Answer Sheet 3

April 19, 2014

[HK] 7.3.4: The Markov matrix is:

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

From *Corollary 7.3.6* in the notes we know that $P_n(\sigma_A) = tr(A^n)$. We have $tr(A^1) = 3$ and this it is clear from *figure 7.3.3* that there are 3 points that can transition to themselves.

$$A^2 = \begin{pmatrix} 2 & 2 & 1 & 1 & 0 \\ 2 & 2 & 2 & 2 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

The trace of which is 7. For A^3 we have:

$$A^3 = \begin{pmatrix} 4 & 4 & 3 & 3 & 1 \\ 4 & 4 & 5 & 4 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 1 & 2 \\ 0 & 0 & 2 & 1 & 2 \end{pmatrix}$$

The trace being 12. If we check for the orbits that start at (4) and return to (4) after 3 transistions (we see from the matrix there should be 2) we find:

$$(4) \rightarrow (4) \rightarrow (4) \rightarrow (4)$$

$$(4) \rightarrow (2) \rightarrow (3) \rightarrow (4)$$

[HK] 7.3.5: Using a similar argument to the one in the notes let $\omega \in C_{\alpha_{1-n} \cdots \alpha_{n-1}}$ then:

$$d'_\lambda(\alpha, \omega) := \sum_{i \in \mathbb{Z}} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}} = \sum_{|i| \geq n} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}}$$

$$|\alpha_i - \omega_i| \leq N - 1 \Rightarrow \sum_{|i| \geq n} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}} \leq \sum_{|i| \geq n} \frac{N-1}{\lambda^{|i|}} = (N-1)(\frac{1}{\lambda^{n-1}} . \frac{2}{\lambda - 1})$$

if $\lambda > 2N-1$ then:

$$(N-1)(\frac{1}{\lambda^{n-1}}.\frac{2}{\lambda-1}) < \frac{1}{\lambda^{n-1}}.\frac{2N-2}{2N-2} = \frac{1}{\lambda^{n-1}}$$

so $C_{\alpha_{1-n}\cdots\alpha_{n-1}} \subset B_{\lambda^{1-n}}(\alpha)$ Let $\omega \notin C_{\alpha_{1-n}\cdots\alpha_{n-1}}$ then for some $|k| \leq n-1$ we have $\omega_k \neq \alpha_k$. In which case

$$d'_\lambda(\alpha,\omega) = \sum_{i\in\mathbb{Z}} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}} \geq \frac{1}{\lambda^{|k|}} \geq \frac{1}{\lambda^{n-1}}$$

so $\omega \notin B_{\lambda^{1-n}}(\alpha)$ Hence $C_{\alpha_{1-n}\cdots\alpha_{n-1}} = B_{\lambda^{1-n}}(\alpha)$

[HK] 7.3.6:

$$d'_\lambda(\alpha,\omega) := \sum_{i\in\mathbb{N}} \frac{|\alpha_i - \omega_i|}{\lambda^i} = \sum_{i\geq n} \frac{|\alpha_i - \omega_i|}{\lambda^i}$$

$$|\alpha_i - \omega_i| \leq N-1 \Rightarrow \sum_{i\geq n} \frac{|\alpha_i - \omega_i|}{\lambda^i} \leq \sum_{i\geq n} \frac{N-1}{\lambda^i} = (N-1)(\frac{1}{\lambda^{n-1}}.\frac{1}{\lambda-1})$$

if $\lambda > N$ (note: naturally a sharper condition here, but in fact also implied by $\lambda > 2N-1$) then:

$$(N-1)(\frac{1}{\lambda^{n-1}}.\frac{1}{\lambda-1}) < \frac{1}{\lambda^{n-1}}.\frac{N-1}{N-1} = \frac{1}{\lambda^{n-1}}$$

so $C_{\alpha_{1-n}\cdots\alpha_{n-1}} \subset B_{\lambda^{1-n}}(\alpha)$ Let $\omega \notin C_{\alpha_{1-n}\cdots\alpha_{n-1}}$ then for some $k \leq n-1$ (least) we have $\omega_k \neq \alpha_k$. In which case

$$d'_\lambda(\alpha,\omega) = \sum_{i\in\mathbb{N}} \frac{|\alpha_i - \omega_i|}{\lambda^i} \geq \frac{1}{\lambda^k} \geq \frac{1}{\lambda^{n-1}}$$

so $\omega \notin B_{\lambda^{1-n}}(\alpha)$ Hence $C_{\alpha_{1-n}\cdots\alpha_{n-1}} = B_{\lambda^{1-n}}(\alpha)$

[HK] 7.3.7: Again similar method to the above:

$$\omega \in C_{\alpha_{1-n}\cdots\alpha_{n-1}} \Rightarrow d''_\lambda(\alpha,\omega) = \frac{1}{\lambda^m} \text{ for some } m > n-1$$

$$\Rightarrow C_{\alpha_{1-n}\cdots\alpha_{n-1}} \subset B_{\frac{1}{\lambda^{n-1}}}(\alpha)$$

$$\omega \notin C_{\alpha_{1-n}\cdots\alpha_{n-1}} \Rightarrow d''_\lambda(\alpha,\omega) = \frac{1}{\lambda^m} \text{ for some } m < n-1$$

$$\Rightarrow \omega \notin B_{\frac{1}{\lambda^{n-1}}}(\alpha)$$

[HK] 7.3.8: We know from *Prop 7.3.12* that if a topological Markov chain with connectivity matrix A is transitive then it is topologically mixing. We also know that there exists $m > 0$ (such that A^m has no zero entry) such that if $\alpha \in \Omega_A$ and $n \in \mathbb{N}$, there exists $\alpha' \in \Omega_A$ such that $\alpha_i \neq \alpha'_i$ for all $|i| \leq n$ and $\alpha_{n+m} \neq \alpha'_{n+m}$. Thus $\sigma_A^{n+m}(\alpha) \neq \sigma_A^{n+m}(\alpha')$ and consequently $d''(\alpha,\alpha') \leq \lambda^{-n}$ and $d''(\sigma_A^{n+m}(\alpha),\sigma_A^{n+m}(\alpha')) = 1$. But with this metric the diameter of Ω_A equals 1, so the maximal sensitivity constant is equal to the diameter (=1).

