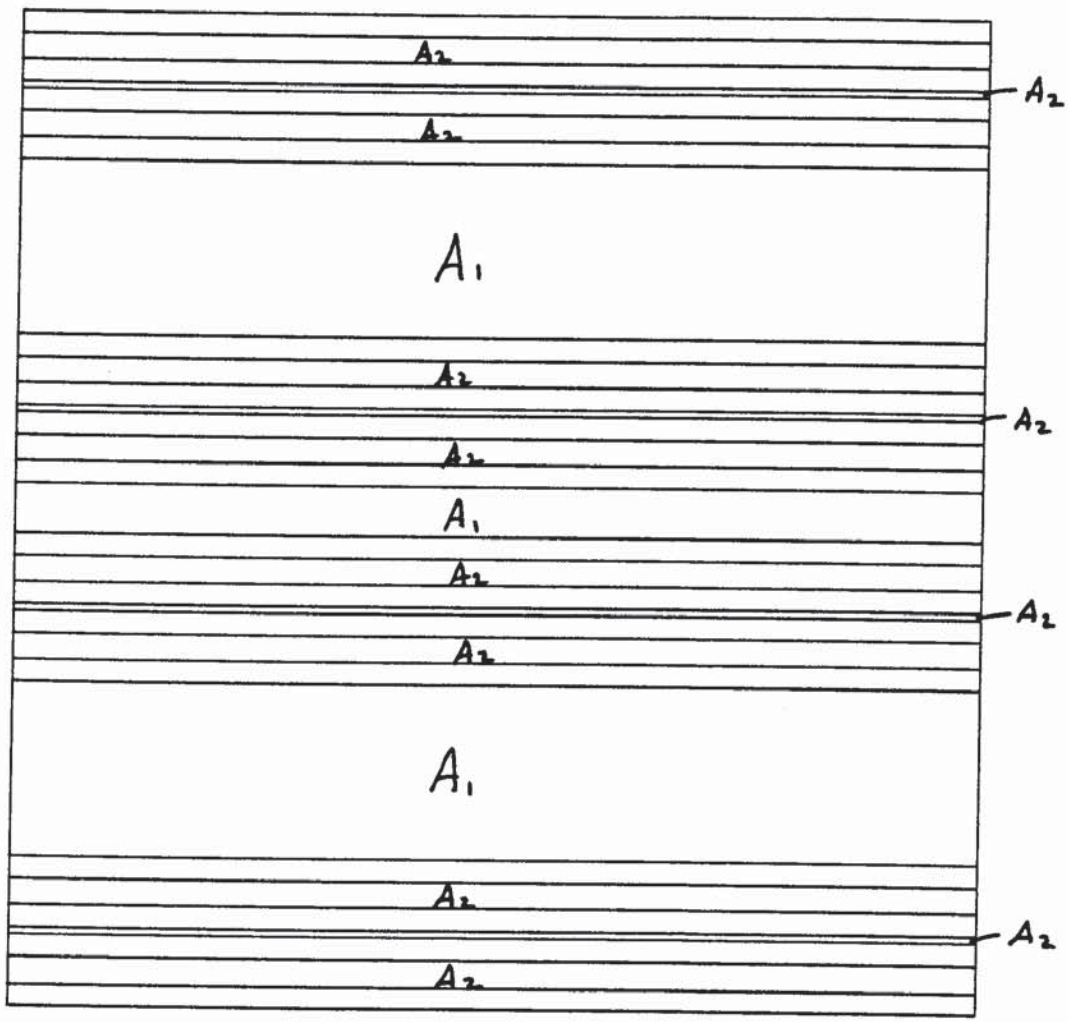

Figure 2.2.18: The Labeling of R_I for $n=2$

EXAMPLE 2.2.19: $\mathcal{A}=\left[\begin{smallmatrix}5&3\\3&2\end{smallmatrix}\right]=\left[\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right]^2$. Let F be any $p\times p$ matrix with row sum 4. We can label the components of R_1 such that the Markov matrix is

$$M=\begin{bmatrix} f_{1,1} & \cdots & f_{1,p} & 2 & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ f_{p,1} & \cdots & f_{p,p} & 2 & 0 & 1 \\ & & & 5 & 2 & 0 \\ & S & & 0 & 0 & 0 \\ & & & 2 & 1 & 0 \end{bmatrix}$$

where $\sum_{j=1}^{p} s_{i,j}=m_{i+1,j}$ where $m_{i+1,j}$ is an entry in M from Example 2.2.9 and $2\leq i\leq 4$. If $p\leq 3$, then we label components in a manner similar to the above examples. We must always be sure, however, that we label in such a way that components with the same label are forward equivalent. If $p>3$, then label all uncovered components of $R_1-\mathcal{A}R_1-\mathcal{A}^2R_1-\ldots-\mathcal{A}^kR_1$ the same, say B, until the number of uncovered components of $R_1-\mathcal{A}R_1-\mathcal{A}^2R_1-\ldots-\mathcal{A}^kR_1$ exceeds p. Figure 2.2.20 is a labeling which gives

$$F=\begin{bmatrix} 0 & 1 & 3 \\ 1 & 1 & 2 \\ 0 & 0 & 4 \end{bmatrix}$$

EXAMPLE 2.2.20: See Figures 2.2.21-2.2.23. $\mathcal{A}=\left[\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right]$. This example shows that rectangles with a non-finite number of components may still have a simple core. The region labeled A in Figure 2.2.21 is broken up as shown in Figure 2.2.23. In this case, the core is simply the origin. The Markov matrix is

$$M=\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

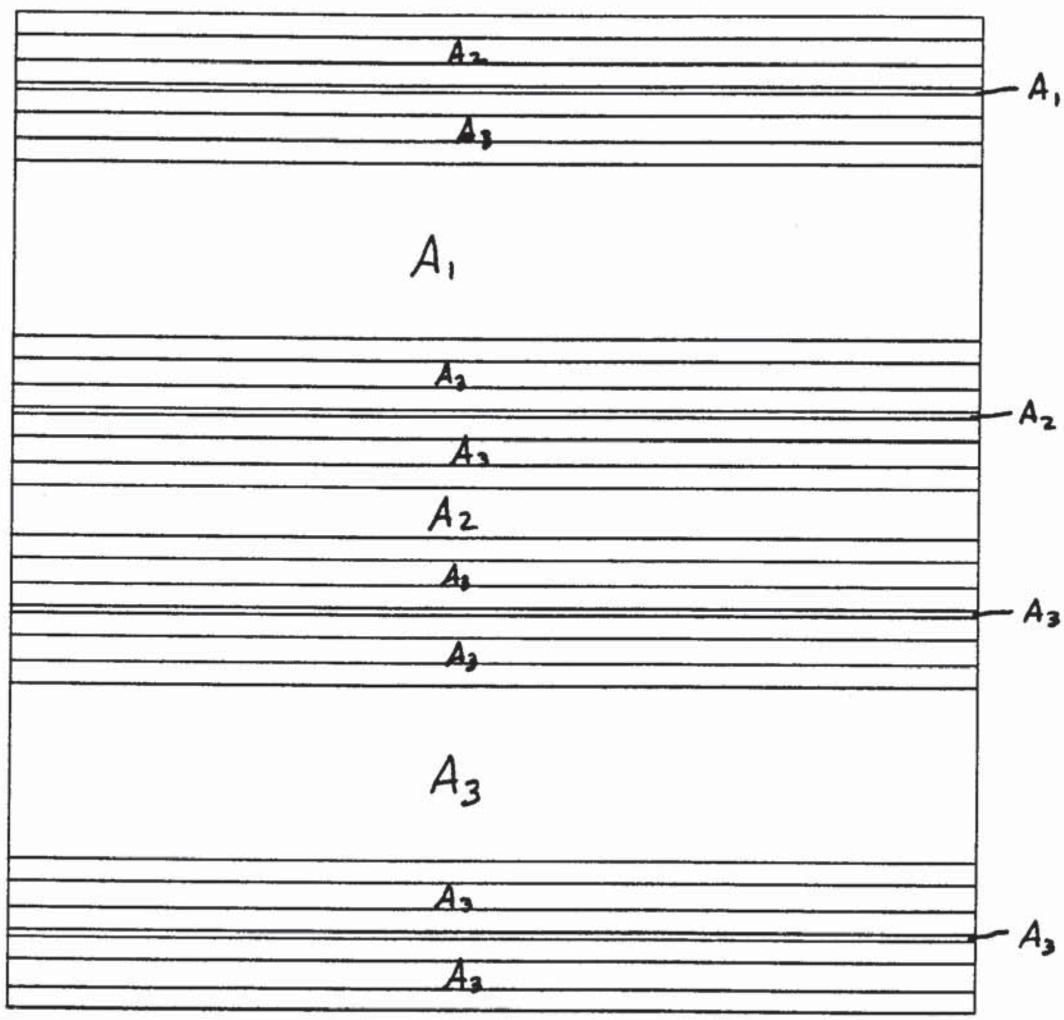

Figure 2.2.21: The Labeling of R_I in Example 2.2.19

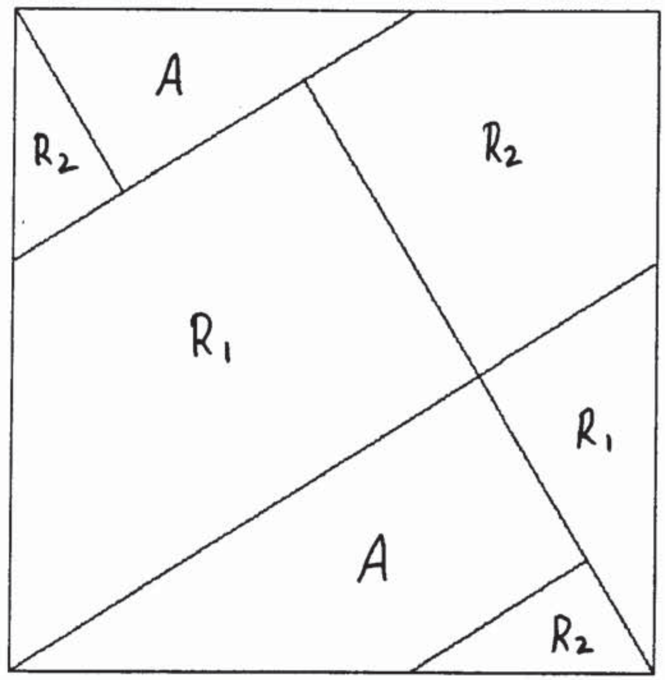

Figure 2.2.22: The Partition 𝒫 of Example 2.2.29

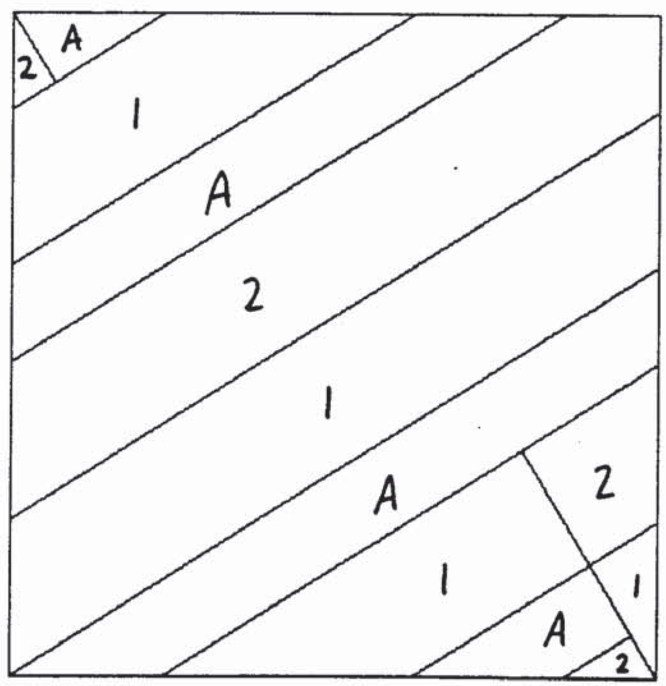

Figure 2.2.23: $\mathcal{A}(\mathcal{P})$

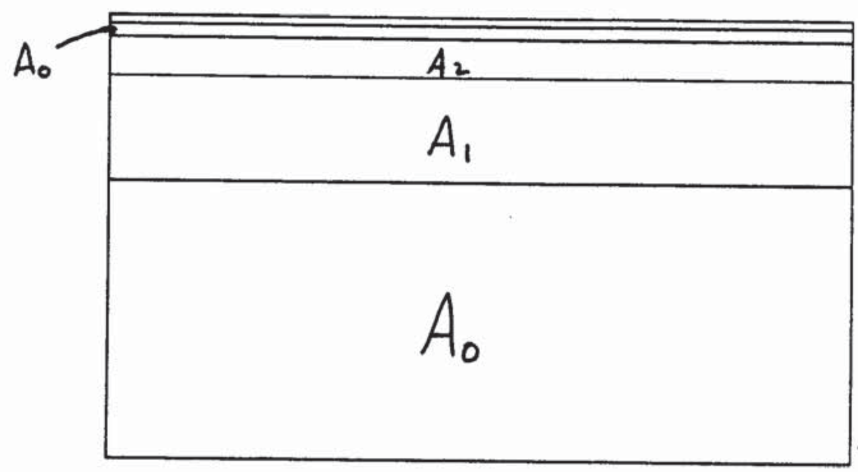

Figure 2.2.24: The Labeling of A

CHAPTER 3 - THE EIGENVALUES OF THE MARKOV MATRIX

SECTION 3.1: PRELIMINARY LEMMAS

The goal of the next two chapters is to say what we can about the Markov matrix M for FCC partitions. The theorem below tells us about the eigenvalues of M. We begin with a series of technical lemmas. First, we need the following lemma due to Matthew Stafford.

LEMMA 3.1.1: [12] Let $h:B\to B$, B finite. Then every eigenvalue of

$$h_*:\widetilde{H}_0(B)\to\widetilde{H}_0(B)$$

is either 0 or a root of unity. The multiplicity of the 0 eigenvalue is equal to the number of pre-periodic points in B. The multiplicity of the eigenvalue 1 is equal to the (# of periodic orbits in B)-1.

Proof: If B is a finite union of periodic orbits, $h|_B$ can be represented by a permutation matrix. (Think of each point in B as a basis element.) Clearly, this permutation is cyclic if and only if B consists of a single periodic orbit.

If B contains pre-periodic points as well, $h|_B$ can be represented by a matrix of the form

$$Q=\begin{bmatrix} P & 0 & \cdots & 0 \\ * & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & 0 \end{bmatrix}$$

where P is a permutation matrix representing the action of h on the periodic points in B. The rows below P represent the action on the pre-periodic points in B. (Here, one must be careful about the ordering of the basis elements: if x and y are pre-periodic and $h(x)=y$, y should be listed before x.) It is clear from the block form of Q that $\chi_Q(t)=\chi_P(t)(-t)^n$, where n is the number of pre-periodic points in B. Note that the multiplicity of the root 1 in $\chi_P(t)$ and thus in $\chi_Q(t)$ is equal to the number of periodic orbits in B.

Let $b_0, b_1, \ldots, b_{p-1}$ be the elements of B. h induces endomorphisms of $H_0(B)\cong\mathbb{Z}^p$ and $\widetilde{H}_0(B)\cong\mathbb{Z}^{p-1}$. With respect to the basis $b_0, b_1, \ldots, b_{p-1}$, the former is represented by the matrix Q described above. Let Q' represent the same transformation in the basis $\sum_{i=0}^{p-1} b_i$, $b_1-b_0, b_2-b_0, \ldots, b_{p-1}-b_0$. Q' is of the form

$$Q'=\begin{bmatrix} 1 & * \\ 0 & F \end{bmatrix}.$$

To prove this, it suffices to show that $V=\mathrm{span}\{b_1-b_0, b_2-b_0, \ldots, b_{p-1}-b_0\}$ is an invariant subspace under h_*. $h_*(b_i-b_0)=h(b_i)-h(b_0)$. If $h(b_i)=h(b_0)$, this difference is $0\in V$. Otherwise, $h_*(b_i-b_0)$ is in the set $(b_i-b_j|\ i\neq j\}$, which is clearly contained in V. Thus h_* maps the basis elements of V into V; it follows that $h_*(V)\subset V$.

The reason for this change of basis can now be make clear: $b_1-b_0, b_2-b_0, \ldots, b_{p-1}-b_0$ form a basis of $\widetilde{H}_0(B)$ and $h_*: \widetilde{H}_0(B) \rightarrow \widetilde{H}_0(B)$ is represented by the submatrix F (of Q′ above) in this basis. Further, $\chi_F(t)=\chi_{Q'}(t)/(1-t)=\chi_Q(t)/(1-t)$. Thus every eigenvalue of F must be 0 or a root of unity. The multiplicity of the eigenvalue 0 for the $p\times p$ matrix Q is not decreased when passing to the $(p-1)\times(p-1)$ matrix F. But the multiplicity of the eigenvalue 1 is one less for F than for Q. This verifies the claim about the multiplicities of the eigenvalues 0 and 1 and completes the proof of the lemma. □

Let $X=\mathbb{T}^2-\partial_u^\circ\mathcal{P}$ and $A=X\cap\partial_s\mathcal{P}=\partial_s\mathcal{P}-\partial_u^\circ\mathcal{P}$. By $\partial_u^\circ\mathcal{P}$ we mean consider $\partial_u\mathcal{P}$ as a finite number of line segments and $\partial_u^\circ\mathcal{P}$ is then the union of the interiors of those line segments. If $\partial_u\mathcal{P}$ is connected, then $\partial_u^\circ\mathcal{P}$ is just $\partial_u\mathcal{P}$ without the endpoints.

DEFINITION 3.1.2: We call a point where $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ intersect a crossing if the line segments cross each other completely. If this is not true, we call the intersection an endpoint.

LEMMA 3.1.3: Let $\mathcal{P}$ be a Markov partition with r rectangles for a hyperbolic toral automorphism $\mathcal{A}$. Let X and A be as defined above. Suppose that $\partial_u\mathcal{P}$ has n_u connected components and $\partial_s\mathcal{P}$ has n_s connected components. Then A has $r-n_u$ connected components.

Proof: We begin by proving the lemma under the assumption that both $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ are connected. In this case, $n_u=n_s=1$ and we are trying to show that A has $r-1$ connected components. The best way to see this is to visualize yourself walking the length of $\partial_s\mathcal{P}$ looking say to the left, counting rectangles once you have crossed them. You will pass each rectangle once. Each rectangle will be counted at a point where $\partial_u\mathcal{P}$ crosses $\partial_s\mathcal{P}$ completely (a crossing) except two; one will be counted at a point where $\partial_u\mathcal{P}$ has an endpoint in $\partial_s\mathcal{P}$ and the last one will see an endpoint of $\partial_s\mathcal{P}$ in $\partial_u\mathcal{P}$. Hence r={number of crossings}+2 and {number of crossings}=$r-2$. Also, the number of components of A is {number of crossings}+1. Therefore, the number of components in A is $r-1$.

Now suppose that $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ are as in the statement of the lemma. Again, walk along each piece of $\partial_s\mathcal{P}$ counting rectangles as you cross them. Be careful to look the same way as you cross each piece. All will again be counted at a crossing except n_u+n_s of them: those will be counted at endpoints. Therefore, r={number of crossings}+n_u+n_s and {number of crossings}=$r-n_u-n_s$. The number of components of A is equal to {number of crossings}+n_s, hence A has $r-n_u$ components and the lemma is proven. □

LEMMA 3.1.4: Let B={m points in $\mathbf{T}^2$, $b_1, b_2, \ldots, b_m$ for $m\neq 0$}. Then

$$\widetilde{H}_1(\mathbf{T}^2-B)=\mathbf{Z}^{m+1}.$$

Proof: Consider the polygonal representation of $\mathbf{T}^2-B$ shown in Figure 3.1.5. Let δ_i={a small disc containing b_i in its interior} and let σ={the boundary square in the polygonal representation of $\mathbf{T}^2-B$}.

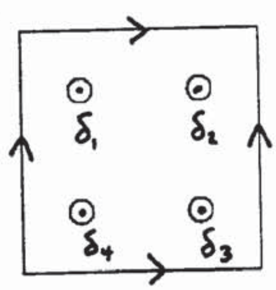

Figure 3.1.5: $\mathbf{T}^2-B$

Let $X_1=\{\mathbf{T}^2-\sigma\}$ and let $X_2=\{\mathbf{T}^2-\bigcup_{i=1}^{m}\delta_i\}$. $X_1\cup X_2=\mathbf{T}^2-B$ so that we can use the Mayer-Vietoris sequence from algebraic topology. Let $X=X_1\cap X_2$. X is homologous to an m-punctured disc and hence to $\bigvee_m S^1$ so $\widetilde{H}_2(X)=0$, $\widetilde{H}_1(X)=\mathbf{Z}^m$, and $\widetilde{H}_0(X)=0$. X_1 is also homologous to $\bigvee_m S^1$ and therefore has the same homology as X. X_2 is homologous to $\bigvee_{m+1} S^1$ hence $\widetilde{H}_1(X_2)=\mathbf{Z}^{m+1}$. Also $\widetilde{H}_2(\mathbf{T}^2-B)=0$. We therefore see the following Mayer-Vietoris sequence:

$$\begin{array}{ccccccccc}\widetilde{H}_2(\mathbf{T}^2-B) & \rightarrow & \widetilde{H}_1(X) & \rightarrow & \widetilde{H}_1(X_1)\oplus\widetilde{H}_1(X_2) & \rightarrow & \widetilde{H}_1(\mathbf{T}^2-B) & \rightarrow & \widetilde{H}_0(X) \\ 0 & \rightarrow & \mathbf{Z}^m & \rightarrow & \mathbf{Z}^m \oplus \mathbf{Z}^{m+1} & \rightarrow & \mathbf{Z}^{m+1} & \rightarrow & 0\end{array}$$

We conclude then that $\widetilde{H}_1(\mathbf{T}^2-B)=\mathbf{Z}^{m+1}$ as desired. □

LEMMA 3.1.6: Let F be the n×n matrix:

$$F_n=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & & \cdots & & -1 \end{bmatrix}$$

The characteristic polynomial of F is $(-1)^n(\lambda^n+\lambda^{n-1}+\ldots+\lambda+1)$.

Proof: By induction. The characteristic polynomial of

$$F_2=\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$$

is $\chi_{F_2}=\lambda^2+\lambda+1$ which is what we wanted. So suppose the lemma is true for n−1. We will show it is true for n. If F is the n×n matrix above, then

$$\chi_{F_n}=\begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\lambda & 1 \\ -1 & \cdots & \cdots & -1 & -\lambda-1 \end{vmatrix}$$

$$=-\lambda\cdot\chi_{F_{n-1}}+(-1)^{n+1}(-1)(1)=(-1)^n((\lambda^n+\lambda^{n-1}+\ldots+\lambda)+1)$$

which is what we wanted. □

LEMMA 3.1.7: Let F_n be the n×n matrix

$$F_n=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \vdots & & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

Then the characteristic polynomial of F_n is $(-1)^n(\lambda^n-1)$.

Proof: By induction. For $F_2=\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, $\chi_{F_2}=\lambda^2-1$. So assume true for n−1. Then

$$\chi_{F_n}=\begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \vdots & \ddots & -\lambda & 1 \\ 1 & 0 & \cdots & 0 & -\lambda \end{vmatrix}$$

$$=-\lambda\begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -\lambda & 1 \\ 0 & \cdots & \cdots & 0 & -\lambda \end{vmatrix}+(-1)^{n+1}$$

$$=(-1)^n(\lambda^n-1). \qquad \square$$

SECTION 3.2: THE EIGENVALUE THEOREM

We now state and prove the main theorem about the eigenvalues of the Markov matrix.

THEOREM 3.2.1: Let $\mathcal{P}$ be a Markov partition for $\mathcal{A}$ where $\mathcal{A}$ is a hyperbolic automorphism of $\mathbb{T}^2$ and let $\mathcal{P}$ satisfy FCC. Let M be the Markov matrix for $\mathcal{P}$. If $\mathrm{Tr}(\mathcal{A})>0$, then the eigenvalues of M are λ_u, λ_s, together with 0's and roots of unity. If $\mathrm{Tr}(\mathcal{A})<0$, then the eigenvalues of M are $-\lambda_u$, $-\lambda_s$, together with 0's and roots of unity.

Proof: First, we will give the proof under the further assumption that both $\partial_u\mathcal{P}$ and $\partial_s\mathcal{P}$ are connected.

Suppose $\mathcal{P}$ has r rectangles, and that $\mathrm{Tr}(\mathcal{A})>0$. Let $X=\mathbb{T}^2-\partial_u^\circ\mathcal{P}$, and $A=X\cap\partial_s\mathcal{P}$. Let $\overline{\mathcal{A}}$ be the map induced by $\mathcal{A}$ on X. By Lemma 3.1.3 A has $r-1$ components and therefore is homologous to $r-1$ points. X is homologous to $\mathbb{T}^2-\{\text{one point}\}$. We know by Lemma 3.1.4 that $\widetilde{H}_1(A)=0$, $\widetilde{H}_1(X)=\mathbb{Z}^2$, $\widetilde{H}_0(A)=\mathbb{Z}^{r-2}$, and $\widetilde{H}_0(X)=0$. The exact relative homology sequence for a pair (X,A) then gives us the following commutative diagram:

$$
\begin{array}{ccccccccc}
\widetilde{H}_1(A) & \rightarrow & \widetilde{H}_1(X) & \xrightarrow{\phi} & \widetilde{H}_1(X,A) & \rightarrow & \widetilde{H}_0(A) & \rightarrow & \widetilde{H}_0(X) \\
0 & \rightarrow & \mathbb{Z}^2 & \rightarrow & \mathbb{Z}^r & \rightarrow & \mathbb{Z}^{r-2} & \rightarrow & 0 \\
 & & \downarrow \overline{\mathcal{A}}_* = \mathcal{A} & & \downarrow M & & \downarrow F & & \\
0 & \rightarrow & \mathbb{Z}^2 & \xrightarrow{\phi} & \mathbb{Z}^r & \rightarrow & \mathbb{Z}^{r-2} & \rightarrow & 0
\end{array}
$$

Because $\mathbb{Z}^m$ is free abelian for $m \geq 0$, the top line of the above sequence is split exact and hence $\widetilde{H}_1(X,A) = \mathbb{Z}^r$. $\overline{\mathcal{A}}_*$ is the map induced on $\widetilde{H}_1(X)$ by $\overline{\mathcal{A}}$ which is in fact $\mathcal{A}$ itself. One set of generators for $\widetilde{H}_1(X,A)$ is a line segment across each rectangle in the unstable direction so the induced map on homology is just M. F is from Lemma 3.1.1. Because the top line of the above sequence is split exact, with possibly a different choice of basis, M can be written in the form:

$$
\begin{bmatrix} \mathcal{A} & 0 \\ * & F \end{bmatrix}.
$$

Thus the eigenvalues of M are exactly the eigenvalues of $\mathcal{A}$ (λ_u and λ_s) together with the eigenvalues of F (roots of unity and zeros) when $\mathrm{Tr}(\mathcal{A})>0$. When $\mathrm{Tr}(\mathcal{A})<0$, the induced map on $\widetilde{H}_1(X,A)$ is $-M$ (because $\lambda_u<0$, $\overline{\mathcal{A}}_*$ reverses the orientation of the generators in $\widetilde{H}_1(X,A)$) so in the diagram replace M with $-M$ and the theorem follows similarly, thus concluding the proof for our case, namely $\partial_u \mathcal{P}$ and $\partial_s \mathcal{P}$ being connected.

We use the above idea and the following lemmas to complete the proof.

LEMMA 3.2.2: Let $B=\Big\{\{\mathcal{A}^i p_1\}_{i=0}^{n_1-1}, \{\mathcal{A}^i p_2\}_{i=0}^{n_2-1}, \ldots, \{\mathcal{A}^i p_m\}_{i=0}^{n_m-1}$ where $p_j \in \mathbf{T}^2$ is an $\mathcal{A}$-periodic point of period $n_j\Big\}$ be a finite set comprised of m periodic orbits of $\mathcal{A}$. Let $\overline{\mathcal{A}}$ be the map induced by $\mathcal{A}$ on $\mathbf{T}^2-B$. Let $\overline{\mathcal{A}}_*$ be the map induced on $\widetilde{H}_1(\mathbf{T}^2-B)$ by $\overline{\mathcal{A}}$. Suppose that λ_u and λ_s have the same sign. Then the eigenvalues of $\overline{\mathcal{A}}_*$ are {the eigenvalues of $\mathcal{A}$} plus {all the n_jth roots of unity except 1 for $0 \le j \le m$ (counting multiplicity) and the multiplicity of the eigenvalue 1 is $m-1$}. If λ_u and λ_s have different signs, then we see {minus all the n_jth roots of unity except -1 for $0 \le j \le m$ (counting multiplicity) and the multiplicity of the eigenvalue -1 is $m-1$.}

Proof: Suppose that λ_u and λ_s have the same sign. We first consider a single orbit of period n with $B=\{\mathcal{A}^i p\}_{i=0}^{n-1}$. By Lemma 3.1.4, we see that $\widetilde{H}_1(\mathbf{T}^2-B)=\mathbf{Z}^{n+1}$. We choose our generators for $\widetilde{H}_1(\mathbf{T}^2-B)$ as follows: α and β, the generators of $\widetilde{H}_1(\mathbf{T}^2)$, and γ_i for $0 \le i \le n-2$ where γ_i is a small circle around $\mathcal{A}^i p$ oriented clockwise; see Figure 3.2.3. This is clearly a set of generators as none of them are homologous. Notice that if γ_{n-1} is a circle around $\mathcal{A}^{n-1}p$, then $\gamma_{n-1}=\alpha+\beta-\alpha-\beta-\gamma_0-\gamma_1-\ldots-\gamma_{n-2}$.

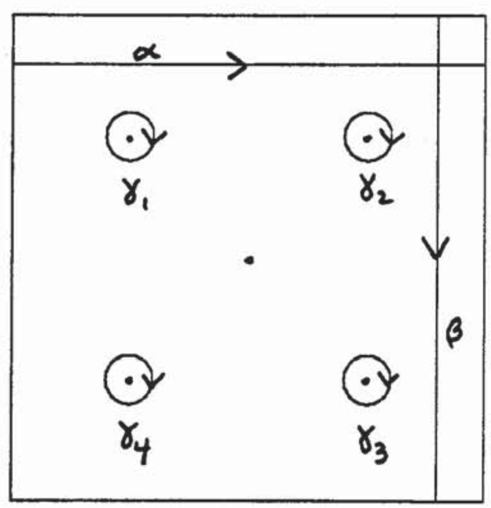

Figure 3.2.3: Generators For $\widetilde{H}_1(\mathbf{T}^2-B)$

The action of $\overline{\mathcal{A}}_*$ on α and β is the same as the action of $\mathcal{A}_*$, that is, $\mathcal{A}$ itself with perhaps a few of the γ_i also. $\overline{\mathcal{A}}_*\gamma_i=\gamma_{i+1}$ for $1\leq i\leq n-2$ which gives $\overline{\mathcal{A}}_*$ the following form:

$$\overline{\mathcal{A}}_*=\begin{bmatrix} \mathcal{A} & 0 & \cdots & \cdots & 0 \\ * & 0 & 1 & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \cdots & 0 & 1 \\ * & -1 & \cdots & \cdots & -1 \end{bmatrix}$$

whose eigenvalues are as desired by Lemma 3.1.6.

Suppose now that B is as given in the statement of the lemma. We choose generators in the same way leaving out only one point in B, WLOG assume this is p_1. We refer to the circle around $\mathcal{A}^i p_j$ as $\gamma_{i,j}$. Clearly $\overline{\mathcal{A}}_*\gamma_{i,j}=\gamma_{i+1,j}$ for $0\leq i\leq n_j-1$ $((n_j-1)+1=0)$ and $2\leq j\leq m$, while the action on the $\gamma_{i,1}$ is as described above. We then see that $\overline{\mathcal{A}}_*$ has the following form:

$$\overline{\mathcal{A}}_*=\begin{bmatrix} \mathcal{A} & 0 & \cdots & 0 \\ * & F_1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ * & 0 & & F_m \end{bmatrix} \text{where}$$

$$F_1=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & & \cdots & & -1 \end{bmatrix} \text{and}$$

$$F_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \vdots & & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

for $2 \leq i \leq m$ which yields the desired eigenvalues by Lemmas 3.1.6 and 3.1.7.

If λ_u and λ_s have different signs, then the orientations of the generators γ_i are reversed so that we see $-F_i$ in each case and hence have minus the eigenvalues. □ (Lemma 3.2.2)

We recall Definition 2.1.7 and expand upon that notion. For $x \in \mathcal{C}$, if $W^u(x) \cap \partial_u \mathcal{P}$ has more than one component, these components differ in nature. One, the component containing x, gets mapped into itself under $\mathcal{A}^{-n}$ for some n while the others do not. We therefore make the following distinction.

DEFINITION 3.2.4: Let I be a component of $W^u(x) \cap \partial_u \mathcal{P}$ (or $W^s(x) \cap \partial_s \mathcal{P}$) for $x \in \mathcal{C}$. We say that I is <u>essential</u> if $x \in I$. Otherwise, we say that I is <u>non-essential</u>.

If $\partial \mathcal{P}$ has only essential components, then $\mathbf{T}^2 - \partial_u^{\circ} \mathcal{P}$ is homologous to $\mathbf{T}^2 - B$ where B is a finite number of periodic orbits and the previous lemma applies. If $\partial \mathcal{P}$ has some non-essential components, we must explore what this will do to $\overline{\mathcal{A}}_*$.

LEMMA 3.2.5: Each non-essential component in the unstable boundary of $\mathcal{P}$ contributes a zero to the eigenvalues of $\overline{\mathcal{A}}_*$.

Proof: We begin by filling in any gaps in $\partial\mathcal{P}$ and proceed by induction. Each component in $\partial_u\mathcal{P}$ is now connected and Lemma 3.2.2 applies. The image of at least one of the former gaps in $\partial_u\mathcal{P}$ does not intersect $\dot{\partial}_u\mathcal{P}$. We can therefore remove a piece of $\partial_u\mathcal{P}$ creating this gap, call it G. We now need a new generator in $\widetilde{H}_1(\mathbf{T}^2-\dot{\partial}_u\mathcal{P})$ and for this generator we choose a loop around the new non-essential component we created, call it γ; see Figure 3.2.6. Now, because $\mathcal{A}\partial_u\mathcal{P}\supset\partial_u\mathcal{P}$, the image of the points in G ($G\subset\mathbf{T}^2-\dot{\partial}_u\mathcal{P}$) must also be in $\mathbf{T}^2-\dot{\partial}_u\mathcal{P}$ while $\mathcal{A}(\mathbf{T}^2-\dot{\partial}_u\mathcal{P})\cap G=\emptyset$ (the preimage of G is contained in $\dot{\partial}_u\mathcal{P}$) hence γ is now contractable to a point and hence equal to zero in $\widetilde{H}_1(\mathbf{T}^2-\dot{\partial}_u\mathcal{P})$. γ has then been mapped to zero in $\widetilde{H}_1(\mathbf{T}^2-\dot{\partial}_u\mathcal{P})$ hence $\overline{\mathcal{A}}_*$ has a new column of zeros and hence a new zero eigenvalue. We now create all the former gaps one by one in a like manner and see that each one gives $\overline{\mathcal{A}}_*$ another zero eigenvalue.

□ (Lemma 3.2.5)

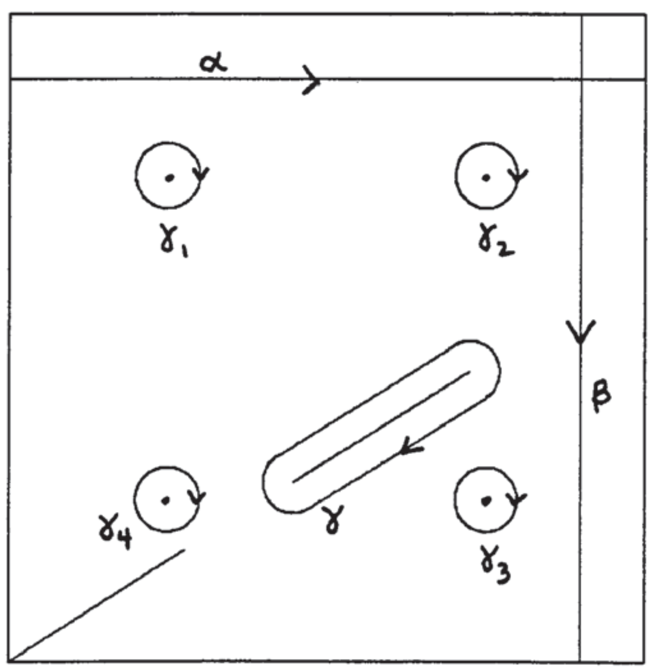

Figure 3.2.6: The New Generator

We now complete the proof of the theorem.

Suppose that $\mathcal{P}$ satisfies FCC and $\mathrm{Tr}(\mathcal{A})>0$. This means that $\mathcal{U}$ and $\mathcal{S}$ are a finite number of periodic orbits and suppose $\partial_u\mathcal{P}$ has n_u components. By Lemma 3.1.4, $\widetilde{H}_1(X)=\mathbf{Z}^{n_u+1}$. $\widetilde{H}_0(A)=\mathbf{Z}^{r-n_u-1}$ by Lemma 3.1.3. By Lemmas 3.2.2 and 3.2.5, $\overline{\mathcal{A}}_*$ has eigenvalues λ_u, λ_s, zeros and roots of unity. We see the following commutative diagram.

$$
\begin{array}{ccccccccc}
\widetilde{H}_1(A) & \rightarrow & \widetilde{H}_1(X) & \xrightarrow{\phi} & \widetilde{H}_1(X,A) & \rightarrow & \widetilde{H}_0(A) & \rightarrow & \widetilde{H}_0(X) \\
0 & \rightarrow & \mathbf{Z}^{n_u+1} & \rightarrow & \mathbf{Z}^r & \rightarrow & \mathbf{Z}^{r-n_u-1} & \rightarrow & 0 \\
 & & \downarrow \overline{\mathcal{A}}_* & & \downarrow M & & \downarrow F & & \\
0 & \rightarrow & \mathbf{Z}^{n_u+1} & \rightarrow & \mathbf{Z}^r & \rightarrow & \mathbf{Z}^{r-n_u-1} & \rightarrow & 0
\end{array}
$$

Therefore, the eigenvalues of M are exactly the eigenvalues of $\mathcal{A}$ together with the eigenvalues of F and the proof is complete for $\mathrm{Tr}(\mathcal{A})>0$ with the proof for $\mathrm{Tr}(\mathcal{A})<0$ following in a similar manner. □

The proof of this theorem gives us the following two results:

COROLLARY 3.2.7: Let $\mathcal{P}$ be a Markov partition for $\mathcal{A}$ with $r>2$ rectangles, and suppose that $\mathcal{P}$ satisfies FCC. Then the Markov matrix M is similar over $\mathbf{Z}$ to an $r\times r$ matrix of the following form:

$$\begin{bmatrix} \mathcal{A} & 0 & \cdots & 0 \\ * & A_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & A_k \end{bmatrix}$$

where the eigenvalues of the non-zero A_i are either $\pm$\{all the pth roots of unity for some p\} or $\pm$\{all the pth roots of unity except 1 for some p\}.