[HK] 7.3.9:

$$sup_{\alpha,\omega\in\Sigma_N}\{d'_\lambda(\alpha,\omega)\} = \sum_{i\in\mathbb{Z}} \frac{N-1}{\lambda^{|i|}} = (N-1).\frac{2}{\lambda-1} = \frac{2N-2}{2N-2} = 1$$

[HK] 7.3.10: Consider $\pi : \Sigma_n \to \Sigma_m$ given by:

$$(\pi(\omega))_i = \quad \omega_i \text{ if } \omega_i \in \{0, ..., m-1\} \tag{1}$$
$$m-1 \text{ if } \omega_i \in \{m, ..., n-1\} \tag{2}$$

π is a factor if π is continuous, surjective and satisfies: $\pi(\sigma_n(\omega)) = \sigma_m(\pi(\omega))$. Coninuity follows from the fact that $\pi(B_{\lambda^{-(n-1)}}(\alpha)) = B_{\lambda^{-(n-1)}}(\pi(\alpha))$. Surjectivity is obvious as Σ_m is fixed under π. Finally:

$$(\pi(\sigma_n(\omega)))_i = (\pi(\omega))_{i+1} = \quad \omega_{i+1} \text{ if } \omega_{i+1} \in \{0, ..., m-1\} \tag{3}$$
$$m-1 \text{ if } \omega_{i+1} \in \{m, ..., n-1\} \tag{4}$$

$$\sigma_m(\pi((\omega)))_i = \quad \sigma_m(\omega)_i \text{ if } \sigma_m(\omega)_i \in \{0, ..., m-1\} \tag{5}$$
$$m-1 \text{ if } \sigma_m(\omega)_i \in \{m, ..., n-1\} \tag{6}$$

$$\sigma_m(\pi((\omega)))_i = \quad \omega_{i+1} \text{ if } \omega_{i+1} \in \{0, ..., m-1\} \tag{7}$$
$$m-1 \text{ if } \omega_{i+1} \in \{m, ..., n-1\} \tag{8}$$

[HK] 7.3.11: For f_4 the left-most point 0 of $[0,1]$ is a fixed point and $f^{-1}(0) = \{0,1\}$ where the point 1 has a nontrivial pre-image, i.e. $1/2 = f^{-1}(1)$. This topological property is to be preserved under topological conjugacy. However, f_λ with $\lambda \in [0,4)$ has again $f^{-1}(0) = \{0,1\}$ with 0 a fixed point but $f^{-1}(1) = \emptyset$.

[HK] 7.3.12: let:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

and

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

Consider $\pi : \Sigma_B \to \Sigma_A$ given by:

$$(\pi(\omega))_i = \quad \omega_i \text{ if } \omega_i \in \{0,2\} \tag{9}$$
$$1 \text{ if } \omega_i = 3 \tag{10}$$

note that the inverse of this is:

$$(\pi^{-1}(\omega))_i = \quad 3 \text{ if } \omega_i = 1 \text{ and } \omega_{i-1} = 2 \tag{11}$$
$$\omega_i \text{ otherwise} \tag{12}$$

To see the Conjucacy note that sequences in Σ_B take the form:

...1...12323....2311.....1

i.e. sets of 1's then sets of 23's. Similarly sequences in Σ_A take the form:

...1...12121....2111.....1

sets of 1's then sets of 21's. π simply switches the 23's with 21's and so the structure of the sequence is clearly preserved under π. π is also clearly a homeomorphism.

[HK] 7.4.2: For $h(\omega) = \bigcap_{n\in\mathbb{N}_0} f^{-n}(\Delta_{\omega_n})$ we require to show:

Injectivness: Assume there exists $x \in \Lambda$ such that $h(\omega^1) = h(\omega^2) = x$ Then by the definition of h, x must be in both the sets $f^{-n}(\Delta_{\omega^1_n})$ and $f^{-n}(\Delta_{\omega^2_n})$ where $n \in \mathbb{N}_0$. But these two sets, for each n, are disjoint unless $\omega^1_n = \omega^2_n$ and hence $\omega^1 = \omega^2$.

Surjectivness: Note that $f(\Lambda) = \Lambda$ and hence if $x \in \Lambda$ then for every $n \in \mathbb{N}_0$, $f^n(x) \in \Delta_1$ or Δ_2. Let, for $i \in \mathbb{N}_0$ ω_i equal the index of whichever set, Δ_1 or Δ_2, $f^n(x)$ is in. Thus:

$$x \in f^{-n}(\Delta_{\omega_n})\forall n \text{ and so as } h \text{ is well defined } h(\omega) = x$$

Continuity: If h not continuous then there exists two sequences ω^i and α^j such that they have the same limit γ in Σ_2^R. But each have different limits in Λ under h. i.e. $h(\omega^i) \to a$ and $h(\alpha^i) \to b$. We have:

$$f^n(h(\omega^i)) \in \Delta_{\omega^i_n} \text{ and } f^n(h(\alpha^i)) \in \Delta_{\alpha^i_n}$$

going to the limit we have:

$$f^n(a) \in \Delta_{\gamma^i_n} \text{ and } f^n(b) \in \Delta_{\gamma^i_n}$$

And as $\bigcap_{n\in\mathbb{N}_0} f^{-n}(\Delta_{\gamma_n})$ is only one point then $a = b$.