Proof: By noting that F in the above proof has as eigenvalues only roots of unity and zeros, the corollary follows directly from the proof of Theorem 3.2.1 and the block triangular form from linear algebra [8]. □

Roy Adler, in an as yet unpublished work, proved that every toral automorphism $\mathcal{A}$ has a Markov partition with two rectangles. He in fact proved the existence of a Markov partition for which the Markov matrix is $\mathcal{A}$ itself ($\mathcal{A} \geq 0$). We use the first result to give us the following corollary which is Corollary 3.2.7. in the 2×2 case:

COROLLARY 3.2.8: Let $\mathcal{P}$ be a Markov partition for $\mathcal{A}$ with two rectangles and Markov matrix M. Then if $\mathrm{Tr}(\mathcal{A}) > 0$, there exists $\phi \in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi = M$. Similarly, if $\mathrm{Tr}(\mathcal{A}) < 0$, there exists $\phi \in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi = -M$.

Proof: Letting X and A be as in Theorem 3.2.1, with $r=2$ and $\mathrm{Tr}(\mathcal{A}) > 0$, we see the following diagram:

$$
\begin{array}{ccccccc}
\widetilde{H}_1(A) & \rightarrow & \widetilde{H}_1(X) & \xrightarrow{\phi} & \widetilde{H}_1(X,A) & \rightarrow & \widetilde{H}_0(A) \\
0 & \rightarrow & \mathbb{Z}^2 & \rightarrow & \mathbb{Z}^2 & \rightarrow & 0 \\
 & & \downarrow \overline{\mathcal{A}}_* = \mathcal{A} & & \downarrow M & & \\
0 & \rightarrow & \mathbb{Z}^2 & \xrightarrow{\phi} & \mathbb{Z}^2 & \rightarrow & 0
\end{array}
$$

ϕ is an isomorphism hence in GL(2,$\mathbb{Z}$) and the corollary is proven for $\mathrm{Tr}(\mathcal{A})>0$ with the proof for $\mathrm{Tr}(\mathcal{A})<0$ being entirely similar. □

SECTION 3.3: EIGENVALUES AND THE CORE

We begin with a few examples of the results presented in the last section.

EXAMPLE 3.3.1: Recall Example 2.2.1. The eigenvalues of M are:

$$\{\tfrac{3+\sqrt{5}}{2}, \tfrac{3-\sqrt{5}}{2}, \tfrac{-1+i\sqrt{3}}{2}, \tfrac{-1-i\sqrt{3}}{2}, -1, 0, 0, 0, 0\}.$$

We see λ_u, λ_s, zeros and second and third roots of unity as expected.

EXAMPLE 3.3.2: Recall Example 2.2.2. The eigenvalues of M are

$$\{\tfrac{3+\sqrt{5}}{2}, \tfrac{3-\sqrt{5}}{2}, 1, 0\}.$$

EXAMPLE 3.3.3: Recall Example 2.2.9. The eigenvalues of M are

$$\{\frac{7+3\sqrt{5}}{2}, \frac{7-3\sqrt{5}}{2}, 1, 1\}.$$

All of the above examples fit the hypotheses of Theorem 3.2.1. Examples 2.2.10, 2.2.17, 2.2.19, and 2.2.21 do not fit the hypotheses.

EXAMPLE 3.3.4: Recall Example 2.2.21. The eigenvalues of M are

$$\{\frac{3+\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}, \frac{-1+i\sqrt{3}}{2}, \frac{-1-i\sqrt{3}}{2}, 0\}.$$

In this case, even though the core is a single fixed point, the eigenvalues reflect a more complicated structure, possibly a period three orbit.

EXAMPLE 3.3.5: Recall Example 2.2.10. The eigenvalues of M are

$$\{\frac{7+3\sqrt{5}}{2}, \frac{7-3\sqrt{5}}{2}, 4, 4, 1, 1\}.$$

EXAMPLE 3.3.6: Recall Example 2.2.17. The eigenvalues of M are

$$\{\frac{7+3\sqrt{5}}{2}, \frac{7-3\sqrt{5}}{2}, \frac{-4+4i\sqrt{3}}{2}, \frac{-4-4i\sqrt{3}}{2}, 1, 1\}.$$

EXAMPLE 3.3.7: Recall Example 2.2.19. Since F has row sum four, four is an eigenvalue of F. The eigenvalues of M are

$$\{\frac{7+3\sqrt{5}}{2}, \frac{7-3\sqrt{5}}{2}, 1, 1, \{\text{all the eigenvalues of F except the number 4}\}\}.$$

To verify this we see that by conjugating by elementary matrices we can get M to the following form:

$$M = \begin{pmatrix} F' & 0 & \dots & \dots & 0 \\ * & 4 & 2 & 0 & 1 \\ \vdots & 2 & 5 & 2 & 0 \\ \vdots & 1 & 0 & 0 & 0 \\ * & 0 & 2 & 1 & 0 \end{pmatrix}.$$

We can then see that the eigenvalues of M are as claimed.

It is clear from the proof of Theorem 3.2.1 that the composition of the core has a lot to do with the eigenvalues of M. A periodic orbit of period p in $\mathcal{U}$ adds the pth roots of unity to the eigenvalues of M (modulo some ones) if $\lambda_u > 0$ and some of the $2p$th roots of unity if $\lambda_u < 0$ while a periodic orbit of period p in $\mathcal{S}$ adds the pth roots of unity if $\lambda_s > 0$ and all the $2p$th roots of unity if $\lambda_s < 0$. Because of this we, given the trace and the determinant of the matrix (which in turn determine the signs of λ_u and λ_s) and the core of the partition, we can determine which eigenvalues can occur. Unfortunately, we cannot determine the exact contents of the core from the eigenvalues.

EXAMPLE 3.3.8: Let $\mathcal{P}_1$ be a partition with a $\mathcal{C} = \mathcal{U} = \mathcal{S} = \{$one period three orbit$\}$ and a crossing at each point in $\mathcal{C}$ and no other crossings. Assume that λ_u and λ_s are positive. From Section 3.2, we can conclude that the eigenvalues of M are λ_u, λ_s, 1, and each of the third roots of unity except 1 will occur three times (once from $\overline{\mathcal{A}}_*$ and twice from F). Let $\mathcal{P}_2$ be a partition for the same system with $\mathcal{U} = \{$one period three orbit$\}$ and $\mathcal{S} = \{$two different period three orbits not in $\mathcal{U}$} and suppose that there are no crossings. This system will have the same eigenvalues as $\mathcal{P}_1$ (the 1 and the third roots of unity twice from F and the roots of unity except one once from $\overline{\mathcal{A}}_*$). We see two partitions with different cores giving the same eigenvalues in the Markov matrix.

By looking at the eigenvalues of M, it is hopeless to try to determine the composition of $\mathcal{U}$ and $\mathcal{S}$. We cannot even determine exactly what is in the core unless we make further assumptions. For example, if one assumes that we have a crossing at each point in the core, FCC, and that both eigenvalues are positive, then each orbit of period p in $\mathcal{C}$ causes the pth roots of unity to surface three times in the eigenvalues of M (modulo 1's); once in $\overline{\mathcal{A}}_*$ and twice in F. Therefore, given a set of eigenvalues we can figure out exactly what was in the core. We can develop similar algorithms for cases where both eigenvalues are not positive and a constant structure is found at each point in $\mathcal{C}$.

A question which arises at this point is whether or not we can detect the presence of large cores or rectangles with infinitely many components. The answer to this question is no. If we consider Examples 2.2.19 and 3.3.7, the eigenvalues of M are $\{\frac{7+3\sqrt{5}}{2}, \frac{7-3\sqrt{5}}{2}, 1, 1, 1, -1\}$. These suggest that there may be some fixed points (we don't know exactly how many because of crossings and such) and a period two orbit. There is nothing to suggest that there is a large core as there is in Example 2.2.17 (roots of unity times four). Hence, the presence of "strange" eigenvalues does indicate that there is a non-finite core but it is not a necessary condition. Sometimes, we can use the number of rectangles present to detect complicated cores. The following proposition deals with the minimum number of rectangles for a given core.

PROPOSITION 3.3.9: Let $\mathcal{A}$ be a hyperbolic toral automorphism. If a core-connected Markov partition $\mathcal{P}$ for $\mathcal{A}$ has $\mathcal{C}$={a single periodic orbit of period p}, then the minimum number of rectangles for $\mathcal{P}$ is $r=2p$ unless both λ_u and λ_s are both negative in which case the minimum number of rectangles is $r=3p$. If $\mathcal{C}$={periodic orbits of periods p_1, p_2,..., p_k} then the minimum number of rectangles in $\mathcal{P}$ is $r=\sum p_i$.

Proof: We know from the proof of Lemma 3.1.4 that if $\partial_u\mathcal{P}$ has n_u components and $\partial_s\mathcal{P}$ has n_s components then $r=n_u+n_s+$\{number of crossings\}. If we assume that $\mathcal{C}$={a single orbit of period p} and at least one of the eigenvalues of $\mathcal{A}$ is positive, then construct a partition with no crossings. We must have each point in this orbit in both $\mathcal{U}$ and $\mathcal{S}$. However, since one of the eigenvalues of $\mathcal{A}$ is positive, we need only have an endpoint at each point in $\mathcal{C}$. Out of necessity, for each point in $\mathcal{C}$ there will be one component in $\partial_u\mathcal{P}$ and one component in $\partial_s\mathcal{P}$, hence we have $2p$ rectangles. If both of the eigenvalues of $\mathcal{A}$ are negative, then out of necessity we have a crossing at each point in $\mathcal{C}$ hence we have $3p$ rectangles.

If $\mathcal{C}$ is a finite number of periodic orbits, construct a partition with $\mathcal{U}\cap\mathcal{S}=\emptyset$ and no crossings. This partition will have one component of $\partial\mathcal{P}$ at each point in $\mathcal{C}$ and hence the number of rectangles is $\sum p_i$. □

If we now examine Example 3.3.4, we see that the eigenvalues of M suggest a single period three orbit in $\mathcal{C}$. However, our previous proposition asserts that the minimum number of rectangles for a single period three orbit is six. Hence, we must have had a partition which is not core-connected.

CHAPTER 4 - THE TWO RECTANGLE CASE

SECTION 4.1: SOME ALGEBRAIC RESULTS

A natural question to ask at this point is whether or not the converse of either Corollary 3.2.7 or Corollary 3.2.8 is true. In this chapter, we examine the 2×2 case and make a conjecture about the r rectangle case. In order to do so we need the following algebraic results.

PROPOSITION 4.1.1: Let $\mathcal{A}\geq 0$ be an element of GL(2,$\mathbb{Z}$) and suppose $\mathcal{A}\neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathcal{A}\neq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $\mathcal{A}$ has a dominant row and a dominant column; i.e. the first column of $\mathcal{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is dominant if $a\geq b$ and $c\geq d$ and similarly for a row being dominant.

Proof: Suppose $\mathcal{A}$ does not have a dominant column. Then either *i*) $a>b$ and $c<d$ or *ii*) $b>a$ and $d<c$.

Case *i*) $a>b$ and $c<d \Rightarrow a\geq b+1 \Rightarrow ad>bc+c \Rightarrow ad-bc>c$. Therefore, since $\det(\mathcal{A})=\pm 1$ and $c\geq 0$ we have that $c=0 \Rightarrow a=d=1 \Rightarrow b=0$ (since $a>b$) $\Rightarrow \mathcal{A}= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ which is a contradiction.

Case *ii*) $b>a$ and $d<c \Rightarrow b\geq a+1 \Rightarrow bc>ad+d \Rightarrow bc-ad>d$. Therefore, since $\det(\mathcal{A})=\pm 1$ and $d\geq 0$ we have that $d=0 \Rightarrow b=c=1 \Rightarrow a=0$ (since $b>a$) $\Rightarrow \mathcal{A}= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ which is a contradiction.

The proof for dominant rows is entirely similar. □

PROPOSITION 4.1.2: Suppose $\mathcal{A}\in GL(2,\mathbb{Z})$, $\mathcal{A}\geq 0$, $\mathcal{A}\neq \begin{bmatrix}1&0\\0&1\end{bmatrix}$. Then

$$\mathcal{A}=\prod_{i=1}^{n}\begin{bmatrix}x_i & 1\\ 1 & 0\end{bmatrix}$$

where $x_i=0$ or 1 for $1\leq i\leq n$.

Proof: We can see the above by observing the following:

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}x & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}c & d\\ y & z\end{bmatrix}$$

where if $a\geq c$ and $b\geq d$ then let $x=1$, $y=a-c$, and $z=b-d$ thus reducing the entries in the dominant row. Otherwise let $x=0$, $y=a$, and $z=b$. We can then view this factorization as a sequence of 1's and 0's $\{x_i\}_{i=1}^{n}$ and this factorization is unique up to two consecutive 0's ($\begin{bmatrix}0&1\\1&0\end{bmatrix}^2=\begin{bmatrix}1&0\\0&1\end{bmatrix}$). □

DEFINITION 4.1.3: We call $\{x_i\}_{i=1}^{n}$ the <u>defining sequence for $\mathcal{A}$</u>. From here on we will denote $\begin{bmatrix}0&1\\1&0\end{bmatrix}$ as x, $\begin{bmatrix}1&1\\1&0\end{bmatrix}$ as y, and $\begin{bmatrix}0&1\\1&1\end{bmatrix}$ as y'. We can also think of the defining sequence as a sequence of x's and y's.

EXAMPLE 4.1.4: Using the above algorithm, we can see that the matrix $\begin{bmatrix}7&3\\2&1\end{bmatrix}$ has defining sequence $yxyxyyxy$ or 10101101.

DEFINITION 4.1.5: We define a <u>shear matrix</u> to be a matrix with ones on the diagonal and a single one off the diagonal while all other off diagonal entries are 0. Note that if a shear matrix has a one as its i,jth entry then its inverse has a minus one as its i,jth entry.

PROPOSITION 4.1.6: Under the same hypotheses as in Proposition 4.1.0, we see that $\mathcal{A}$ also factors as follows: If $\mathcal{A}\neq x$, then $\mathcal{A}=\prod_{i=1}^{m} M_i$ where M_i is a 2×2 shear matrix for $1\leq i\leq m-1$ and M_m is either a 2×2 shear matrix or y or y'.

Proof: We can see that $xy=y'x$, $yx=\begin{bmatrix}1&1\\0&1\end{bmatrix}$, a shear, and $y'x=\begin{bmatrix}1&0\\1&1\end{bmatrix}$, a shear. Given the defining sequence for $\mathcal{A}$, $\{x_i\}_{i=1}^{n}$, we replace 1's with y's and 0's with x's. Now using the first relation above, put the sequence in the form $s_1xs_2x\ldots s_kxs_{k+1}s_{k+2}\ldots s_m$ where s_i can be either y or y' for $1\leq i\leq m$. (If there are no x's in the sequence, we let $k=0$.) If $m=k$ or $m=k+1$ then we are done. If $m>k+1$ then insert xx between s_{k+2j-1} and s_{k+2j} for $1\leq j\leq\lfloor\frac{m-k}{2}\rfloor$ (where $\lfloor r\rfloor$={the largest integer smaller than r}), replace xs_{k+2j} with sx where s is either y or y' and we are done. □

EXAMPLE 4.1.7: For the matrix $\begin{bmatrix}7&3\\2&1\end{bmatrix}$, we rewrite the defining sequence from Example 4.1.4 as $yxyxyxy'y$ by replacing the sixth and seventh entries, yx, with xy'. We then insert xx between the y and y' yielding $yxyxyxy'xxy=yxyxyxy'xy'x=\begin{bmatrix}1&1\\0&1\end{bmatrix}^3\begin{bmatrix}1&0\\1&1\end{bmatrix}^2$.

We state the following definition which is used in the proof of the next proposition:

DEFINITION 4.1.8: Two integral matrices S and T are said to be <u>shift</u> <u>equivalent</u> <u>over</u> <u>$\mathbb{Z}$</u> if there exist integral matrices U and V and $l\in\mathbb{N}$ such that

$$SU=UT,\ VS=TV,\ UV=S^l,\text{ and } VU=T^l.$$

l is called the <u>lag</u> of the shift equivalence. We can similarly define shift equivalence over $\mathbb{Z}_+$.

PROPOSITION 4.1.9: Suppose $\mathcal{A}$ and $\mathcal{B}$ are non-negative, hyperbolic matrices in GL(2,$\mathbb{Z}$). The following are equivalent:

i) $\mathcal{A}$ and $\mathcal{B}$ are similar over $\mathbb{Z}$

ii) There is a matrix $P \geq 0$, $P \in GL(2,\mathbb{Z})$, such that $P^{-1}\mathcal{A}P = \mathcal{B}$

iii) The defining sequence for $\mathcal{A}$ is a cyclic permutation of the defining sequence for $\mathcal{B}$ up to two consecutive 0's.

Proof: *ii*)$\Rightarrow$*i*) is obvious.

iii)$\Rightarrow$*ii*) If the defining sequence for $\mathcal{A}$ is $\{x_i\}_{i=1}^{n}$ and the defining sequence for $\mathcal{B}$ is $x_k x_{k+1} \ldots x_n x_1 x_2 \ldots x_{k-1}$ for $1 < k \leq n$ then $P = x_k x_{k+1} \ldots x_n$ is the desired matrix.

i)$\Rightarrow$*iii*). One way to do this is to see that *i*) $\Rightarrow$ shift equivalent over $\mathbb{Z}$ $\Rightarrow$ shift equivalent over $\mathbb{Z}_+$ $\Rightarrow$ *iii*). To show the first implication, if there is a $\phi \in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi = \mathcal{B}$ then $S = \mathcal{A}$, $T = \mathcal{B}$, $U = \phi$, and $V = \phi^{-1}\mathcal{A}$ is a shift equivalence with lag 1. The second implication is a result of Kim and Roush [7]. We then need to prove the third implication. This proof is due to Charlie Jacobson. Suppose $\mathcal{A}$ and $\mathcal{B}$ are shift equivalent and that $\mathcal{A}$ and $\mathcal{B}$ have defining sequences $\{a_i\}_{i=1}^{n}$ and $\{b_i\}_{i=1}^{m}$. Then there exist U, V and l as in the definition above. Note that $\det(U) = \pm 1$ and $\det(V) = \pm 1$ so that U and V have defining sequences $\{u_i\}_{i=1}^{p}$ and $\{v_i\}_{i=1}^{q}$ respectively. By the uniqueness of these factorizations, $v_1 = a_1$ and $u_p = a_n$ so that not both v_1 and u_p are x. Then the defining sequence for UV is $\{u_1, u_2, \ldots, u_p, v_1, v_2, \ldots, v_q\}$. Therefore, $p+q=mk$. Similarly $p+q=nk$, so $m=n$. Since $UV = \mathcal{B}^l$, the uniqueness of the defining sequence implies that $u_p = b_r$ for some $1 < r < m$. Since $VU = \mathcal{A}^l$, $a_1 = b_{r+1}$, $a_2 = b_{r+2}$, $\ldots$, $a_{n-r} = b_n$, $a_{n-r+1} = b_1$, $\ldots$, $a_n = b_r$. Hence, the defining sequence for $\mathcal{A}$ is a cyclic permutation of the defining sequence for $\mathcal{B}$ and the proposition is proven. □

Note: This proposition can also be proven directly by letting $w=\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ and showing that the group generated by x, y, and w has a solvable word problem and is in fact GL(2,$\mathbb{Z}$). Then using this machinery and a canonical form for any element of GL(2,$\mathbb{Z}$) involving x, y, and w, show *i*) $\Leftrightarrow$ new canonical form is cyclic permutation $\Leftrightarrow$ *iii*).

EXAMPLE 4.1.10: The defining sequence for $\begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}$ is *yxyyxyyx* which is a cyclic permutation of the sequence for $\begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$ in Example 4.1.4 hence the two matrices are similar over $\mathbb{Z}$. The defining sequence for $\begin{bmatrix} 7 & 6 \\ 1 & 1 \end{bmatrix}$ is *yxyxyxyxyxyy* and therefore $\begin{bmatrix} 7 & 6 \\ 1 & 1 \end{bmatrix}$ is not similar over $\mathbb{Z}$ to either of the above matrices although they are similar over $\mathbf{R}$.

SECTION 4.2: PARTITIONS WITH TWO RECTANGLES

We are now ready to begin examining partitions with two rectangles with the following lemma.

LEMMA 4.2.1: Let $\mathcal{P}$ be a Markov partition with two rectangles for $\mathcal{A}$ with Markov matrix M and suppose there is a shear matrix S and a matrix $M' \geq 0$ such that $S^{-1}MS=M'$. Then there is a Markov partition $\mathcal{P}'$ for $\mathcal{A}$ with Markov matrix M'.

Proof: We give the proof for $S=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\mathrm{Tr}(\mathcal{A})>0$ with the proofs for $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $\mathrm{Tr}(\mathcal{A})<0$ being similar. Suppose first that $\mathcal{C}(\mathcal{P})$=origin. Consider $S^{-1}MS=\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $=\begin{bmatrix} a-c & a-c+b-d \\ c & c+d \end{bmatrix}=M'$. In order for this matrix to be non-negative, row 1 must be the dominant row in M. Row 1 being dominant means that $\mathcal{A}(R_1)$ crosses each rectangle at least as many times as $\mathcal{A}(R_2)$ which says that R_1 is strictly longer than R_2 in the unstable direction. If R_1 is strictly longer than R_2, we can divide R_1 into two pieces, R_1' and L as shown in Figure 4.2.3. To divide R_1 in this manner, draw R_1 and R_2 with their stable sides farthest from the origin side by side (Figure 4.2.2). We then extend segment $\overline{CB}$ to $\overline{CD}$ and now have a Markov partition $\mathcal{T}$ with three rectangles, R_1', L, and R_2. The Markov matrix for $\mathcal{T}$ is

$$M(\mathcal{T})=\begin{bmatrix} a-c & a-c & b-d \\ c & c & d \\ c & c & d \end{bmatrix}.$$

For row 3, $\mathcal{A}(R_2)$ will cross R_2 the same number of times as in $\mathcal{P}$, and L and R_1' the number of times it crossed R_1 in $\mathcal{P}$. For row 2, $\mathcal{A}(L)$ will cross the same rectangles as $\mathcal{A}(R_2)$; because of the expansive nature of the unstable boundary, the boundary between R_2 and L (segment $\overline{AB}$) gets mapped to the interior of any rectangle it crosses (see Example 1.2.9). $\mathcal{A}(R_1')$ will then cross everything that R_1 used to cross that isn't crossed by $\mathcal{A}(L)$. We have then seen that $S^{-1}MS$ being non-negative (in particular, multiplication on the left by S^{-1}) means that we can split R_1 into two rectangles, one of which has the same "personality" as R_2. We now can remove segment $\overline{AB}$ from $\mathcal{T}$ and get a new partition $\mathcal{P}'$ (Figure 4.2.4). This is possible because $\mathcal{A}(\overline{AB})\cap\partial_u\mathcal{P}=\emptyset$. We claim that the Markov matrix for $\mathcal{P}'$ is M'. The rectangles for $\mathcal{P}'$ are $R_2'=R_2\cup L$ and R_1'. R_2' will be crossed by everything which crossed either R_2 or L hence column 2 of M' is $\Big\{\{\text{column 2 of } M(\mathcal{T})\}+\{\text{column 3 of } M(\mathcal{T})\}$ with either the second or third entry deleted$\Big\}$. Column 1 of M' is $\Big\{$column 1 of $M(\mathcal{T})$ with either the second or third entry deleted$\Big\}$. Hence,

$$M'=\begin{bmatrix} a-c & a-c+b-d \\ c & d+c \end{bmatrix}.$$

This concludes the proof for $\mathcal{C}(\mathcal{P})=\{\text{origin}\}$. If $\mathcal{C}=\{\text{any fixed point not the origin}\}$, then by translation the above proof works. We then must deal with the case when $\mathcal{C}(\mathcal{P})=\{$two fixed

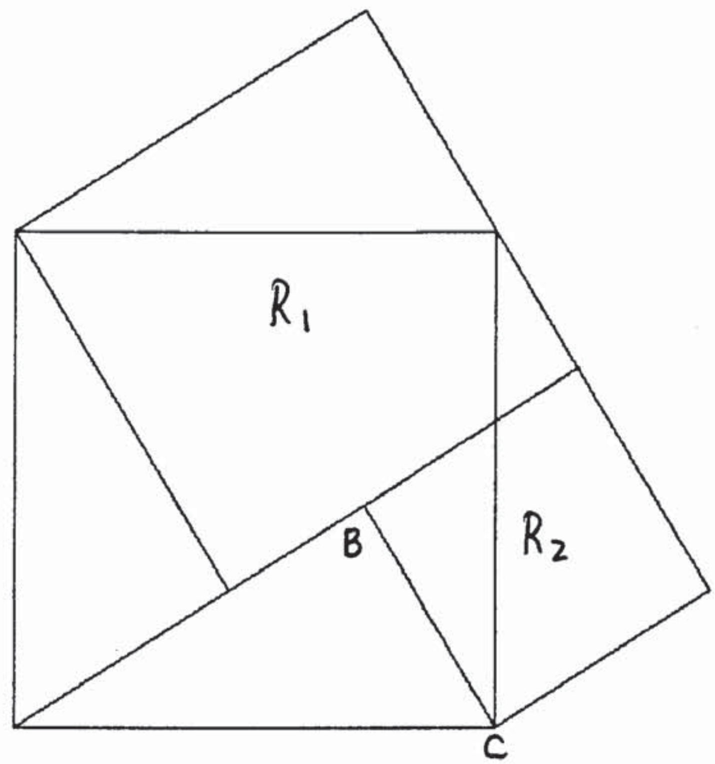

Figure 4.2.2: Lining Up the Partition

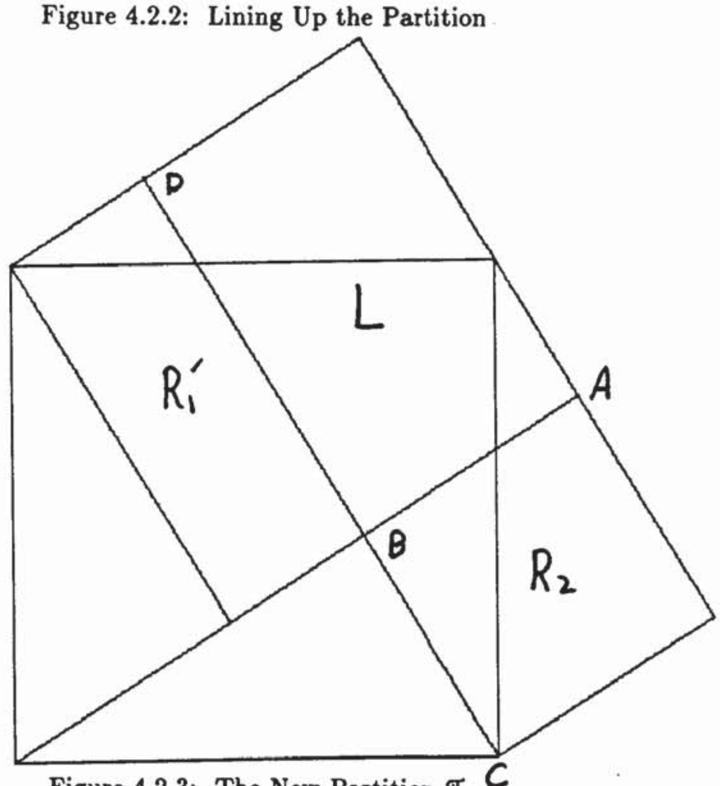

Figure 4.2.3: The New Partition $\mathfrak{T}$

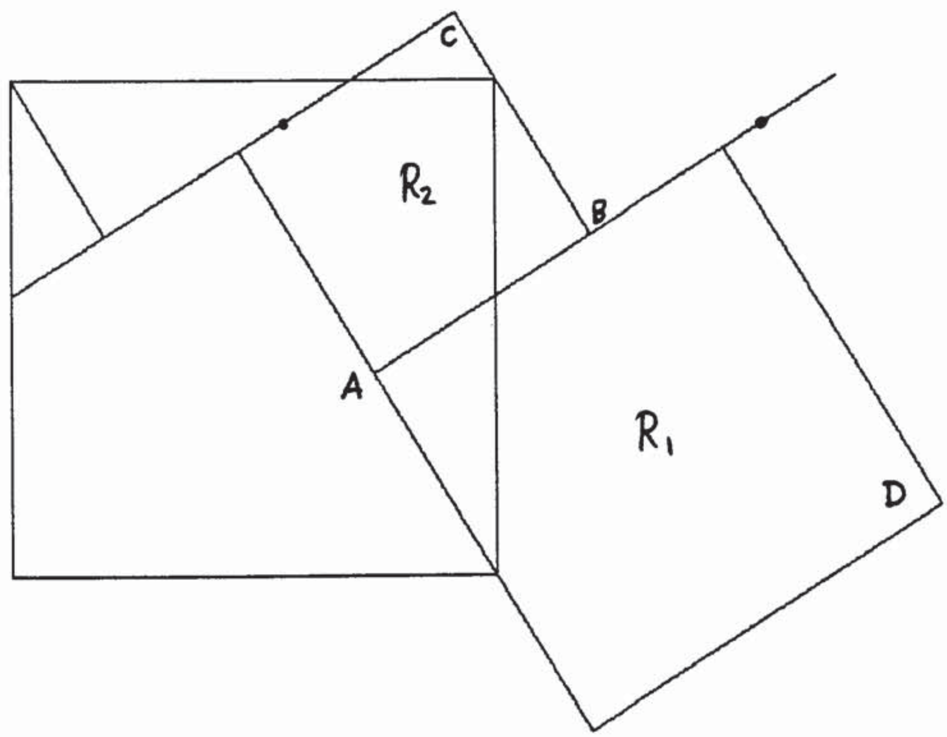

Figure 4.2.4: The New Partition $\mathcal{P}'$

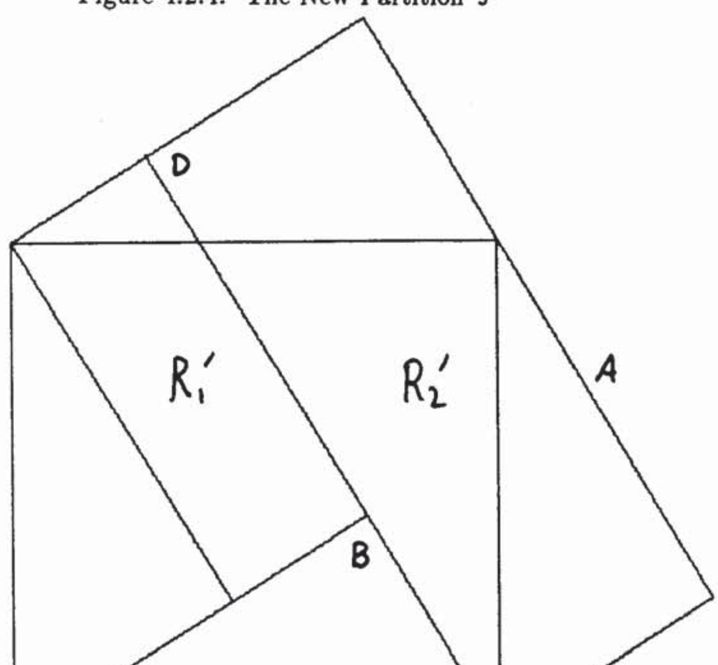

Figure 4.2.5: Lining Up the Partition

points}. Notice that if there are more than two fixed points in the core or any periodic point of higher period, then there are more than two rectangles (Proposition 3.3.9). We will assume in our diagrams that one of these fixed points is the origin. The only part of the above proof which we must modify is how to divide R_1. First, we must have $\mathcal{U}$={one fixed point} and $\mathcal{S}$={the other fixed point}. We will assume WLOG that $\mathcal{S}=0$ and $\mathcal{U}$={the other fixed point}, call it Q. We want to draw $\mathcal{P}$ so that there is no fixed point on $\overline{AB}$ as in Figure 4.2.5. It is then possible to extend $\overline{CB}$ to $\overline{CD}$ and obtain a partition $\mathcal{T}$, then remove $\overline{AB}$ as above. □

EXAMPLE 4.2.6: Consider the partition in Figure 4.2.8 for the matrix $\left[\begin{smallmatrix}3&2\\1&1\end{smallmatrix}\right]$. The Markov matrix for the partition is the original matrix $\left[\begin{smallmatrix}3&2\\1&1\end{smallmatrix}\right]$. Since $\left[\begin{smallmatrix}1&-1\\0&1\end{smallmatrix}\right]\left[\begin{smallmatrix}3&2\\1&1\end{smallmatrix}\right]\left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]=$ $\left[\begin{smallmatrix}1&2\\1&3\end{smallmatrix}\right]$ which is non-negative, we should be able to perform the above operation. We extend segment $\overline{CB}$ to $\overline{CD}$ to get a new partition shown in Figure 4.2.10. The Markov matrix for this partition is

$$M=\begin{bmatrix}1&2&2\\1&2&2\\0&1&1\end{bmatrix}$$

We can then remove segment $\overline{AB}$ to obtain the partition in Figure 4.2.11 and the Markov matrix for this partition is $M=\left[\begin{smallmatrix}1&2\\1&3\end{smallmatrix}\right]$.

We use the lemma to prove the following proposition which will in turn give us the desired theorem.

PROPOSITION 4.2.7: Let $\mathcal{P}$ be a Markov partition with two rectangles for $\mathcal{A}$ with Markov matrix M and suppose there is a matrix $M'\geq 0$ and $\phi \in GL(2,\mathbf{Z})$ such that $\phi^{-1}M\phi= M'$. Then there exists a Markov partition $\mathcal{P}'$ for $\mathcal{A}$ with Markov matrix M'.

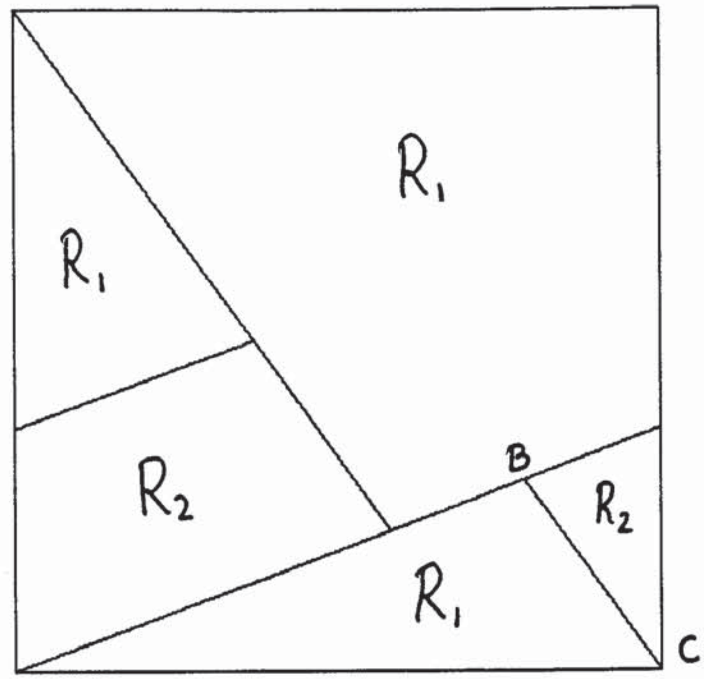

Figure 4.2.8: The Partition $\mathcal{P}$ of Example 4.2.6

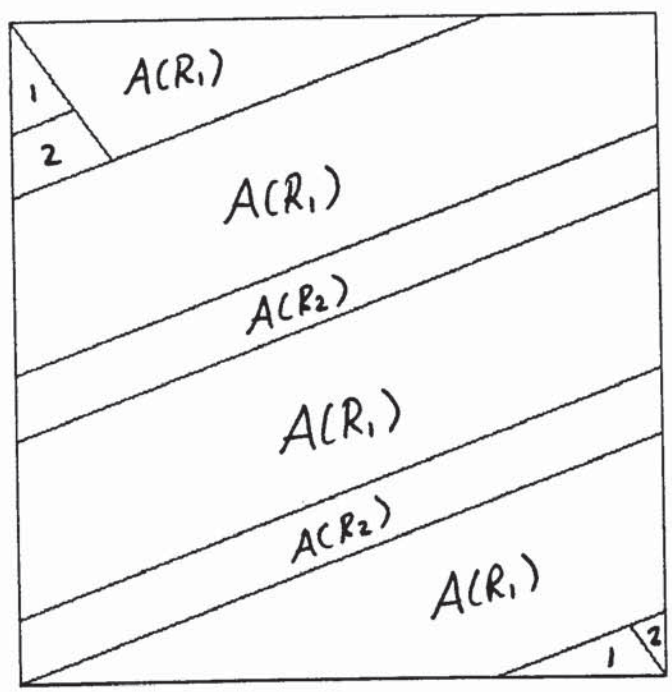

Figure 4.2.9: $\mathcal{A}(\mathcal{P})$

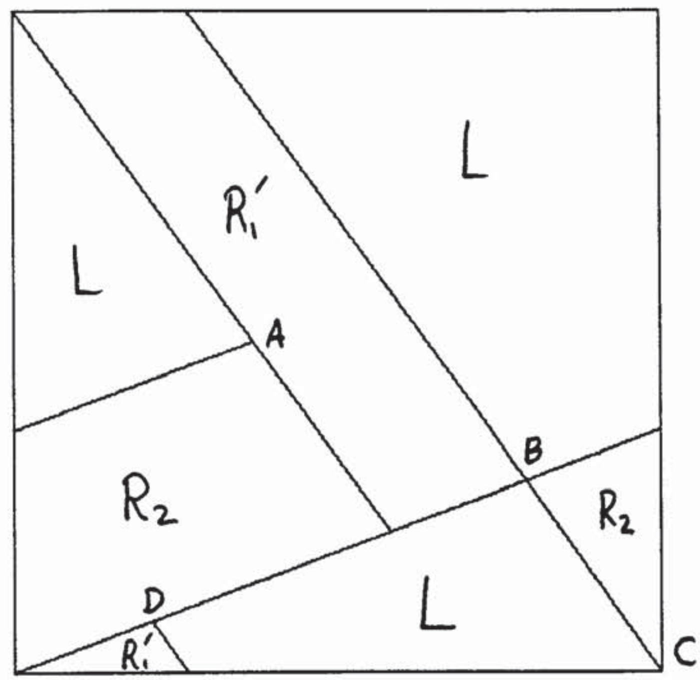

Figure 4.2.10: Extending $\overleftrightarrow{CB}$

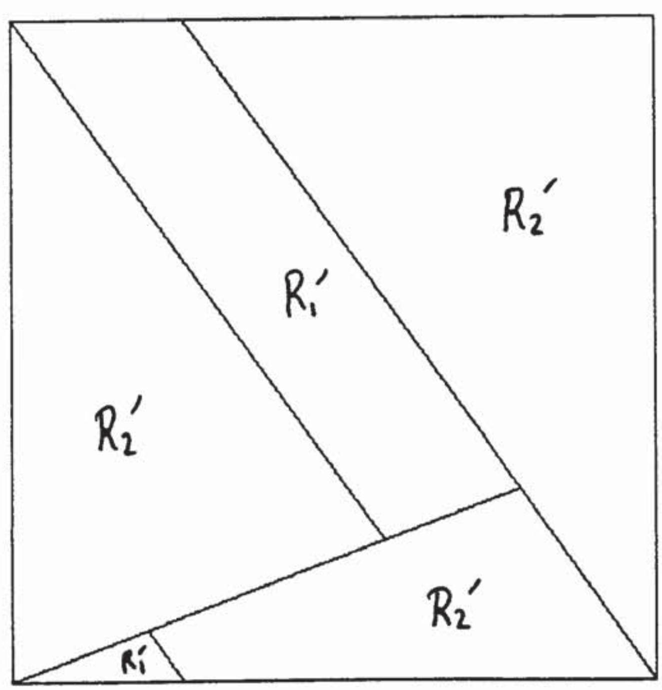

Figure 4.2.11: The New Partition

Proof: Suppose such a ϕ exists. Then by Proposition 4.1.9 there exists a matrix $P \in GL(2,\mathbb{Z})$ with $P \geq 0$ such that $P^{-1}MP = M'$. We know by Proposition 4.1.6 that $P = M_1M_2 \ldots M_k$ where one of the following is true: *i*) $k=1$ and $M_1 = x$; *ii*) M_i is a shear matrix for $1 \leq i \leq k$; or *iii*) M_i is a shear for $1 \leq i \leq k-1$ and M_k is either y or y' where x and y are as in Definition 4.1.3.