[HK] 7.4.3: See for instance Clark Robinson (1994) "Dynamical Systems: Stability, Symbolic Dynamics, and Chaos", section 7.5.1 (attached).

[HK] 7.4.4 Methodology similar to 7.4.3 and notes in [HK] chapter 7. I am happy to post an answer if anyone worked this out.

[HK] 7.4.5: Not important; perhaps you can verify that you can do it for some simple examples.

uous maps from $\mathbf{R}^n$ to itself. Now we let

$$C^0_{b,per}(\mathbf{R}^n) = \{v \in C^0_b(\mathbf{R}^n) : v(\mathbf{x}+\mathbf{w}) = v(\mathbf{x}) \text{ for all } \mathbf{w} \in \mathbf{Z}^n, \mathbf{x} \in \mathbf{R}^n\}$$

and

$$C^1_{b,per}(\mathbf{R}^n) = C^0_{b,per}(\mathbf{R}^n) \cap C^1(\mathbf{R}^n).$$

Then, if $\hat{G} \in C^1_{b,per}(\mathbf{R}^n)$. Because of the periodicity, there is a uniform bound on $\|D\hat{G}_{\mathbf{x}}\|$ for all $\mathbf{x} \in \mathbf{R}^n$.

For $\hat{G} \in C^1_{b,per}(\mathbf{R}^n)$ and $v \in C^0_{b,per}(\mathbf{R}^n)$, as in the proof of Hartman-Grobman we let

$$\begin{aligned} \Theta(\hat{G}, v) &= \mathcal{L}^{-1}\{\hat{G} \circ (id + v) \circ L_A^{-1}\}, \qquad \text{where} \\ \mathcal{L}(v) &= (id - (L_A)_{\#})v = v - L_A \circ v \circ L_A^{-1}. \end{aligned}$$

A direct check shows that $\Theta(\hat{G}, \cdot)$ preserves $C^0_{b,per}(\mathbf{R}^n)$. (We leave this verification to the exercises. See Exercise 7.22.) Exactly as in the proof of the Hartman-Grobman Theorem, if $\mathrm{Lip}(\hat{G})$ is small relative to the distance of the contraction and expansion rates away from one, $\Theta(\hat{G}, \cdot)$ has a fixed point $v_{\hat{G}} \in C^0_{b,per}(\mathbf{R}^n)$. Letting $H_G = id + v_{\hat{G}}$ we get that

$$\begin{aligned} \mathcal{L}(v_{\hat{G}}) &= \hat{G} \circ (id + v_{\hat{G}}) \circ L_A^{-1} \\ H_G &= id + v_{\hat{G}} = L_A \circ L_A^{-1} + L_A \circ v_{\hat{G}} \circ L_A^{-1} + \mathcal{L}(v_{\hat{G}}) \\ &= L_A \circ H_G \circ L_A^{-1} + \hat{G} \circ H_G \circ L_A^{-1} \\ &= G \circ H_G \circ L_A^{-1} \end{aligned}$$

on $\mathbf{R}^n$. For $\mathbf{w} \in \mathbf{Z}^n$, $H_G(\mathbf{x}+\mathbf{w}) = H_G(\mathbf{x}) + \mathbf{w}$ so H_G induces a map h_g on $\mathbf{T}^n$ that satisfies $g \circ h_g \circ f_A^{-1} = h_g$.

Next we check that h_g is one to one. If $h_g(\mathbf{x}) = h_g(\mathbf{y})$, $\bar{\mathbf{x}}$ is a lift of $\mathbf{x}$, and $\bar{\mathbf{y}}$ is a lift of $\mathbf{y}$, then $H_G(\bar{\mathbf{x}}) = H_G(\bar{\mathbf{y}}) + \mathbf{w} = H_G(\bar{\mathbf{y}} + \mathbf{w})$ for some $\mathbf{w} \in \mathbf{Z}^n$. Replacing $\bar{\mathbf{y}}$ with $\bar{\mathbf{y}}' = \bar{\mathbf{y}} + \mathbf{w}$ we get another lift of $\mathbf{y}$ with $H_G(\bar{\mathbf{x}}) = H_G(\bar{\mathbf{y}}')$. Because H_G is one to one, $\bar{\mathbf{x}} = \bar{\mathbf{y}}'$ and $\mathbf{x} = \mathbf{y}$. (The proof that H_G is one to one uses the fact that L_A is expansive: $H_G \circ L_A^n(\bar{\mathbf{x}}) = H_G \circ L_A^n(\bar{\mathbf{y}}')$ for all n so $\bar{\mathbf{x}} = \bar{\mathbf{y}}'$.) Thus h_g is one to one.

By invariance of domain, $h_g(\mathbf{T}^n)$ is open in $\mathbf{T}^n$. Since it is also close, $h_g(\mathbf{T}^n) = \mathbf{T}^n$, and h_g is onto. This completes the proof that h_g is a homeomorphism, that f_A is structurally stable, and the proof of the theorem. □

REMARK 5.1. In Section 9.7 we prove that all Anosov diffeomorphisms are structurally stable. Manning (1974) proved that any Anosov diffeomorphism on a torus is topologically conjugate to a hyperbolic toral automorphism. One conjecture which is still unknown is whether being Anosov implies that all points are nonwandering (or chain recurrent).