Case *i*): Conjugating M by x corresponds to relabeling the rectangles; i.e. R_1 as R_2 and R_2 as R_1.

Case *ii*): We know by Lemma 4.2.1 that conjugating M by a shear matrix is "legal" as far as partitions go. Therefore, since $P^{-1}MP = M_k^{-1}M_{k-1}^{-1} \ldots M_1^{-1}MM_1 \ldots M_{k-1}M_k$ we need only show that $M_j^{-1} \ldots M_1^{-1}MM_1 \ldots M_j \geq 0$ for $1 \leq j \leq k$. It is sufficient to show that $M_1^{-1}MM_1 \geq 0$. So, suppose $P^{-1}MP \geq 0$ with P as above. Note at this point that multiplication on the right by a shear matrix does not change the dominant row of M and hence MP and M have the same dominant row. Therefore since $M_k^{-1} \ldots M_1^{-1}MP \geq 0$, we know that $M_1^{-1}MP \geq 0$. Since M_1^{-1} is the inverse of a shear matrix and MP and M have the same dominant row, $M_1^{-1}M \geq 0$ also. Hence $M_1^{-1}MM_1 \geq 0$ which concludes case *ii*).

Case *iii*): If $M_1 = y$ or y' then let $P' = xP$. Now P' is a product of shear matrices and reduces to case *ii*). We see, however, that $xP'^{-1}MP'x$ is just $P^{-1}MP$ and hence by case *i*) we have proven our proposition. □

We now give necessary and sufficient conditions on the Markov matrix for a hyperbolic toral automorphism.

THEOREM 4.2.12: Let $\mathcal{A}$ be a hyperbolic automorphism of $\mathbf{T}^2$, and let M be a non-negative 2×2 integer matrix.

i) If $\mathrm{Tr}(\mathcal{A})>0$, then there exists a Markov partition $\mathcal{P}$ for $\mathcal{A}$ with Markov matrix M $\Leftrightarrow$ there exists $\phi \in GL(2,\mathbf{Z})$ such that $\phi^{-1}\mathcal{A}\phi= M$.

ii) If $\mathrm{Tr}(\mathcal{A})<0$, then there exists a Markov partition $\mathcal{P}$ for $\mathcal{A}$ with Markov matrix M $\Leftrightarrow$ there exists $\phi \in GL(2,\mathbf{Z})$ such that $\phi^{-1}\mathcal{A}\phi= -M$.

Proof:

($\Rightarrow$) In both cases this is Corollary 3.2.8.

($\Leftarrow$) Suppose $\mathrm{Tr}(\mathcal{A})>0$, and suppose there exists $\phi \in GL(2,\mathbf{Z})$ such that $\phi^{-1}\mathcal{A}\phi= M$. Let $\mathcal{P}'$ be any Markov partition for $\mathcal{A}$ with two rectangles. $\mathcal{P}'$ gives rise to a Markov matrix, call it M'. By Corollary 3.2.8, there exists $\psi \in GL(2,\mathbf{Z})$ such that $\psi^{-1}\mathcal{A}\psi= M'$. Then $\phi^{-1}\psi M'\psi^{-1}\phi= M \Rightarrow (\psi^{-1}\phi)^{-1}M'(\psi^{-1}\phi)= M \Rightarrow$ by Proposition 4.2.7 there exists a partition $\mathcal{P}$ for $\mathcal{A}$ with Markov matrix M. If $\mathrm{Tr}(\mathcal{A})<0$, let $\mathcal{P}'$ be any Markov partition for $\mathcal{A}$ with two rectangles and Markov matrix $-M$ and Proposition 4.2.7 gives us the theorem as above. □

SECTION 4.3: A CONJECTURE ABOUT r RECTANGLES

We know that the Markov matrix M for any hyperbolic automorphism of $\mathbf{T}^2$ must be aperiodic (there is some integer $n\geq 1$ such that $M^n>0$) because the image of every rectangle is dense in $\mathbf{T}^2$. It is unknown whether or not this enough for the converse of Corollary 3.2.7 to be true. We know by Chapter 3 that there are some restrictions as to the frequency with which the roots of unity can occur. For that reason, we give the conjecture under the assumptions that $\mathrm{Tr}(\mathcal{A})>0$, $\det(\mathcal{A})=1$, and a crossing at each point in $\mathbb{C}$ with similar conjectures for the other cases unstated.

CONJECTURE 4.3.1: Let $\mathcal{A}$ be a hyperbolic automorphism of $\mathbf{T}^2$ ($\mathrm{Tr}(\mathcal{A})>0$ and $\det(\mathcal{A})=1$) and let M be an $r\times r$ integer matrix, $r>2$. Then there is an FCC Markov partition $\mathcal{P}$ with r rectangles for $\mathcal{A}$ with Markov matrix M $\Leftrightarrow$ M is aperiodic and M is similar over $\mathbf{Z}$ to a matrix of the following form:

$$\begin{bmatrix} \mathcal{A} & 0 & \cdots & 0 \\ * & A_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & A_k \end{bmatrix}$$

with the following restrictions:

i) If more than one of the A_i are non-zero, then at least one but at most two of the non-zero A_i have eigenvalues equal to {the pth roots of unity excluding 1 for some p}, and the rest of the non-zero A_i have eigenvalues equal to {the pth roots of unity for some p}. Further, r must be large enough by the standards of Lemma 3.3.9 and $\mathcal{A}$ must have period p orbits for each p appearing above. Also, the roots of unity for each p must occur $3m$ times for some m.

ii) If only A_1 is non-zero, then $A_1=[1]$.

CHAPTER 5 - A GENERATING SET

SECTION 5.1: CONTINUED FRACTIONS

In this chapter we will show that there is a finite set of partitions, all having the same core, which generate all other partitions with the same core under the action of GL(2,$\mathbb{Z}$). This leads to a generating set for all partitions with two rectangles. One of the tools we will use is the theory of continued fractions. The goal of this section is to introduce a few of the ideas from this theory. For a more detailed introduction see Olds [9].

Given any real number $\alpha \geq 0$, we can write α in the following form:

$$\alpha = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \ddots}}}$$

where $a_i \in \mathbb{Z}$. To do this, we write $\alpha = a_1 + a_1'$ where a_1 is the integer part of α and $0 \leq a_1' < 1$. If $a_1' = 0$, we are done. If not, $a_1' = \frac{1}{a_1''}$ where $a_1'' > 1$. We then repeat the process on $a_1'' = a_2 + a_2'$. As a shorthand, we write $\alpha = [a_1, a_2, a_3, a_4, \ldots]$.

DEFINITION 5.1.1: We call $[a_1, a_2, a_3, \ldots]$ the <u>continued fraction expansion of α</u>.

If this sequence terminates at some point, then $\alpha \in \mathbf{Q}$ and we write $\alpha = [a_1, a_2, a_3, \ldots, a_n]$. If the a_i's repeat at some point, we write $\alpha = [a_1, a_2, a_3, \ldots, \overline{a_k, \ldots a_l}]$ where $a_k \ldots a_l$ is the block which is repeated. If α is negative, then $\alpha=-a_1+\beta$ where a_1 is a positive integer and $0\leq\beta<1$. β has a continued fraction expansion $\beta=[0, b_2, b_3, \ldots]$ and we say that the continued fraction expansion for α is $[-a_1, b_2, b_3, \ldots]$. Of course, the continued fraction expansion for any integer z is simply $[z]$.

Recall that the characteristic polynomial for a hyperbolic matrix $\mathcal{A}\in GL(2,\mathbb{Z})$ is $\chi=\lambda^2-\mathrm{Tr}(\mathcal{A})\lambda+\det(\mathcal{A})$.

PROPOSITION 5.1.2: Let T be a positive integer.

a) If λ is the larger (in modulus) root of $\lambda^2-T\lambda-1$ then λ has continued fraction expansion $[\overline{T}]$.

b) If λ is the larger root of $\lambda^2-T\lambda+1$ then λ has continued fraction expansion $[(T-1), \overline{1, (T-2)}]$.

c) If λ is the larger root of $\lambda^2+T\lambda+1$ then λ has continued fraction expansion $[-T, (T-1), \overline{1, (T-2)}]$.

d) If λ is the larger root of $\lambda^2-T\lambda-1$ then λ has continued fraction expansion $[-(T+1), 1, (T-1), \overline{T}]$.

Proof: a) $\lambda^2-T\lambda-1=0 \Rightarrow \lambda^2=T\lambda+1 \Rightarrow \lambda=T+\frac{1}{\lambda} \Rightarrow \lambda=[\overline{T}]$.

b) $\lambda=\frac{T+\sqrt{T^2-4}}{2}=\frac{T}{2}+\frac{\sqrt{T^2-4}}{2}=(T-1)+1-\frac{T}{2}+\frac{\sqrt{T^2-4}}{2}=(T-1)+\frac{2-T}{2}+\frac{\sqrt{T^2-4}}{2}=$

$$(T-1)+\frac{(2-T)+\sqrt{T^2-4}}{2}=(T-1)+\frac{1}{x} \text{ where } x=\frac{2}{(2-T)+\sqrt{T^2-4}}.$$

$$x=\frac{-2}{(T-2)-\sqrt{T^2-4}}=\frac{(T-2)\{(T-2)-(T+2)\}}{2(T-2)\{(T-2)-\sqrt{T^2-4}\}}=\frac{(T-2)^2-(T^2-4)}{2(T-2)\{(T-2)-\sqrt{T^2-4}\}}$$

$$=\left\{\frac{(T-2)-\sqrt{T^2-4}}{(T-2)-\sqrt{T^2-4}}\right\}\cdot\left\{\frac{(T-2)+\sqrt{T^2-4}}{2(T-2)}\right\}=\frac{(T-2)+\sqrt{T^2-4}}{2(T-2)} \Rightarrow \text{x is satisfies}$$

$$(T-2)x^2-(T-2)x-1=0 \Rightarrow x=1+\frac{1}{(T-2)x} \Rightarrow x=[1, \overline{(T-2), 1}].$$

Therefore, $\lambda=(T-1)+\frac{1}{x} \Rightarrow \lambda=[(T-1), \overline{1,(T-2)}]$.

c) If λ satisfies $\lambda^2+\lambda+1=0$ then $-\lambda$ satisfies $\lambda^2-\lambda+1=0$ hence $-\lambda=[T-1, \overline{1, T-2}]$ by part b). With this in mind we proceed as follows.

$$\lambda^2=-T\lambda-1 \Rightarrow \lambda=-T-\frac{1}{\lambda} \Rightarrow \lambda=-T+\frac{1}{-\lambda} \Rightarrow \lambda=[-T, T-1, \overline{1, T-2}].$$

d) If λ satisfies $\lambda^2+T\lambda-1=0$ then $-\lambda$ satisfies $\lambda^2-T\lambda-1=0$ hence $-\lambda=[\,\overline{T}]$.

$$\lambda^2+T\lambda-1=0 \Rightarrow T\lambda^2+T^2\lambda-T=0 \Rightarrow \lambda^2T-\lambda=-T^2\lambda+T-\lambda \Rightarrow$$

$$\lambda(\lambda T-1)=(-\lambda T^2+T-\lambda T+1)+\lambda T-\lambda-1 \Rightarrow \lambda(\lambda T-1)=-(T+1)(\lambda T-1)+\lambda(T-1)-1 \Rightarrow$$

$$\lambda=-(T+1)+\frac{\lambda(T-1)-1}{\lambda T-1} = -(T+1)+\frac{1}{\frac{\lambda T-\lambda-1+\lambda}{\lambda(T-1)-1}}=-(T+1)+\frac{1}{1+\frac{\lambda}{\lambda(T-1)-1}} \Rightarrow$$

$$\lambda=-(T+1)+\frac{1}{1+\frac{1}{(T-1)+\frac{1}{-\lambda}}} \Rightarrow \lambda=[-(T+1), 1, T-1, \overline{T}].$$ □

We have seen that the eigenvalues of $\mathcal{A}$ have continued fraction expansions which fall into one of four categories. The continued fraction expansions for the eigenvectors, however, have no such regularity. The following series of propositions will demonstrate this fact in one specific case which will be relevant to future examples.

DEFINITION 5.1.3: We define the <u>Fibonacci</u> <u>numbers</u> as follows:

$F_0=0$; $F_1=1$; and $F_i=F_{i-1}+F_{i-2}$ for $i\geq 2$ where F_i denotes the ith Fibonacci number.

LEMMA 5.1.4: $F_n^2=(-1)^{n+1}+F_{n+1}F_{n-1}$ for $n\geq 1$.

Proof: We proceed by induction. Clearly true for $n=1,2,3,4$. So assume true for $k<n$.

$$F_n^2-F_{n+1}F_{n-1}=F_n(F_{n-1}+F_{n-2})-F_{n+1}F_{n-1}=F_{n-1}(F_n-F_{n+1})+F_nF_{n-2}.$$

Since $F_{n+1}=F_n+F_{n-1}$, we see that $F_n-F_{n+1}=-F_{n-1}$ so the above equality becomes

$$F_n^2-F_{n+1}F_{n-1}=F_{n-1}(-F_{n-1})+F_nF_{n-2}=(-1)(F_{n-1}^2-F_nF_{n-2}).$$

By induction $F_{n-1}^2-F_nF_{n-2}=(-1)^n$ hence

$$F_n^2-F_{n+1}F_{n-1}=(-1)((-1)^n+F_nF_{n-2}-F_nF_{n-2})=(-1)^{n+1}. \qquad \square$$

LEMMA 5.1.5: $[\underbrace{1,1,1,\ldots,1}_{(n-1)},a,x]=\dfrac{(F_na+F_{n-1})x+F_n}{(F_{n-1}a+F_{n-2})x+F_{n-1}}$ for $n\geq 2$ and $a\in\mathbb{Z}_+$.

Proof: When $n=2$, $1+\dfrac{1}{a+\frac{1}{x}}=\dfrac{(a+1)x+1}{ax+1}$. Therefore, assume true for $k\leq n$ and prove for $n+1$.

$$[\underbrace{1,1,1,\ldots,1}_{n},a,x]=1+\cfrac{1}{1+\cfrac{1}{\ddots\,\cfrac{}{1+\cfrac{1}{a+\frac{1}{x}}}}}=1+\cfrac{1}{\cfrac{(F_na+F_{n-1})x+F_n}{(F_{n-1}a+F_{n-2})x+F_{n-1}}}$$

$$=\frac{(F_na+F_{n-1})x+F_n}{(F_na+F_{n-1})x+F_n}+\frac{(F_{n-1}a+F_{n-2})x+F_{n-1}}{(F_na+F_{n-1})x+F_n}=\frac{(F_{n+1}a+F_n)x+F_n}{(F_na+F_{n-1})x+F_n}. \qquad \square$$

PROPOSITION 5.1.6: The matrix $B_n=\begin{bmatrix} F_n & F_n a+F_{n-1} \\ F_{n+1} & F_{n+1}a+F_n \end{bmatrix}$ for $n\geq 1$ and $a\in\mathbb{Z}_+$ has the following properties:

a) $\det(B_n)=(-1)^{n+1}$.

b) $\nexists\, R\in GL(2,\mathbb{Z})$ such that $R^k=B_n$ for any positive integers $n\geq 1$ or $k\geq 2$.

c) $m_u=[\overline{\underbrace{1, 1, \ldots, 1}_{n}, a}]$ where m_u is the slope of the unstable eigenvector of B_n.

Proof: a) $\det(B_n)=\begin{vmatrix} F_n & F_n a+F_{n-1} \\ F_{n+1} & F_{n+1}a+F_n \end{vmatrix}=F_n^2-F_{n+1}F_{n-1}=(-1)^{n+1}$ by Lemma 5.1.4.

b) Since B_n is a non-negative, integer matrix we factor B_n using the methods of Section 4.1.

$$\begin{bmatrix} F_n & F_n a+F_{n-1} \\ F_{n+1} & F_{n+1}a+F_n \end{bmatrix}= x\cdot\begin{bmatrix} F_{n+1} & F_{n+1}a+F_n \\ F_n & F_n a+F_{n-1} \end{bmatrix}$$

$$=xy\cdot\begin{bmatrix} F_n & F_n a+F_{n-1} \\ F_{n-1} & F_{n-1}a+F_{n-2} \end{bmatrix}=xy^{n-1}\cdot\begin{bmatrix} F_2 & F_2 a+F_1 \\ F_1 & F_1 a+F_0 \end{bmatrix}$$

$$=xy^{n-1}\cdot\begin{bmatrix} 1 & a+1 \\ 1 & a \end{bmatrix}=xy^n\cdot\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}=xy^n(yx)^a.$$

Therefore by the uniqueness of the factorization of B_n, part b) is proven.

c) We give the proof for n odd (hence $\det(B_n)=1$) with the proof for n even being entirely similar.

$$\lambda_u=\frac{F_{n+1}a+2F_n+\sqrt{(F_{n+1}a+2F_n)^2-4}}{2}\text{ hence}$$

$$m_u = \frac{F_{n+1}a + \sqrt{(F_{n+1}a + 2F_n)^2 - 4}}{2(F_n a + F_{n-1})} = \frac{F_{n+1}a + \sqrt{(F_{n+1}a)^2 + 4F_{n+1}F_n a + 4F_n^2 - 4}}{2(F_n a + F_{n-1})}$$

$$= \frac{F_{n+1}a + \sqrt{(F_{n+1}a)^2 + 4F_{n+1}F_n a + 4(F_{n+1}F_{n-1} + 1) - 4}}{2(F_n a + F_{n-1})}$$

$$= \frac{F_{n+1}a + \sqrt{(F_{n+1}a)^2 + 4F_{n+1}F_n a + 4F_{n+1}F_{n-1}}}{2(F_n a + F_{n-1})} = \frac{F_{n+1}a + \sqrt{(F_{n+1}a)^2 + 4F_{n+1}(F_n a + F_{n-1})}}{2(F_n a + F_{n-1})} \Rightarrow$$

$$m_u \text{ satisfies } (F_n a + F_{n-1})m_u^2 - F_{n+1}am_u - F_{n+1} = 0 \Rightarrow$$

$$(F_n a + F_{n-1})m_u^2 + F_n m_u = (F_{n+1}a + F_n)m_u + F_{n+1} \Rightarrow$$

$$m_u = \frac{(F_{n+1}a + F_n)m_u + F_{n+1}}{(F_n a + F_{n-1})m_u + F_n} \Rightarrow$$

$$m_u = [\overline{\underbrace{1, 1, \ldots, 1}_{n}, a}] \text{ by Lemma 5.1.5.} \qquad \square$$

Suppose $\alpha \in \mathbf{R}$ has continued fraction expansion $\alpha = [a_1, a_2, \ldots]$. It is clear then that $c_1 = [a_1]$, $c_2 = [a_1, a_2]$, ..., $c_i = [a_1, a_2, \ldots, a_i]$ are good rational approximations to α and that c_i is a better approximation that c_{i-1}.