7.5.1 Markov Partitions for Hyperbolic Toral Automorphisms

We want to connect the dynamics of a hyperbolic toral automorphism, $f : \mathbf{T}^n \to \mathbf{T}^n$, with that of a subshift of finite type, i.e., to see how symbolic dynamics can be applied to a hyperbolic toral automorphism. We need to find (and define) the replacements for the geometric boxes of the horseshoe which are used to define the symbol sequences. The theory which we give is for all dimensions, but the examples are all in two dimensions where the situation is simpler.

Example 5.2. We introduce the ideas of rectangles, a Markov partition, and the semi-conjugacy using the toral automorphism f_A induced by the matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. We want to subdivide the total space into *rectangles* (which can be taken to be actual parallelepipeds in two dimensions but not higher dimensions). The eigenvalues are $\lambda_u = (1 + 5^{1/2})/2$ with eigenvector $\mathbf{v}^u = (2, 5^{1/2} - 1)$ and $\lambda_s = (1 - 5^{1/2})/2$ with eigenvector $\mathbf{v}^s = (2, -5^{1/2} - 1)$. Note that $\lambda_u + \lambda_s = 1 = \mathrm{tr}(A) > 0$. Thus $\lambda_u = \mathrm{tr}(A) - \lambda_s$, and the fact that the trace of A is a positive integer insures that $\lambda_u > 0$. Then $\lambda_u\lambda_s = \det(A) = -1 < 0$, so this insures that $\lambda_s < 0$. Also, $\mathbf{v}^u$ has positive slope and $\mathbf{v}^s$ has negative slope.

To form the rectangles for A, we look in the covering space, $\mathbb{R}^2$. From the origin and other lattice points take the part of the unstable manifold of this point in $\mathbb{R}^2$ that crosses the fundamental domain above and to the right of the lattice point. See Figure 5.1. Next, extend the stable manifold from the lattice point downward to the point $\mathbf{a}$ where it hits the part of the unstable line segment drawn above. Similarly, extend the stable manifold upward from a lattice point to the point $\mathbf{b}$ where it hits the part of the unstable manifold drawn above. Finally, extend the unstable manifold to the point $\mathbf{c}$ where it hits the line segment $[\mathbf{a}, \mathbf{b}]_s$ in the stable manifold. These line segments, $[\mathbf{a}, \mathbf{b}]_s$ in $W^s(\mathbf{0})$ and $[\mathbf{0}, \mathbf{c}]_u$ in $W^u(\mathbf{0})$ (and their translates in $\mathbb{R}^2$), define two rectangles R_1 and R_2 in $\mathbb{T}^2$. See Figure 5.1.

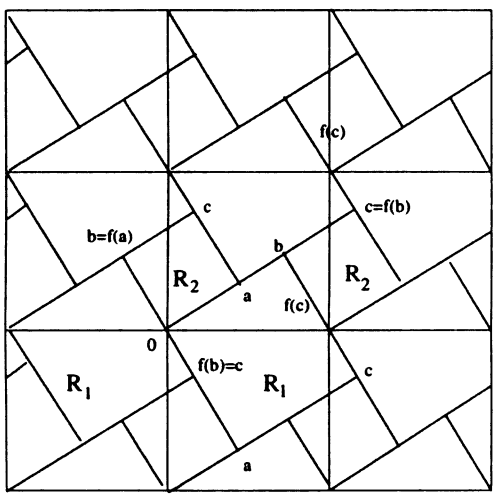

FIGURE 5.1. Rectangles for Example 5.2

To find the images of the rectangles, we first consider the images of the points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$: $f_A(\mathbf{a}) = \mathbf{b}$, $f_A(\mathbf{b}) = \mathbf{c}$, and $f_A(\mathbf{c}) \in [\mathbf{0}, \mathbf{b}]_s$, where $[\mathbf{x}, \mathbf{y}]_s$ is a line segment in the stable manifold from $\mathbf{x}$ to $\mathbf{y}$. See Figure 5.1. Using these images, it follows that

$$f_A(R_1) \text{ crosses } R_1 \text{ and } R_2,$$
$$f_A(R_2) \text{ crosses } R_1.$$

See Figure 5.2. The pair of rectangles $\{R_1, R_2\}$ have the properties of a *Markov partition* for f_A: (i) the collection of rectangles covers $\mathbb{T}^2$, (ii) the interiors of R_1 and R_2 are

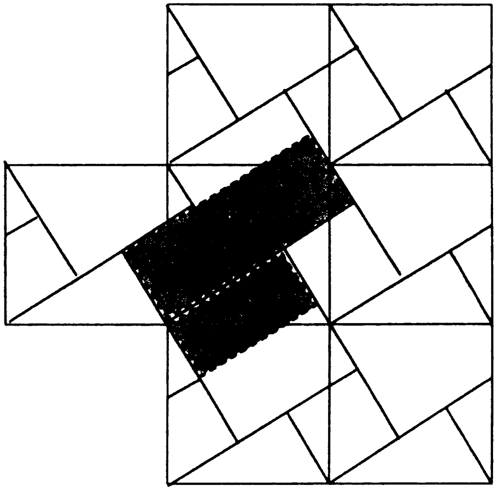

FIGURE 5.2. Images of Rectangles for Example 5.2

disjoint, and (iii) if $f_A(\mathrm{int}(R_i)) \cap \mathrm{int}(R_j) \neq \emptyset$, then $f_A(R_i)$ reaches all the way across R_j in the unstable direction and does not cross the edges of R_j is the stable direction. (There is a fourth condition which we only discuss implicitly below in terms of the semi-conjugacy.) We give the general definition below.

We define a transition matrix which indicates which itineraries for the orbit of a point are allowable: for a transition from rectangle R_i to R_j to be allowable, it must be possible for an orbit of a point to pass from the interior of R_i to the interior of R_j. (We disregard the fact that the image of the boundary of R_2 hits the boundary of R_2.) In this example the transition matrix is given by

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

Notice that this transition matrix B is the same matrix as the original matrix A which induced the toral automorphism. The shift space for B is the two sided subshift of finite type

$$\Sigma_B = \{\mathbf{s} : \mathbf{Z} \to \{1,2\} : b_{s_i s_{i+1}} = 1\}$$

with shift map $\sigma_B = \sigma|\Sigma_B$.

To define the symbolic dynamics, we can not get a continuous map (conjugacy or semiconjugacy) h from $\mathbf{T}^2$ to Σ_B because $\mathbf{T}^2$ is connected and Σ_B is a totally disconnected Cantor set. Also for a point $\mathbf{p} \in \partial(R_i)$ there are at least two choices of rectangles to which $\mathbf{p}$ belongs. Therefore, there is no way to assign a unique symbol sequence to points on the boundary of a rectangle. Instead, we define a map going the other direction, $h : \Sigma_B \to \mathbf{T}^2$. We want h to be a semiconjugacy (continuous, onto, and $f_A \circ h = h \circ \sigma_B$). To do this we define $h : \Sigma_B \to \mathbf{T}^2$ by

$$h(\mathbf{s}) = \bigcap_{n=0}^{\infty} \mathrm{cl}\Big(\bigcap_{j=-n}^{n} f_A^{-j}(\mathrm{int}(R_{s_j}))\Big).$$

We take the images of the interiors because $R_1 \cap f_A^{-1}(R_2)$ does not always equal $\mathrm{cl}(\mathrm{int}(R_1) \cap f_A^{-1}(\mathrm{int}(R_2))$ but can have extra points whose images are on the boundary of R_2. (We must put up with the annoyance to be able to use fewer rectangles.) Using the general theory, Theorem 5.3 proves that this h is a semi-conjugacy. In fact, it proves that h is at most four to one.

In order to give the precise definitions of rectangle and Markov partition, it is necessary to indicate what we mean by the component of a stable or unstable manifold for a point in a rectangle. As we have done before, we use the notation of $\text{comp}_{\mathbf{z}}(S)$ to be the connected component of the set S containing the point $\mathbf{z}$. We think of $W^\sigma(\mathbf{z}, R)$ as equal to $\text{comp}_{\mathbf{z}}(R \cap W^\sigma(\mathbf{z}))$ for $\sigma = u, s$ and R one of the rectangles (if it is connected). However, this definition does not quite work, because even in the rectangles for Example 5.2 there is a difficulty: in $\mathbb{T}^2$, R_1 touches itself along the projection of the line segment from $\mathbf{0}$ to $\mathbf{c}$, $\pi([\mathbf{0}, \mathbf{c}]_s)$. When the total ambient manifold is a torus, a better definition of the stable and unstable manifolds in a rectangle uses the covering space $\mathbb{R}^2$ as follows. Let $\bar{R}$ be one lift of a rectangle R in $\mathbb{T}^2$ to a rectangle in $\mathbb{R}^2$, so $\pi : \bar{R} \to R$ is onto and one to one in the interior. Let $\bar{\mathbf{z}}$ be a lift of $\mathbf{z}$, $\bar{\mathbf{z}} \in \bar{R}$ and $\pi(\bar{\mathbf{z}}) = \mathbf{z}$. For $\sigma = u, s$, define

$$W^\sigma(\mathbf{z}, R) = \pi(W^\sigma(\bar{\mathbf{z}}) \cap \bar{R}).$$

Note in Example 5.2, for $\mathbf{z} = \mathbf{0}$ and rectangle R_1, there are two choices for the lift $\bar{R}_1$ which touches the origin $\bar{\mathbf{0}}$ in $\mathbb{R}^2$. (There is one choice above and to the right of $\bar{\mathbf{0}}$ and one below and to the left.) Making either of these choices, $W^\sigma(\mathbf{0}, R_1) = \pi(W^\sigma(\bar{\mathbf{0}}, \bar{R}_1))$ is a proper subset of $\text{comp}_{\mathbf{0}}(R_1 \cap W^\sigma(\mathbf{0}))$. In fact, $\text{comp}_{\mathbf{0}}(R_1 \cap W^\sigma(\mathbf{0}))$ is the union of the two choices for $W^\sigma(\mathbf{0}, R_1)$.

We now use the motivation of the rectangles defined above for the specific example to give a general definition of both a rectangle and a Markov partition.