DEFINITION 5.1.7: We call c_i as defined above the <u>ith convergent of α</u>. If $c_i > \alpha$ we call c_i an <u>upper convergent of α</u>. If $c_i < \alpha$ we call c_i a <u>lower convergent of α</u>.

It can be shown that if i is even, then c_i is an upper convergent and if i is odd, c_i is a lower convergent.

We have a general formula for computing the convergents of a number α given the continued fraction expansion of $\alpha=[a_1, a_2, \ldots]$. If $c_i=[a_1, a_2, \ldots, a_i]=\frac{p_i}{q_i}$ then

$$c_{i+1}=[a_1, \ldots, a_{i+1}]=\frac{a_{i+1}p_i+p_{i-1}}{a_{i+1}q_i+q_{i-1}}=\frac{p_{i+1}}{q_{i+1}}.$$

A proof of this fact and the fact that p_i and q_i are relatively prime can be found in Olds [9]. In order to use this recursive formula, we must have values for p_0, q_0, p_{-1}, and q_{-1}. If we assign the values

$$p_0=1, \qquad p_{-1}=0,$$

$$q_0=0, \qquad q_{-1}=1,$$

we see that c_1 and c_2 agree with our earlier notion $c_1=a_1$, etc. and we can use our recursive formula to compute c_i for any $i\geq 1$.

EXAMPLE 5.1.8: The convergents of $\frac{1+\sqrt{5}}{2}=[\bar{1}]$ are $c_i=\frac{p_i}{q_i}=\frac{F_{i+1}}{F_i}$. To show this, we compute the convergents.

$$c_1=\frac{1\cdot p_0+p_{-1}}{1\cdot q_0+q_{-1}}=\frac{1+0}{0+1}=\frac{p_1}{q_1}$$

$$c_2=\frac{1\cdot p_1+p_0}{1\cdot q_1+q_0}=\frac{1+1}{1+0}=\frac{2}{1}=\frac{p_2}{q_2}$$

$$c_i=\frac{1\cdot p_{i-1}+p_{i-2}}{1\cdot q_{i-1}+q_{i-2}}=\frac{p_i}{q_i}=\frac{F_{i+1}}{F_i}.$$

Convergents have been proven to be the best rational approximations to an irrational number as seen in the following theorem. The proof can be found in Olds [9].

THEOREM 5.1.9: Let $c_n = \frac{p_n}{q_n}$ be the nth convergent to an irrational number α. Then

$$\frac{1}{2q_n q_{n+1}} < \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2},\ n \geq 1.$$

This estimate can be made even sharper: Of any two consecutive convergents $\frac{p_n}{q_n}$ and $\frac{p_{n+1}}{q_{n+1}}$, at least one, call it $\frac{p}{q}$ satisfies

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}.$$

Further, if α is an irrational number, and if $\frac{p}{q}$ is a rational fraction in lowest terms with $q \geq 1$, such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$$

then $\frac{p}{q}$ is necessarily one of the convergents of α.

Convergents also have a geometric significance. We identify each convergent $\frac{p_i}{q_i}$ with the integer lattice point (q_i, p_i). If we consider the line y=αx, the convergents of α are good approximations to the slope of the line y=αx. In fact, they are the best approximations in following sense: if we place a peg at each point in the integer lattice and lay two strings on the portion of the line y=αx which is in the first quadrant, then separate the two strings at the origin, the pegs which the strings touch are exactly the points which correspond to the convergents of α [9]. See Figure 5.1.10. Let $\mathcal{A} \in GL(2,\mathbb{Z})$ be a matrix which has $\begin{bmatrix} 1 \\ \alpha \end{bmatrix}$ as its unstable eigenvector. It is clear from the above discussion that the image of a convergent of α under $\mathcal{A}$ must be another convergent. It is also clear that if $k \geq 1$ is a fixed integer and

$$\mathcal{A}\begin{bmatrix} q_i \\ p_i \end{bmatrix} = \begin{bmatrix} q_{i+k} \\ p_{i+k} \end{bmatrix}$$

for some $i \geq 1$ then it is true for all $i \geq 1$. The next proposition shows that even for matrices which are not powers of other matrices, k can be large.

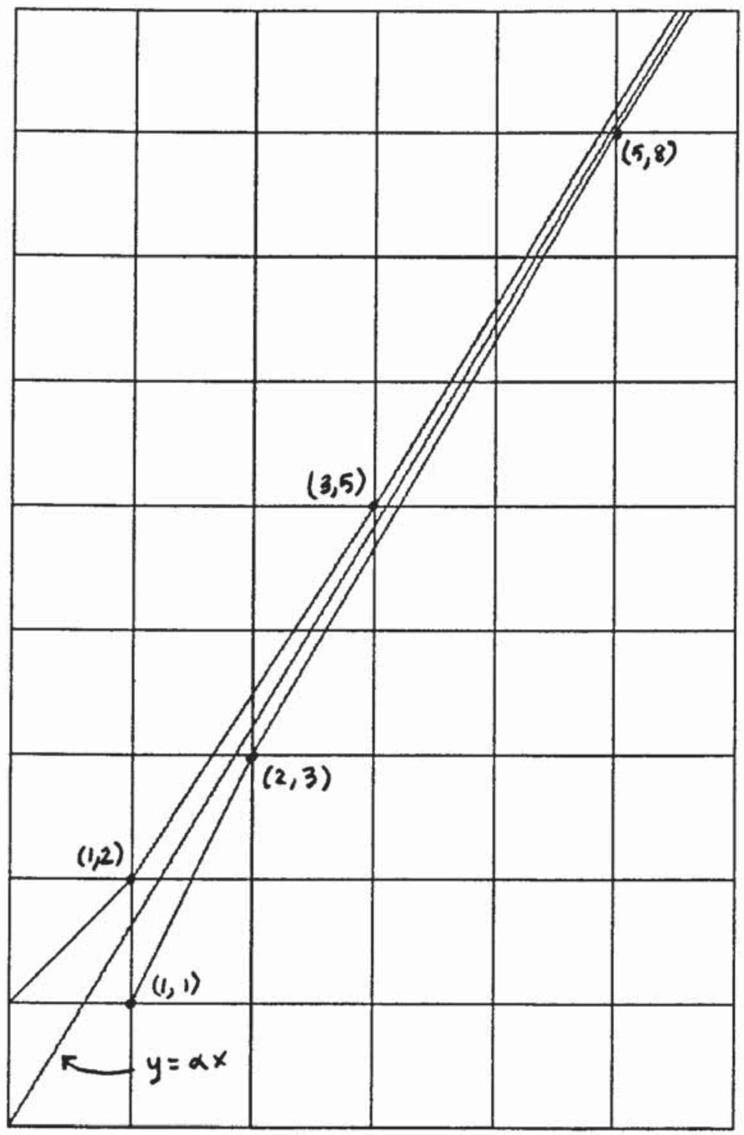

Figure 5.1.10 $\alpha = \frac{1+\sqrt{5}}{2}$

PROPOSITION 5.1.11: Let B_n be as given in Proposition 5.1.6. Then

$$B_n\begin{bmatrix} q_i \\ p_i \end{bmatrix}=\begin{bmatrix} q_{i+n+1} \\ p_{i+n+1} \end{bmatrix}$$

if the p_i and q_i are from the ith convergent of m_u, the slope of the unstable eigenvector of B_n.

Proof: We prove this by direct computation. Recall from Proposition 5.1.6 that $m_u=[\overline{\underbrace{1, 1, \ldots, 1}_{n}, a}]$. From the calculations in Example 5.1.9, we know that

$$\frac{p_i}{q_i}=\frac{F_{i+1}}{F_i} \text{ for } 1\leq i\leq n.$$

We compute the next few convergents:

$$\frac{p_{n+1}}{q_{n+1}}=\frac{a\cdot p_n+p_{n-1}}{a\cdot q_n+q_{n-1}}=\frac{a\cdot F_{n+1}+F_n}{a\cdot F_n+F_{n-1}}$$

$$\frac{p_{n+2}}{q_{n+2}}=\frac{1\cdot p_{n+1}+p_n}{1\cdot q_{n+1}+q_n}=\frac{(a+1)F_{n+1}+F_n}{(a+1)F_n+F_{n-1}}$$

$$\frac{p_{n+3}}{q_{n+3}}=\frac{1\cdot p_{n+2}+p_{n+1}}{1\cdot q_{n+2}+q_{n+1}}=\frac{(2a+1)F_{n+1}+2F_n}{(2a+1)F_n+2F_{n-1}}.$$

We observe that $B_n\begin{bmatrix} q_2 \\ p_2 \end{bmatrix}=B_n\begin{bmatrix} 1 \\ 2 \end{bmatrix}=\begin{bmatrix} (2a+1)F_n+2F_{n-1} \\ (2a+1)F_{n+1}+2F_n \end{bmatrix}=\begin{bmatrix} q_{n+3} \\ p_{n+3} \end{bmatrix}.$ □

SECTION 5.2: THE DEFINING POINT REPRESENTATION

We now introduce one characterization of core-connected partitions. It is worthy to note that this construction was developed to aid in graphing Markov partitions of $\mathbb{T}^2$ on a computer.

Let us first consider the case $\mathcal{C}=\{\text{the origin}\}$. When asked to draw a simple Markov partition for $\mathbb{T}^2$, one usually draws a lines extending out from the origin in the directions of the eigenvectors of the toral automorphism $\mathcal{A}$. They are drawn so that each line has its endpoint in another line so that $\mathbf{T}^2$ is divided up into parallelograms. It is then necessary to show that this construction satisfies all the conditions set forth in Definition 1.1.16. This partition is core-connected and has $\mathcal{C}=\{0\}$. In constructing this partition, the main consideration was that each line had its endpoints in another line. Of course, if any of the eigenvalues of the matrix $\mathcal{A}$ are negative, we must be a bit more careful about the lengths of these lines. We shall explore the above construction a bit further. From now on we will assume that all partitions are core-connected.

If we draw the lines described above in $\mathbf{R}^2$ instead of $\mathbf{T}^2$, we see two line segments which intersect at 0. Figures 5.2.2 and 5.2.3 show a partition and its preimage under $\pi:\mathbf{R}^2\to\mathbf{T}^2$. Notice that we have chosen a single preimage at 0 even though there are preimages at each point in $\mathbb{Z}\times\mathbb{Z}$.

DEFINITION 5.2.1: Let $q\in\mathcal{C}$. Then $\pi^{-1}q\cap[0,1)\times[0,1)$ is exactly one point which we will call the <u>central</u> <u>preimage</u> <u>of</u> <u>q</u> and which we will denote <u>q'</u>. Further, let I be an essential component of $\partial\mathcal{P}$, hence there is a point $q\in$I such that $q\in\mathcal{C}$. We define the <u>central</u> <u>preimage</u> <u>of</u> <u>I</u>, which we denote <u>I$'$</u>, to be the preimage of I which contains q'.

In other words, we want the preimage of I which has the point in I from the core in the region $[0,1)\times[0,1)$. This is the preimage shown in Figures 5.2.2 and 5.2.3.

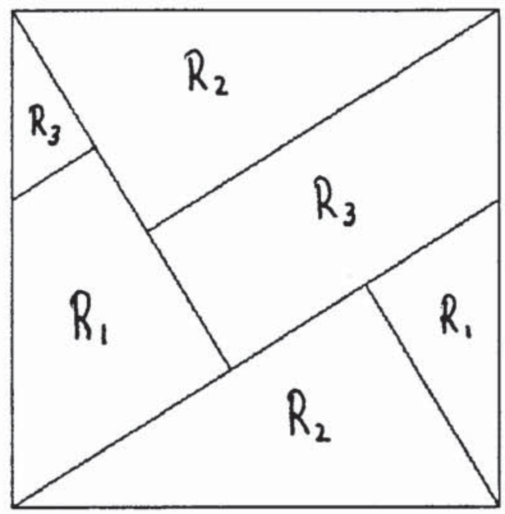

Figure 5.2.2: A Partition $\mathcal{P}$

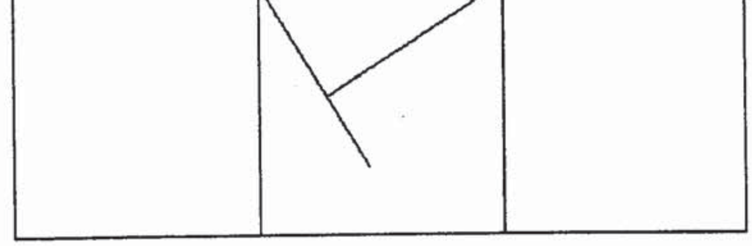

Figure 5.2.3: A Preimage of $\mathcal{P}$

Let us examine Figure 5.2.3 more closely. We see that the length of the unstable boundary is determined by two points in the integer lattice: one which determines how far $\partial_u \mathcal{P}$ extends into the first quadrant (in this example) and one which determines how far $\partial_u \mathcal{P}$ extends into the third quadrant. Similarly, there are two such points for $\partial_s \mathcal{P}$. These pairs of integer points determine the partition. We make this notion more formal.

Let $\mathcal{A}$ be a toral automorphism with eigenvalues and eigenvectors as defined in Section 1.2. Let $q \in \mathcal{C}$ and suppose L_u is the line in $\mathbf{R}^2$ with slope $\vec{v}_u$ through $q' = (x_q, y_q)$. We can write L_u as $L_u(t) = (x_q, y_q) + t\vec{v}_u$. We will assume WLOG that $\vec{v}_u$ is a unit vector in the upper half plane. We can similarly write $L_s(t) = (x_q, y_q) + w\vec{v}_s$ where $\vec{v}_s$ is a unit vector in the upper half plane.

DEFINITION 5.2.4: For $q \in \mathcal{C}$, we define

$$\partial_u^q = W^u(q) \cap \partial_u \mathcal{P}$$

$$\partial_s^q = W^s(q) \cap \partial_s \mathcal{P}.$$

We will refer to the above as the <u>unstable (stable) boundary at *q*</u>. We also define

$$\partial_{u+}^q = \pi(\{L_u(t) | t \geq 0\}) \cap \partial_u \mathcal{P}$$

$$\partial_{u-}^q = \pi(\{L_u(t) | t \leq 0\}) \cap \partial_u \mathcal{P}$$

$$\partial_{s+}^q = \pi(\{L_s(w) | w \geq 0\}) \cap \partial_s \mathcal{P}$$

$$\partial_{s-}^q = \pi(\{L_s(w) | w \leq 0\}) \cap \partial_s \mathcal{P}.$$

We refer to the above as the <u>unstable (stable) positive (negative) boundary at *q*</u>.

Figure 5.2.5 has these entities labeled. If $\mathcal{C}$ consists of a single fixed point, we will omit the superscript q. We will always omit any reference to a specific partition although such is certainly implied.

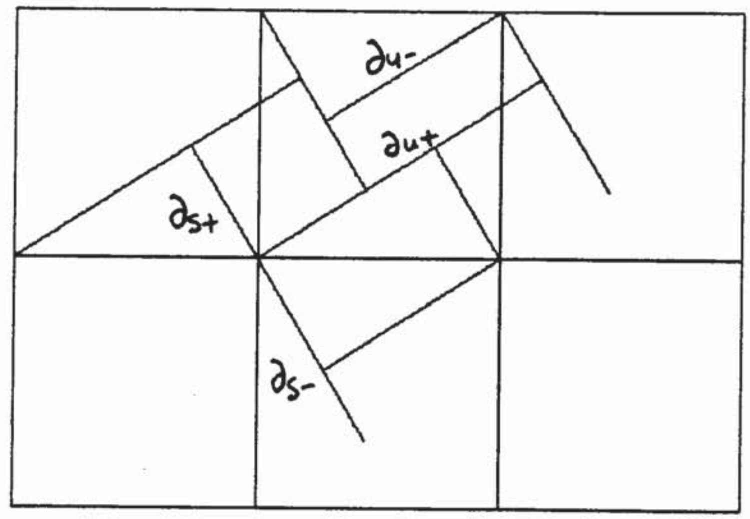

Figure 5.2.5: Boundary Parts Labeled

DEFINITION 5.2.6: The lengths of ∂^q_{u+}, ∂^q_{u-}, ∂^q_{s+}, and ∂^q_{s-} are determined by integer points which we will denote $\underline{p^q_*}$ for $*=\{u+, u-, s+, s-\}$.

DEFINITION 5.2.7: Define

$$t^q_* = (p^q_* - q)\cdot\vec{v}_u \text{ for } *=\{u+, u-, s+, s-\} \text{ and}$$

$$w^q_* = (p^q_* - q)\cdot\vec{v}_s \text{ for } *=\{u+, u-, s+, s-\}.$$

where we consider $(p^q_* - q)$ as a vector and use the standard dot product.

Figure 5.2.8 has all of these labeled. Notice that $|t^q_{u+}| = \text{length}(\partial^q_{u+})$, $|t^q_{u-}| = \text{length}(\partial^q_{u-})$, $|w^q_{s+}| = \text{length}(\partial^q_{s+})$, and $|w^q_{s-}| = \text{length}(\partial^q_{s-})$.

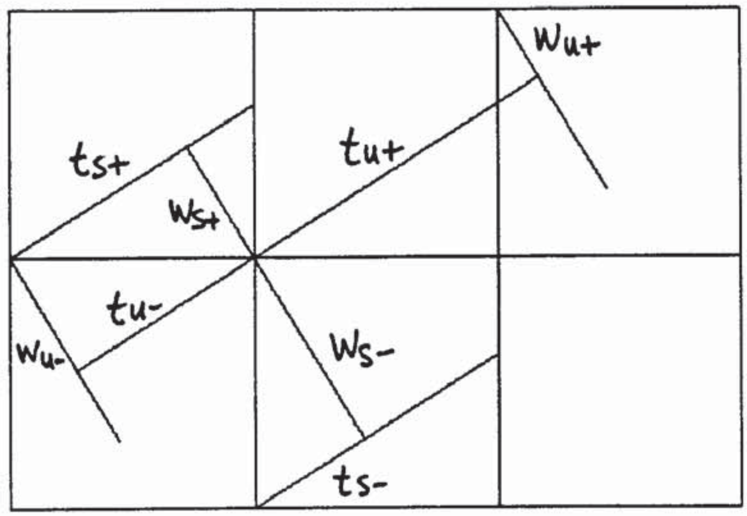

Figure 5.2.8: t$_*$ and w$_*$ Labeled

If we are given a set of two pairs of integer points, when do they actually define a partition? In order for ∂_{u+} to have its endpoints in other line segments, the length of the line from p_{u+} to ∂_{u+} must be shorter than the length of ∂_{s+} if p_{u+} is below ∂_{u+} or shorter than the length of ∂_{s-} if p_{u+} is above ∂_{u+}. Further, if $\lambda_s<0$, then $\mathcal{A}(\partial_{s+})$ must be contained in ∂_{s-} and similarly for $\mathcal{A}(\partial_{s-})$ and the unstable boundary components if $\lambda_u<0$. We make this precise in the following proposition.