Definition. For a hyperbolic toral automorphism on the n-torus, $\mathbb{T}^n$, we proceed as follows. Let R be a subset of $\mathbb{T}^n$ and $\mathbf{z} \in R$. Let $\bar{R}$ is a lift of R to $\mathbb{R}^n$ and $\bar{\mathbf{z}} \in \bar{R}$ be a lift of $\mathbf{z}$, i.e., $\pi : \bar{R} \subset \mathbb{R}^n \to R$ is a homeomorphism, and $\pi(\bar{\mathbf{z}}) = \mathbf{z}$. If R is connected then $\bar{R}$ should be taken to be connected; if R is not connected, then care must be taken to choose the points in $(\pi)^{-1}(R)$ in a reasonable manner, e.g. $\bar{R}$ should be in one fundamental region of $\pi : \mathbb{R}^n \to \mathbb{T}^n$. For $\sigma = u, s$, let

$$W^\sigma(\mathbf{z}, R) = \pi(W^\sigma(\bar{\mathbf{z}}) \cap \bar{R}).$$

We do not give a completely precise definition for the general case of a hyperbolic invariant set Λ. An isolated hyperbolic invariant set has a property called a *local product structure* provided for $\epsilon > 0$ small enough, there is a $\delta > 0$ such that if $d(\mathbf{x}, \mathbf{y}) < \delta$ for $\mathbf{x}, \mathbf{y} \in \Lambda$, then $W^u_\epsilon(\mathbf{x}) \cap W^s_\epsilon(\mathbf{y})$ is a single point in Λ. Let Λ be a hyperbolic invariant set with a local product structure, let R be a subset of Λ that has diameter less than δ, and let $\mathbf{z} \in R$. Then

$$W^\sigma(\mathbf{z}, R) \equiv R \cap W^\sigma_\epsilon(\mathbf{z})$$

using the local stable and unstable manifolds of size ϵ. This general case was considered by Bowen (1970a, 1975). Also see Section 9.6.

Definition. Let f be a diffeomorphism with a hyperbolic invariant set with a local product structure. (This includes the case where f is a hyperbolic toral automorphism.) A nonempty set R of $\mathbb{T}^n$ (or of Λ) is a *(proper) rectangle* provided

(i) $R = \text{cl}(\text{int}(R))$ (where the interior is relative to Λ) so that it is closed, and

(ii) $\mathbf{p}, \mathbf{q} \in R$ implies that $W^s(\mathbf{p}, R) \cap W^u(\mathbf{q}, R)$ is exactly one point, and this point is in R. If we are considering a hyperbolic toral automorphism, then the same lift must be used for R to determine both $W^s(\mathbf{p}, R)$ and $W^u(\mathbf{q}, R)$.

REMARK 5.2. In his general definition, Bowen defines $W^s_\epsilon(\mathbf{p}) \cap W^u_\epsilon(\mathbf{q}) \equiv [\mathbf{p},\mathbf{q}]$. He then demands that for $\mathbf{p},\mathbf{q} \in R$ that $[\mathbf{p},\mathbf{q}]$ is exactly one point, and that this point is in R. Note, if we use Bowen's definition then R_1 is not a rectangle in Example 5.2 because there are points $\mathbf{p}$ and $\mathbf{q}$ in R_1 near $\mathbf{0}$, for which $W^s_\epsilon(\mathbf{p}) \cap W^u_\epsilon(\mathbf{q})$ is in R_2 and not in R_1. Using the fact that the manifold is a torus and our definition of the subsets of the stable and unstable manifolds using lifts, the sets R_1 and R_2 given in the above example are indeed rectangles.

Below, we define a collection of rectangles (a Markov partition) which have the properties needed to use them to define symbolic dynamics. The definitions use the notion of the interior and boundary of a rectangle. A point $\mathbf{p} \in R$ is a *boundary point of R* if arbitrarily near to $\mathbf{p}$ there is a point $\mathbf{q}$ in Λ such that $\mathbf{q} \notin R$. (This is the usual pointset boundary of a subset.) If $\mathbf{p}$ is a boundary point of R, it follows for such $\mathbf{q}$, that either $W^s(\mathbf{p}) \cap W^u(\mathbf{q})$ or $W^s(\mathbf{q}) \cap W^u(\mathbf{p})$ is not in R. Let $\partial(R)$ be the set of all boundary points of R, and the interior of R be the complement of $\partial(R)$ in R, $\mathrm{int}(R) = R \setminus \partial(R)$.

Definition. Assume that $f : M \to M$ is a diffeomorphism which has an isolated hyperbolic invariant set Λ with a local product structure. (This includes the case where f is a hyperbolic toral automorphism with $\Lambda = M$.) A *Markov partition* for f is a finite collection of rectangles, $\mathcal{R} = \{R_j\}_{j=1}^m$, that satisfies the following four conditions. (All interiors are taken relative to Λ.)

(i) The collection of rectangles cover Λ, $\Lambda = \bigcup_{j=1}^m R_j$.

(ii) If $i \neq j$ then $\mathrm{int}(R_i) \cap \mathrm{int}(R_j) = \emptyset$ (so $\mathrm{int}(R_i) \cap R_j = \emptyset$).

(iii) If $\mathbf{z} \in \mathrm{int}(R_i)$ and $f(\mathbf{z}) \in \mathrm{int}(R_j)$ then

$$f(W^u(\mathbf{z}, R_i)) \supset W^u(f(\mathbf{z}), R_j) \qquad \text{and}$$
$$f(W^s(\mathbf{z}, R_i)) \subset W^s(f(\mathbf{z}), R_j).$$

(iv) (The rectangles are small enough.) If $\mathbf{z} \in \mathrm{int}(R_i) \cap f^{-1}(\mathrm{int}(R_j))$ then

$$\mathrm{int}(R_j) \cap f\big(W^u(\mathbf{z}, \mathrm{int}(R_i))\big) = W^u\big(f(\mathbf{z}), \mathrm{int}(R_j)\big) \quad \text{and}$$
$$\mathrm{int}(R_i) \cap f^{-1}\big(W^s(f(\mathbf{z}), \mathrm{int}(R_j))\big) = W^s(\mathbf{z}, \mathrm{int}(R_i))$$

where $W^\sigma(\mathbf{z}', \mathrm{int}(R_k)) = W^\sigma(\mathbf{z}', \mathrm{int}(R_k)) \cap \mathrm{int}(R_k)$ for $\sigma = u, s$, any point $\mathbf{z}'$, and rectangle R_k.

Definition. Once we have a Markov partition, we want to set up the symbolic dynamics of the subshift of finite type by means of a transition matrix. Given a Markov partition $\mathcal{R} = \{R_j\}_{j=1}^m$, the *transition matrix* $B = (b_{ij})$ is defined by

$$b_{ij} = \begin{cases} 1 & \text{if } \mathrm{int}(f(R_i)) \cap \mathrm{int}(R_j) \neq \emptyset \\ 0 & \text{if } \mathrm{int}(f(R_i)) \cap \mathrm{int}(R_j) = \emptyset. \end{cases}$$

The *shift space for B* is defined as

$$\Sigma_B = \{\mathbf{s} : \mathbf{Z} \to \{1, \dots, m\} : b_{s_i s_{i+1}} = 1\}.$$

Letting σ be the shift map on the full m-shift, $\Sigma_m = \{1, \dots, m\}^{\mathbf{Z}}$, define $\sigma_B = \sigma|\Sigma_B : \Sigma_B \to \Sigma_B$.

Example 5.3. For the geometric horseshoe, let $R_i = H_i \cap \Lambda$ for $i = 1, 2$. These two rectangles form a Markov partition with transition matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

For the hyperbolic invariant set created for a homoclinic point, the sets $R_i = \Lambda \cap A_i$ form a Markov partition with transition matrix B given in the proof of Theorem 4.4(b).

REMARK 5.3. Notice that we do not demand that a rectangle be connected, although the examples we give for hyperbolic toral automorphisms are connected. There are examples where a rectangle has countably many components even for a Markov partition of a total space which is connected. In general, for a Markov partition of a hyperbolic invariant set, the total space is often not connected or even locally connected, so a rectangle certainly could not be connected in this case.

REMARK 5.4. Let $\partial(R_i)$ be the boundary of R_i relative to Λ. Conditions (i) and (ii) in the definition of a Markov partition imply that $\partial(R_i) = \{\mathbf{p} \in R_i : \mathbf{p} \in R_j \text{ for some } j \neq i\}$. This holds because clearly $\text{int}(R_i) \cap \{\mathbf{p} \in R_i : \mathbf{p} \in R_j \text{ for some } j \neq i\} = \emptyset$, so $\partial(R_i) \supset \{\mathbf{p} \in R_i : \mathbf{p} \in R_j \text{ for some } j \neq i\}$. Next, if $\mathbf{p} \in \partial(R_i)$, then there are $\mathbf{q}_k \in R_{j_k}$ with $j_k \neq i$ and $\mathbf{q}_k$ converging to $\mathbf{p}$. Because there are a finite number of rectangles, by taking a subsequence we can take all the $j_k = j$ to be the same. Because R_j is closed it follows that $\mathbf{p} \in R_j$. This proves that $\partial(R_i) \subset \{\mathbf{p} \in R_i : \mathbf{p} \in R_j \text{ for some } j \neq i\}$.

REMARK 5.5. Condition (iii) in the definition of a Markov partition insures that if the image of a rectangle hits the interior of another rectangle, then it goes all the way across in the unstable direction and is a subset in the stable direction (goes all the way across in the stable direction when looking at the inverse). Note that if a point $\mathbf{z}$ is on the boundary of a rectangle R_i, then the image of R_i can abut on another rectangle R_j without even going into the interior of rectangle R_j. (Thus the condition (iii) does not necessarily hold for the points on the boundary.)

REMARK 5.6. Condition (iv) is not included in Bowen's definition because he only used small rectangles. It is added to our list to make the point determined by a sequence of rectangles allowed by the transition matrix well defined. This condition prohibits the image of a rectangle R_i from crossing a rectangle R_j twice. Note that it does allow the image to intersect the boundary a second time. (See $f(R_1)$ and R_2 in Figure 5.2.)

We could strength Condition (iv) to the following assumption:
(iv)′ for $\mathbf{z} \in \text{int}(R_i) \cap f^{-1}(\text{int}(R_j))$,

$$\begin{aligned} R_j \cap f(W^u(\mathbf{z}, R_i)) &= W^u(f(\mathbf{z}), R_j) \qquad \text{and} \\ R_i \cap f^{-1}(W^s(f(\mathbf{z}), R_j)) &= W^s(\mathbf{z}, R_i). \end{aligned}$$

This condition does not allow the image of a rectangle R_i to cross the rectangle R_j once and then intersect the boundary a second time. Therefore the partition constructed in Example 5.2 satisfies assumption (iv) but not assumption (iv)′. The advantage of assumption (iv)′ over (iv) is that the definition of the conjugacy in Theorem 5.3 without taking interiors and closures. See Remark 5.10.

For some purposes, people allow the image of a rectangle to cross more than one time. If multiple crossings are allowed, then (1) Condition (iv) is not included in the definition and (2) the transition matrix must be allowed to have integer entries which are larger than one, i.e., we get an adjacency matrix as defined in Section 7.3.1 on subshifts for matrices with nonnegative integer entries. More precisely, assume there is a partition by rectangles $\{R_i\}_{i=1}^n$ which satisfies conditions (i–iii) for a Markov partition but not necessarily condition (iv). To such a partition, we can associate an adjacency matrix $A = (a_{ij})$ where the entry a_{ij} equals the number of times that the image $f(R_i)$ crosses the rectangle R_j. Thus if $a_{ij} = 2$, then $f(R_i)$ crosses R_j twice. We do not pursue this connection. See Franks (1982).