PROPOSITION 5.2.9: Let $\mathcal{A}$ be a hyperbolic toral automorphism. Given two pairs of integer points $\{\{p_{u+}, p_{u-}\}, \{p_{s+}, p_{s-}\}\}$, they define a core-connected Markov partition with $\mathfrak{C}=\{0\}$ if and only if all of the below are satisfied:

i) p_* is on the integer lattice for $* = \{u+, u-, s+, s-\}$.

ii) t_{u+}, $w_{s+} \geq 0$, t_{u-}, $w_{s-} \leq 0$ with at most one of them equal to 0.

iii) If $\lambda_u < 0$, then $|\lambda_u t_{u+}| \geq |t_{u-}|$ and $|\lambda_u t_{u-}| \geq |t_{u+}|$.

If $\lambda_s < 0$, then $|\lambda_s t_{s+}| \leq |t_{s-}|$ and $|\lambda_s t_{s-}| \leq |t_{s+}|$.

iv) If $w_{u+} \leq 0$, then $|w_{s-}| \geq |w_{u+}|$; if $w_{u+} \geq 0$, then $|w_{s+}| \geq |w_{u+}|$.

If $w_{u-} \leq 0$, then $|w_{s-}| \geq |w_{u-}|$; if $w_{u-} \geq 0$, then $|w_{s+}| \geq |w_{u-}|$.

If $t_{s+} \leq 0$, then $|t_{u-}| \geq |t_{s+}|$; if $t_{s+} \geq 0$, then $|t_{u+}| \geq |t_{s+}|$.

If $t_{s-} \leq 0$, then $|t_{u-}| \geq |t_{s-}|$; if $t_{s-} \geq 0$, then $|t_{u+}| \geq |t_{s-}|$.

Proof: Condition *i)* is clearly necessary since each defining point must be equivalent to 0 when mapped into $\mathbf{T}^2$. Condition *ii)* insures that our points make sense in that p_{u+} actually defines ∂_{u+} etc. Condition *iii)* insures that the image of $\partial_u \mathcal{P}$ covers itself under $\mathcal{A}$ and the image of $\partial_s \mathcal{P}$ is mapped by $\mathcal{A}$ into itself. Condition *iv)* makes sure that we actually have rectangles and that there is no "overhang". If any of these is not satisfied, then we do not have a partition by the argument above. If they are all satisfied, we do have a partition. □

We now suppose that $\mathcal{C} = \{$a finite number of periodic orbits$\}$. At each point in $\mathcal{C}$, we have a similar situation as above. The only difference is that each defining points now must be equivalent to some point in $\mathcal{C}$ and that ∂^{α}_{u+} must cover ∂^{β}_{u+} if $\mathcal{A}(\alpha) = \beta$ and $\lambda_u > 0$, and so forth.

PROPOSITION 5.2.10: Let $\mathcal{A}$ be a hyperbolic toral automorphism and let $B = \{$a finite number of periodic orbits$\}$. Given two pairs of points $\{\{p^{\alpha}_{u+}, p^{\alpha}_{u-}\}, \{p^{\alpha}_{s+}, p^{\alpha}_{s-}\}\}$ for each $\alpha \in B$, they define a core-connected Markov partition with $\mathcal{C} = B$ if and only if all of the below are satisfied:

i) $\pi p_*^\alpha \in \mathcal{C}$ $\forall \alpha \in \mathcal{C}$ and $* = \{$u+, u-, s+, s-$\}$.

ii) t_{u+}^α, $w_{s+}^\alpha \geq 0$, t_{u-}^α, $w_{s-}^\alpha \leq 0$ for each α and

a) at most one is zero for every α

b) at most two are zero for any fixed α in which case the two which are zero are both stable or both unstable.

iii) If $\lambda_u > 0$, then $\left|\lambda_u t_{u+}^\alpha\right| \geq \left|t_{u+}^\beta\right|$ and $\left|\lambda_u t_{u-}^\alpha\right| \geq \left|t_{u-}^\beta\right|$ if $\mathcal{A}(\alpha) = \beta$.

If $\lambda_s > 0$, then $\left|\lambda_s t_{s+}^\alpha\right| \leq \left|t_{s+}^\beta\right|$ and $\left|\lambda_s t_{s-}^\alpha\right| \leq \left|t_{s-}^\beta\right|$ if $\mathcal{A}(\alpha) = \beta$.

If $\lambda_u < 0$, then $\left|\lambda_u t_{u+}^\alpha\right| \geq \left|t_{u-}^\beta\right|$ and $\left|\lambda_u t_{u-}^\alpha\right| \geq \left|t_{u+}^\beta\right|$ if $\mathcal{A}(\alpha) = \beta$.

If $\lambda_s < 0$, then $\left|\lambda_s t_{s+}^\alpha\right| \leq \left|t_{s-}^\beta\right|$ and $\left|\lambda_s t_{s-}^\alpha\right| \leq \left|t_{s+}^\beta\right|$ if $\mathcal{A}(\alpha) = \beta$.

iv) If $\pi p_{u+}^\alpha = \beta$ then

if $w_{u+}^\alpha \leq 0$, then $\left|w_{s-}^\beta\right| \geq \left|w_{u+}^\alpha\right|$; if $w_{u+}^\alpha \geq 0$, then $\left|w_{s+}^\beta\right| \geq \left|w_{u+}^\alpha\right|$.

If $\pi p_{u-}^\alpha = \beta$ then

if $w_{u-}^\alpha \leq 0$, then $\left|w_{s-}^\beta\right| \geq \left|w_{u-}^\alpha\right|$; if $w_{u-}^\alpha \geq 0$, then $\left|w_{s+}^\beta\right| \geq \left|w_{u-}^\alpha\right|$.

If $\pi p_{s+}^\alpha = \beta$ then

if $t_{s+}^\alpha \leq 0$, then $\left|t_{u-}^\beta\right| \geq \left|t_{s+}^\alpha\right|$; if $t_{s+}^\alpha \geq 0$, then $\left|t_{u+}^\beta\right| \geq \left|t_{s+}^\alpha\right|$.

If $\pi p_{s-}^\alpha = \beta$ then

if $t_{s-}^\alpha \leq 0$, then $\left|t_{u-}^\beta\right| \geq \left|t_{s-}^\alpha\right|$; if $t_{s-}^\alpha \geq 0$, then $\left|t_{u+}^\beta\right| \geq \left|t_{s-}^\alpha\right|$.

Proof: The reasoning is the same as in the previous proposition. □

SECTION 5.3: THE GENERATING SET $\mathfrak{G}$

In this section we prove the existence of a finite set of partitions with two rectangles all having the same core which generate all other two-rectangle partitions with that particular core. With this, we will have a finite set of partitions which generates all partitions with two rectangles.

We will assume throughout this section that the unstable eigenvector of $\mathcal{A}$ is in the first quadrant. If this is not so and $\mathcal{A}$ has its stable eigenvector in the first quadrant, we work with $\mathcal{A}^{-1}$ since by Proposition 1.4.2 any Markov partition for $\mathcal{A}^{-1}$ is also a partition for $\mathcal{A}$. If $\mathcal{A}$ has both stable and unstable eigenvectors in the second quadrant, the $\mathcal{A}$ is similar over $\mathbb{Z}$ to either a non-negative or non-positive matrix by Corollary 3.2.8 and this class of matrices has the desired unstable eigenvector by the Perron-Frobenius Theorem [10]. Further, by Proposition 1.4.3 there is a one-to-one correspondence between the partitions for $\mathcal{A}$ and the partitions for $\mathcal{B}$ if $\mathcal{A}$ and $\mathcal{B}$ are similar over $\mathbb{Z}$.

First, we examine the convergents to the slope of $\vec{v}_u$ which for now we will call α, as these arguments apply to all real numbers. In some cases, there are other fractional (reduced) approximations $\frac{p}{q}>\alpha$ such that $\left|\alpha-\frac{p}{q}\right|<\left|\alpha-\frac{p_i}{q_i}\right|$ but $\left|\alpha-\frac{p}{q}\right|>\left|\alpha-\frac{p_{i+2}}{q_{i+2}}\right|$ where $\frac{p_i}{q_i}$ is an upper convergent to α and $q_i<q<q_{i+1}$ yet $\frac{p}{q}$ is not a convergent. For example, if $\alpha=\frac{\sqrt{15}-1}{7}$, the first few convergents are $\frac{1}{2}$, $\frac{2}{5}$, $\frac{7}{17}$, $\frac{16}{39}$, and $\frac{55}{134}$. The first, third, and fifth of these are upper convergents and the second and fourth are lower convergents. However, we notice that $\frac{2}{5}<\frac{9}{22}<\frac{16}{39}$ and $\frac{55}{134}<\frac{39}{95}<\frac{23}{56}<\frac{7}{17}<\frac{5}{12}<\frac{3}{7}<\frac{1}{2}$. We make the following definition.

DEFINITION 5.3.1: Let $\frac{p_i}{q_i}$ be the convergents to an irrational number α. If there rational numbers (reduced) $\frac{p_{i,1}}{q_{i,1}}, \frac{p_{i,2}}{q_{i,2}},\ldots, \frac{p_{i,j}}{q_{i,j}}$ such that

i) $q_{i,1}<q_{i,2}<\ldots<q_{i,j}$ for each i and

ii) $\frac{p_i}{q_i}=\frac{p_{i,1}}{q_{i,1}}>\frac{p_{i,2}}{q_{i,2}}>\ldots>\frac{p_{i,j}}{q_{i,j}}=\frac{p_{i+2}}{q_{i+2}}>\alpha$

then we call $\frac{p_{i,1}}{q_{i,1}}, \frac{p_{i,2}}{q_{i,2}},\ldots, \frac{p_{i,j}}{q_{i,j}}$ the <u>ith</u> <u>upper</u> <u>pseudo-convergents</u> <u>to</u> <u>α</u>.

If there rational numbers (reduced) $\frac{p_{i,1}}{q_{i,1}}, \frac{p_{i,2}}{q_{i,2}}, \ldots, \frac{p_{i,j}}{q_{i,j}}$ such that

$i)$ $q_{i,1} < q_{i,2} < \ldots < q_{i,j}$ for each i and

$$ii)\ \frac{p_i}{q_i} = \frac{p_{i,1}}{q_{i,1}} < \frac{p_{i,2}}{q_{i,2}} < \ldots < \frac{p_{i,j}}{q_{i,j}} = \frac{p_{i+2}}{q_{i+2}} < \alpha$$

then we call $\frac{p_{i,1}}{q_{i,1}}, \frac{p_{i,2}}{q_{i,2}}, \ldots, \frac{p_{i,j}}{q_{i,j}}$ the <u>ith</u> <u>lower</u> <u>pseudo-convergents</u> <u>to</u> <u>α</u>.

Note that convergents are pseudo-convergents. Also note that if $\mathcal{A}$ has α as the slope of its unstable eigenvector, pseudo-convergents map to pseudo-convergents under $\mathcal{A}$. We already know that convergents map to convergents. Hence, the number of upper pseudo-convergents between c_i and c_{i+2} is constant for all upper convergents c_i and similarly for lower convergents.

EXAMPLE 5.3.2: For the matrix $\begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}$, $m_u = \frac{\sqrt{15}-1}{7}$. If we look at the pseudo-convergents for m_u, we see

$$\begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 17 \\ 7 \end{bmatrix};\ \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 17 \\ 7 \end{bmatrix} = \begin{bmatrix} 134 \\ 55 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 56 \\ 23 \end{bmatrix};\ \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 12 \\ 5 \end{bmatrix} = \begin{bmatrix} 95 \\ 39 \end{bmatrix}.$$

We observe that convergents do indeed map to convergents and pseudo-convergents to pseudo-convergents.

The best way to compute the ith pseudo-convergents to a number α uses the Farey sequences.

DEFINITION 5.3.3: Let $\mathcal{F}_n$ be the (ordered) sequence of all rational numbers in reduced form between 0 and 1 inclusive with denominator less than or equal to n. We call $\mathcal{F}_n$ the <u>nth</u> <u>Farey sequence</u>.

EXAMPLE 5.3.4: We compute that

$$\mathcal{F}_1=\{\tfrac{0}{1},\tfrac{1}{1}\};\quad \mathcal{F}_2=\{\tfrac{0}{1},\tfrac{1}{2},\tfrac{1}{1}\};$$

$$\mathcal{F}_3=\{\tfrac{0}{1},\tfrac{1}{3},\tfrac{1}{2},\tfrac{2}{3},\tfrac{1}{1}\};$$

$$\mathcal{F}_5=\{\tfrac{0}{1},\tfrac{1}{5},\tfrac{1}{4},\tfrac{1}{3},\tfrac{2}{5},\tfrac{1}{2},\tfrac{3}{5},\tfrac{2}{3},\tfrac{3}{4},\tfrac{4}{5},\tfrac{1}{1}\}.$$

To compute the *i*th pseudo-convergents for a given irrational α, we look at $\mathcal{F}_{q_{i+2}}$. Suppose c_i is an upper convergent and we are computing the upper pseudo-convergents. We look at the numbers between c_i and c_{i+2} in $\mathcal{F}_{q_{i+2}}$. If there are none, there are no upper pseudo-convergents. If there are, find the one with the lowest denominator, call it ξ_1. It is a pseudo-convergent. Now look between ξ_1 and c_{i+2}. If there are no numbers, we have found all the *i*th upper pseudo-convergents. If there are numbers, again find the one with the lowest denominator. It is a pseudo-convergent, call it ξ_2. Continue this process until ξ_k and c_{i+2} are adjacent in $\mathcal{F}_{q_{i+2}}$ and we have found all the *i*th upper pseudo-convergents. A similar process will find all the *i*th lower pseudo-convergents.

DEFINITION 5.3.5: We combine the convergents and pseudo-convergents into one sequence, $\underline{\Gamma}$, where

$$\Gamma=\{c_i \text{ where } c_i \text{ is a convergent or a pseudo-convergent and } |\alpha-c_i|>|\alpha-c_{i+1}|\}.$$

Further, define

$$\Gamma_u=\{c_i \text{ where } c_i\in\Gamma \text{ and } c_i>\alpha\} \text{ and}$$

$$\Gamma_l=\{c_i \text{ where } c_i\in\Gamma \text{ and } c_i<\alpha\}.$$

We now apply this to Markov partitions. We will identify pseudo-convergents with points on the integer lattice as we did previously. For now, we will assume that $\mathfrak{C}=\{0\}$ for our partitions and that $\lambda_u>0$ so that we can have the length of ∂_{u-} equal to zero. Suppose we choose a length l_u for ∂_{u+} by choosing p_{u+} so that ∂_{u+} is long. If p_{u+} is not a pseudo-convergent, then there is a pseudo-convergent ξ such that a line segment of length l_u from ξ in the direction of $\vec{v}_u$ intersects the line segment from p_{u+} in the direction of $\vec{v}_s$ to the line $y=m_u x$. This implies that in a partition with the length of ∂_{u+} equal to l_u, there would be a crossing and by the proof of Lemma 3.1.3 such a partition would have more than two rectangles. Figure 5.3.6 shows why (2,2) can not be a defining point for a partition with two rectangles for $\left[\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right]$. Hence, in order to have a partition with two rectangles and our above assumptions, p_{u+} must be a pseudo-convergent.

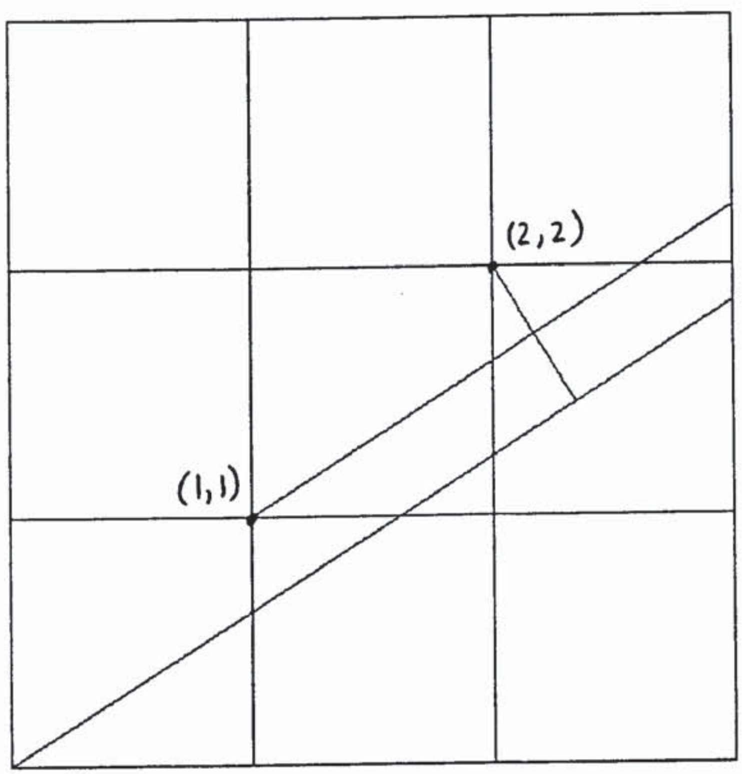

Figure 5.3.6: Why (2,2) Can Not Be a Defining Point

Conversely, suppose that p_{u+} is an upper pseudo-convergent. Let ξ_1 be the largest lower pseudo-convergent which precedes p_{u+} in Γ; hence $\left|w_{s+}^{\xi_1}\right|>\left|w_{s-}^{p_{u+}}\right|$. Let ξ_2 be the upper pseudo-convergent which precedes p_{u+} in Γ_u; hence $\left|w_{s-}^{\xi_2}\right|>\left|w_{s-}^{p_{u+}}\right|$. If p_{u+} is a lower pseudo-convergent, choose the lower pseudo-convergent which precedes p_{u+} in Γ_l and the smallest upper pseudo-convergent which precedes p_{u+} in Γ. We would then replace $w_{s-}^{p_{u+}}$ with $w_{s+}^{p_{u+}}$ in the above inequalities. Let $p_{s+}=-\xi_1$ and $p_{s-}=-\xi_2$. The reason we make these choices is so that we have no crossings and hence two rectangles. If we chose a defining point which precedes either ξ_1 or ξ_2 in Γ, then the stable boundary from one of them would intersect the line $y=m_u x$ and we would have a crossing. If we chose a point which followed p_{u+} in Γ, then we would not have the endpoints of ∂_s in ∂_u. The reason we take minus these quantities is that we want ∂_{s+} to be the length of the line segment from ξ_1 to $y=m_u x$ and so forth. Letting $p_{u-}=0$, the points $\{p_{u+}, p_{u-}, p_{s+}, p_{s-}\}$ are defining points for a Markov partition in that they satisfy the conditions set forth in Proposition 5.2.9. They satisfy these conditions by construction. Because we had no choice when selecting the other defining points, the length of ∂_{u+} defines every partition with two rectangles, $\mathfrak{C}=\{0\}$, and the length of ∂_{u-} equal to zero.

Once again we suppose that $\mathcal{A}$ is a hyperbolic toral automorphism and that $\vec{v}_u$ has pseudo-convergents as defined above. For convenience, we will assume that i is a large enough number so that stable lines from c_i intersect $y=\alpha x$ closer to the origin than stable lines from c_{i+1}. This is not necessarily the case for small values of i and eigenvectors far from perpendicular. Because these two-rectangle partitions are defined by the length of ∂_{u+}, if $c_i \overset{\mathcal{A}}{\mapsto} c_{i+k}$ then the partition defined by c_i will be mapped to the partition defined by c_{i+k}. This observation leads us to the following proposition.

PROPOSITION 5.3.7: Let $\mathcal{M}_u$ denote the set of two rectangle Markov partitions for $\mathcal{A}$ with the length of ∂_{u-} equal to zero and $\mathcal{C}=\{0\}$. There is a finite set $\mathcal{G}_u$ of partitions in $\mathcal{M}_u$ such that any partition in $\mathcal{M}_u$ is the image of an element of $\mathcal{G}_u$ under some power of $\mathcal{A}$.

Proof: Let i be large and suppose that $c_i \overset{\mathcal{A}}{\mapsto} c_{i+k}$. Let $\mathcal{P}_i$ be the partition defined by c_i. Let $\mathcal{G}_u=\{\mathcal{P}_j\}_{j=i}^{i+k-1}$. We claim that $\mathcal{G}_u$ is such a generating set. Consider $\mathcal{P}_n$ with $n \geq i+k$. We can write $n=mi+j$ with $m \geq 1$ and $0 \leq j \leq k-1$. Hence, $\mathcal{P}_n$ is the image of $\mathcal{P}_{i+j}$ under m iterates of $\mathcal{A}$. Therefore $\mathcal{G}_u$ generates all partitions with unstable positive length larger than that of $\mathcal{P}_{i+k-1}$. Suppose then that $\mathcal{P}_n$ has positive unstable length less that that of $\mathcal{P}_i$. Under enough iterates of $\mathcal{A}$, the unstable positive length of $\mathcal{A}^l\mathcal{P}_n$ will be larger than that of $\mathcal{P}_{i+k-1}$ hence $\mathcal{A}^l\mathcal{P}_n=\mathcal{A}^m\mathcal{P}_{i+j}$ for $0 \leq j \leq k-1$ and we are done. □

The set $\mathcal{G}_u$ can be quite large. Consider Propositions 5.1.6 and 5.1.11. If n is large, we have the ith convergent mapped to the $(i+n+1)$st convergent. Therefore, $\mathcal{G}_u$ contains n partitions.

Because $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\mathcal{A}\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}=\mathcal{A}$, by Proposition 1.4.3 $-\mathcal{P}$ is always a partition for $\mathcal{A}$. Therefore, by the previous proposition, we have shown that $\mathcal{G}_u$ as defined above generates all partitions with either the length of ∂_{u+} or the length of ∂_{u-} equal to zero under powers of $\mathcal{A}$ and $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. The question arises as to when $\mathcal{G}_u$ actually generates all partitions with two rectangles and $\mathcal{C}=\{0\}$. This is equivalent to asking whether or not we can map the unstable manifold of $\mathcal{A}$ to the stable manifold of $\mathcal{A}$ with an element of GL(2.$\mathbb{Z}$). The way to visualize this if the eigenvectors are perpendicular is to ask whether or not we can rotate the partition by an angle of radian measure $\frac{\pi}{2}$. If $\lambda_s<0$, then we cannot rotate in this fashion and still have a Markov partition. In our rotated partition, the former unstable boundary is now the stable boundary. Because length$(\partial_{u-})=0$ in the former partition, length$(\partial_{s-})=0$ in the new partition and $\mathcal{A}\partial_{s-}$ must cover ∂_{s+} which is impossible. However, if $\lambda_s>0$, we can rotate in this fashion in some cases. Being able to rotate in this way means that we are mapping a partition for $\mathcal{A}$ to a partition for $\mathcal{A}^{-1}$ while maintaining the identity of the unstable and stable directions. (The unstable eigenvector for $\mathcal{A}$ is the stable eigenvector for $\mathcal{A}^{-1}$.) In other words, we are searching for a $\phi\in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi=\mathcal{A}^{-1}$. The following proposition uses this notion. It is sufficient to prove the proposition for non-negative matrices since each element of $GL(2,\mathbb{Z})$ is similar over $\mathbb{Z}$ to either a non-negative matrix or a non-positive matrix.

PROPOSITION 5.3.8: For $\mathcal{A}\geq 0$, there exists a matrix $\phi\in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi=\mathcal{A}^{-1}$ $\Leftrightarrow$ $\mathcal{A}$ is shift equivalent over $\mathbb{Z}$ to $\mathcal{A}^{T}$.

Proof: ($\Rightarrow$) Suppose $\exists$ $\phi\in GL(2,\mathbb{Z})$ such that $\phi^{-1}\mathcal{A}\phi=\mathcal{A}^{-1}$. Let $\mathcal{P}$ be a partition for $\mathcal{A}$ with Markov matrix $\mathcal{A}$. Then by Proposition 1.4.3, $\phi^{-1}\mathcal{P}$ is a partition for $\mathcal{A}^{-1}$ with Markov matrix $\mathcal{A}$ $\Rightarrow$ by Proposition 1.4.2 $\phi^{-1}\mathcal{P}$ is a partition for $\mathcal{A}$ with Markov matrix $\mathcal{A}^{T}$ $\Rightarrow$ $\exists$ $\psi\in GL(2,\mathbb{Z})$ such that $\psi^{-1}\mathcal{A}\psi=\mathcal{A}^{T}$ $\Rightarrow$ by Proposition 4.1.9 and its proof that $\mathcal{A}$ and $\mathcal{A}^{T}$ are shift equivalent over $\mathbb{Z}$.

($\Leftarrow$) Suppose $\mathcal{A}$ is shift equivalent over $\mathbb{Z}$ to $\mathcal{A}^{T}$. By Proposition 4.1.9 and its proof, $\exists\psi\in GL(2,\mathbb{Z})$ such that $\psi^{-1}\mathcal{A}\psi=\mathcal{A}^{T}$. Let $\mathcal{P}$ be a partition for $\mathcal{A}$ with Markov matrix $\mathcal{A}$. $\psi^{-1}\mathcal{P}$ is a partition for $\mathcal{A}^{T}$ with Markov matrix $\mathcal{A}$ $\Rightarrow$ $\psi^{-1}\mathcal{P}$ is a partition for $(\mathcal{A}^{T})^{-1}=(\mathcal{A}^{-1})^{T}$ with matrix $\mathcal{A}^{T}$ $\Rightarrow$ $\exists\phi\in GL(2,\mathbb{Z})$ such that $\phi^{-1}(\mathcal{A}^{-1})^{T}\phi=\mathcal{A}^{T}$ $\Rightarrow$ $\left((\phi^{T})\mathcal{A}^{-1}(\phi^{-1})^{T}\right)^{T}=\mathcal{A}^{T}$ $\Rightarrow$ $\phi^{T}\mathcal{A}^{-1}(\phi^{T})^{-1}=\mathcal{A}$. $\square$

What we have now shown is that if the hypotheses of the above proposition are met, $\mathcal{G}_u$ generates all two-rectangle partitions with $\mathcal{C}=\{0\}$ under elements of $GL(2,\mathbb{Z})$. If the hypotheses are not met, we use the construction in Proposition 5.3.7 on $\vec{v}_s$ to find an analogous set $\mathcal{G}_s$. (If $\vec{v}_s$ is not in the first quadrant, rotate by $\frac{-\pi}{2}$ so that it is easier to find the "closest" integer lattice points then rotate back.) $\mathcal{G}_s$ will be a generating set for all partitions with length(∂_{s-}) or length(∂_{s+}) equal to zero. We have proven the following:

PROPOSITION 5.3.9: Let $\mathcal{M}_1$ be the set of all two rectangle Markov partitions for $\mathcal{A}$ with $\mathcal{C}=\{$a single fixed point$\}$. There is a finite set of partitions $\mathcal{G}$ such that any partition in $\mathcal{M}_1$ is the image of an element of $\mathcal{G}$ under an element of $GL(2,\mathbb{Z})$ which commutes with $\mathcal{A}$.

Proof: By the above arguments, if $\mathcal{C}=\{0\}$, let $\mathcal{G}=\mathcal{G}_u\cup\mathcal{G}_s$ and we are done. If $\mathcal{P}$ is a partition with $\mathcal{C}$ consisting of any other fixed point of $\mathcal{A}$, $\mathcal{P}$ is a translation of a partition with $\mathcal{C}=\{0\}$. Since $\mathcal{A}$ has a finite number of fixed points, let

$$\mathcal{G}=\bigcup_{\mathrm{Fix}(\mathcal{A})}\{\mathcal{G}_u\cup\mathcal{G}_s\}$$

and we are done. □

We have completed the case when $\mathcal{C}=\{0\}$. We know from Proposition 3.3.9 that the only other way we can have two rectangles is to have two fixed points in $\mathcal{C}$ with $\mathcal{U}\cap\mathcal{S}=\emptyset$. The procedure we will use in this case is essentially the same as in the $\mathcal{C}=\{0\}$ case except for the fact that our lines do not necessarily pass through the origin hence our computations will not be aided by the use of pseudo-convergents. In the case $\mathcal{C}=\{0\}$, a length of unstable positive boundary determined a partition. In the case where $\mathcal{C}=\{q_1, q_2\}$, we will still use a length of unstable positive boundary as a reference point but now it will define a finite number of partitions. WLOG, we assume that $\mathcal{U}=q_1$ and $\mathcal{S}=q_2$. We now develop some ideas which will aid us in our computations. We will abuse notation and refer to $q_i \in \mathbf{T}^2$ and $\pi^{-1} q_i \in [0,1) \times [0,1) \subset \mathbf{R}^2$ as q_i.

Define Δ' to be $\Delta' = \{$all points (a,b) in $\mathbf{Q} \times \mathbf{Q}$ such that $\pi(a,b) = q_2\}$. For $c \in \Delta'$, define $\hat{c}$ to be the point on the line $l(t) = q_1 + t\vec{v}_u$ closest to c. We can write $\hat{c}$ as $\hat{c} = q_1 + t_{\hat{c}} \vec{v}_u$.

DEFINITION 5.3.10: Define Δ as

$\Delta = \{$the sequence of all $c \in \Delta'$ such that the rectangle R with diagonal corners q_1 and c satisfies $\mathrm{int}(R) \cap \Delta' = \emptyset$. We order Δ by the following rule: $t_{\hat{c}_i} < t_{\hat{c}_{i+1}}$ for all i.$\}$

Further, define

$$\Delta_u = \{c_i \in \Delta \text{ such that } c_i \text{ is above the line } l(t) = q_1 + t\vec{v}_u\} \text{ and}$$

$$\Delta_l = \{c_i \in \Delta \text{ such that } c_i \text{ is below the line } l(t) = q_1 + t\vec{v}_u\}.$$

An equivalent definition for the pseudo-convergents (in the geometric sense) of a real number α is $\{$all the points p on the integer lattice such that the rectangle R with diagonal corners **0** and p contains no integer lattice points in its interior$\}$. This is the motivation for the above definition. The points in Δ behave in the same way as pseudo-convergents under the action of $\mathcal{A}$; that is, $\mathcal{A}: \Delta \to \Delta$ for points away from the origin. Near the origin, particularly in the square $[0,1] \times [0,1]$, we may need to translate the image of a point in Δ across the x or y-axis. Also, defining points for ∂_{u+} in the two rectangle case must be in Δ for reasons exactly analogous to those for pseudo-convergents.

PROPOSITION 5.3.11: Let $\mathcal{M}_2$ be the set of all Markov partitions for $\mathcal{A}$ with two rectangles and $\mathcal{U}=\{q_1\}$ and $\mathcal{S}=\{q_2\}$ with $q_1 \neq q_2$. There is a finite set $\mathcal{G}$ of partitions in $\mathcal{M}_2$ such that any partition in $\mathcal{M}_2$ is the image of an element of $\mathcal{G}$ under a power of $\mathcal{A}$.

Proof: Suppose we choose $p_{u+}^{q_1} = c_j \in \Delta$ so that ∂_{u+} is long and suppose $\lambda_u > 0$. We know that $\pi p_{u+}^{q_1} \in \mathcal{S}$ which means that $p_{u+}^{q_1}$ is equivalent modulo one to q_2. Extend lines in the direction of $\vec{v}_s$ and $-\vec{v}_s$ from q_2 (in $\mathbf{T}^2$) until both endpoints of this line lie in ∂_{u+} and no crossings. From this we can calculate $p_{s+}^{q_2}$ and $p_{s-}^{q_2}$. Extend the line from q_1 to make ∂_{u-} with no crossings and we have a partition, call it $\mathcal{P}_{j,1}$. It is possible now that we can extend ∂_{u-} and shorten ∂_s in exactly the manner of Proposition 4.2.1 to obtain another partition $\mathcal{P}_{j,2}$ which also has $p_{u+}^{q_1}$ as before. We can continue this process only a finite number of times because eventually when extending ∂_s we will intersect ∂_u, have a crossing, and more than two rectangles. This process generates partitions $\mathcal{P}_{j,i}$ for $1 \leq i \leq n_1$. The image of $\mathcal{P}_{j,i}$ is a partition $\mathcal{P}_{j+k,i}$ for $1 \leq i \leq n_1$. We repeat the above process for $\mathcal{P}_{j+m,i}$, which has $p_{u+}^{q_1} = c_{j+m}$, for $1 \leq i \leq n_{j+m}$ and $0 \leq m \leq k-1$. If $\lambda_u < 0$, then $\mathcal{A}^2(\mathcal{P}_{j,i})$ is a partition $\mathcal{P}_{j+k,i}$ which we are certain has positive unstable length greater than that of $\mathcal{P}_{j,i}$. We take $\mathcal{A}^2$ of the partition because we are characterizing a partition by its unstable positive length and $\mathcal{A}^2$ maps ∂_{u+} across ∂_{u+}. So if $\lambda_u < 0$, we set $\mathcal{P}_{j+k,i} = \mathcal{A}^2(\mathcal{P}_{j,i})$ and do exactly what we did above. (Note: If $\lambda_u < 0$, some of these partitions $\mathcal{P}_{j,i}$ may not be Markov; namely, $\mathcal{A}(\partial_u)$ might not cover ∂_u. In this case, any image of $\mathcal{P}_{j,i}$ will not be Markov either so we discard $\mathcal{P}_{j,i}$.) By an argument exactly analogous to that for the $\mathcal{C}=\{0\}$ case, we see that

$$\mathcal{G} = \bigcup_{\alpha=j}^{j+m-1} \left\{ \bigcup_{\beta=1}^{n_\alpha} \mathcal{P}_{\alpha,\beta} \right\}$$

generates all partitions with this specific $\mathcal{U}$ and $\mathcal{S}$. □

We have proven the following theorem.

THEOREM 5.3.12: Let $\mathcal{M}$ denote the set of Markov partitions with two rectangles. There exists a finite set $\mathcal{G} \subset \mathcal{M}$ such that any element of $\mathcal{M}$ is the image of an element of $\mathcal{G}$ under a matrix in $GL(2,\mathbb{Z})$ which commutes with $\mathcal{A}$.

Proof: Proposition 5.3.9 tells us that a finite set generates all partitions with $\mathcal{C}=\{$a single fixed point$\}$. Proposition 5.3.11 tells us that a finite set generates all partitions with $\mathcal{C}=\{$two fixed points$\}$. Because $\mathcal{A}$ has a finite number of fixed points there are a finite number of possible core combinations hence a finite set will generate all partitions with two fixed points in the core. The union of these two sets generates all partitions with two rectangles. □

CHAPTER 6 - CONCLUSIONS

SECTION 6.1: SUMMARY

We study the Markov matrix M as defined in this paper to determine which SFT's are semiconjugate to the system $(\mathbf{T}^2, \mathcal{A}, \mathcal{P})$. If we assume FCC, we can say quite a bit about M. We have established necessary conditions on the eigenvalues of M and in the two rectangle case we know precisely which SFT's are semiconjugate to our system; namely, those which, when represented as an edge shift with two nodes, have transition matrix similar over to $\mathbb{Z}$ to $\mathcal{A}$. Another way of describing this set of matrices is the strong shift equivalence class (2×2 matrices) of $\mathcal{A}$ if $\mathcal{A}\geq0$ or of $\pm(\phi^{-1}\mathcal{A}\phi)\geq0$ ($\phi\in GL(2,\mathbb{Z})$) if $\mathcal{A}$ is not non-negative. It can be shown that if $M\geq0$ is any $n\times n$ integer matrix which is strong shift equivalent to $\mathcal{A}\geq0$ (or $\phi^{-1}\mathcal{A}\phi\geq0$) then there exists an FCC partition for $\mathcal{A}$ with matrix M [11]. Hence Conjecture 4.3.1 is equivalent to saying that the conditions given describe strong shift equivalence classes. As seen in this paper, there are Markov matrices which are not strong shift equivalent to $\mathcal{A}$ but these can be broken down into equivalence classes. Further, under the FCC assumption, we can learn something of the composition of the core and under further assumptions determine the exact contents of the core.

If we do not assume FCC, we witness very interesting behavior in the core. If we see a rectangle which crosses itself n times under $\mathcal{A}$ (and $\pi:\hat{R}\rightarrow R$ is 1-1 everywhere), we can divide R in such a way that we construct a new partition with the full n shift in the core. In such a case, the eigenvalues of M tell us nothing about the core. In fact, if we do not assume FCC we cannot in general conclude that we have it by looking at M. We can in some in some cases conclude that a partition is not FCC.

Finally, if we concentrate on two rectangle core-connected partitions, we can find a finite set of partitions which generate all partitions with two rectangles under elements of GL(2,$\mathbb{Z}$) which commute with $\mathcal{A}$. It is probable that such a generating set exists for partitions with n rectangles for each n.

Appendix - Homology

In this appendix we state two theorems used in this exposition. They can be found in most introductory algebraic topology texts.

THEOREM: There is a homomorphism $\partial_*: H_p(X, A) \rightarrow H_{p-1}(A)$ defined for $A \subset X$ and all p such that the sequence

$$\cdots \rightarrow H_p(A) \xrightarrow{i_*} H_p(X) \xrightarrow{\pi_*} H_p(X, A) \xrightarrow{\partial_*} H_{p-1}(A) \rightarrow \cdots$$

is exact where i and π are inclusions. The same result holds in reduced homology if $A \neq \emptyset$.

THEOREM (Mayer-Vietoris): Let $X = X_1 \cup X_2$; suppose $\{X_1, X_2\}$ is an excisive couple. Let $A = X_1 \cap X_2$. Then there is an exact sequence

$$\cdots \rightarrow H_p(A) \rightarrow H_p(X_1) \oplus H_p(X_2) \rightarrow H_p(X) \rightarrow H_{p-1}(A) \rightarrow \cdots$$

called the Mayer-Vietoris sequence of $\{X_1, X_2\}$. A similar sequence holds in reduced homology of $A \neq \emptyset$.

References

[1] R. Adler, *The Torus and the Disk*, IBM J. of R. & D. 31 (1987), 224-234.

[2] R. Adler and B. Weiss, *Similarity of Automorphisms of the Torus*, Mem. Amer. Math. Soc. 98 (1970).

[3] J. Ashley, B. Kitchens, and M. Stafford, *Boundaries of Markov Partitions*, preprint (1990).

[4] K. Berg, *On the Conjugacy Problem for K-Systems*, Ph.D. thesis, University of Minnesota, 1967.

[5] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Springer Lecture Notes in Math. 470 (1975).

[6] M. Hirsch and C. Pugh, *Stable Manifolds and Hyperbolic Sets*, Proc.Symp. Pure Math. 14 (1970), 133-163.

[7] K. Kim and F. Roush, *Some Results on the Decidability of Shift Equivalence*, J. Comb. Info. & Syst. Sci. 4 (1979), 123-146.

[8] M. Newman, *Integral Matrices*, Academic Press, New York, 1972.

[9] C. Olds, *Continued Fractions*, Mathematical Association of America, Washington D. C., 1963.

[10] E. Seneta, *Non-negative Matrices*, Allen and Unwin, London, 1973.

[11] M. Snavely, *The Torus and Strong Shift Equivalence*, in progress.

[12] M. Stafford, *Markov Partitions for the Doubling Map*, Ph.D. thesis, Northwestern University, 1989.