REMARK 5.7. Adler and Weiss (1970) gave a method of constructing simple Markov partitions for hyperbolic toral automorphisms on $\mathbb{T}^2$. Assume A is a 2×2 adjacency matrix with all positive entries and which induces a hyperbolic toral automorphism on $\mathbb{T}^2$. Then, there is always has a partition by two rectangles, $\{R_1, R_2\}$, such that (1) the partition satisfies all the properties of a Markov partition except (iv), and (2) the image $f(R_i)$ has a_{ij} geometric crossings of R_j. The recent theses by Snavely (1990) and Rykken (1993) give more details on constructing such a Markov partition.

REMARK 5.8. It should be noted however that even for Markov partitions for hyperbolic toral automorphisms in $\mathbb{T}^n$ with $n \geq 3$, the boundaries of the rectangles are not smooth. Thus the "rectangles" are much different than the simple two dimensional example leads one to believe. See Bowen (1978b).

REMARK 5.9. Bowen (1970a) proved that any hyperbolic invariant set with a local product structure has a Markov partition. We prove this result in Section 9.6. In this chapter, we restrict ourselves to finding Markov partitions for hyperbolic toral automorphisms on $\mathbb{T}^2$ and the solenoid which is defined in Section 7.7.

We can now state the main result.

Theorem 5.3. *Let $\mathcal{R} = \{R_j\}_{j=1}^m$ be a Markov partition for a hyperbolic toral automorphism on $\mathbb{T}^2$. Let (Σ_B, σ_B) be the shift space and $h : \Sigma_B \to \mathbb{T}^2$ be defined by*

$$h(\mathbf{s}) = \bigcap_{n=0}^{\infty} \mathrm{cl}\Big(\bigcap_{j=-n}^{n} f^{-j}(\mathrm{int}(R_{s_j}))\Big).$$

Then h is a finite to one semiconjugacy from σ_B to f. In fact h is at most m^2 to one where m is the number of rectangles in the partition.

REMARK 5.10. If we used assumption (iv)′ given in Remark 5.6 above, then we could just use the intersection of the images $f^{-j}(R_{s_j})$ to define h,

$$h(\mathbf{s}) = \bigcap_{j=-\infty}^{\infty} f^{-j}(R_{s_j}).$$

This latter intersection is usually used to define the conjugacy. The problem is that $f(\mathrm{int}(R_i)) \cap \mathrm{int}(R_j)$ can be nonempty and $f(R_i)$ abut on the boundary of R_j at points for which there are no nearby interior points, so

$$\mathrm{cl}\left(f(\mathrm{int}(R_i)) \cap \mathrm{int}(R_j)\right) \neq f(R_i) \cap R_j.$$

See Example 5.2. We allow such intersections on the boundary in order to find Markov partitions with fewer rectangles. This forces us to use this slightly more complicated definition of h given above.

REMARK 5.11. This theorem is used in Section VIII.1.2 to prove that the topological entropy of F_A can be calculated by the largest eigenvalue of B.

PROOF. By condition (iv), $\mathrm{cl}(\mathrm{int}(R_{s_k}) \cap f^{-1}(\mathrm{int}(R_{s_{k+1}})))$ is a nonempty subrectangle that reaches all the way across in the stable direction. By induction,

$$\mathrm{cl}\Big(\bigcap_{j=k}^{k+i} f^{-j}(\mathrm{int}(R_{s_j}))\Big)$$

is a nonempty subrectangle that reaches all the way across in the stable direction for any $k \in \mathbb{Z}$ and $i \in \mathbb{N}$. The width of this set in the unstable direction decreases exponentially at the rate given by the inverse of the minimum expansion constant. Thus

$$\bigcap_{n=0}^{\infty} \mathrm{cl}\,\Big(\bigcap_{j=0}^{n} f^{-j}(\mathrm{int}(R_{s_j}))\Big) = W^s(\mathbf{p}_s, R_{s_0})$$

for some $\mathbf{p}_s \in R_{s_0}$. Similarly,

$$\bigcap_{n=0}^{-\infty} \mathrm{cl}\,\Big(\bigcap_{j=n}^{0} f^{-j}(\mathrm{int}(R_{s_j}))\Big) = W^u(\mathbf{p}_u, R_{s_0})$$

for some $\mathbf{p}_u \in R_{s_0}$. Therefore,

$$\bigcap_{n=-\infty}^{\infty} \mathrm{cl}\,\Big(\bigcap_{j=-n}^{n} f^{-j}(\mathrm{int}(R_{s_j}))\Big) = W^s(\mathbf{p}_s, R_{s_0}) \cap W^u(\mathbf{p}_u, R_{s_0})$$

is a unique point $\mathbf{p} = h(\mathbf{s})$. This shows that h is a well defined map.

By arguments like those used for the horseshoe, h is continuous, onto, and a semi-conjugacy.

If $f^j(\mathbf{p}) \in \mathrm{int}(R_{s_j})$ for all j, then $h^{-1}(\mathbf{p})$ is a unique symbol sequence, $\mathbf{s}$, because $f^j(\mathbf{p}) \notin R_k$ for $k \neq s_j$. Thus, h is one to one on the residual subset (in the sense of Baire category)

$$\bigcap_j f^{-j}\Big(\bigcup_i \mathrm{int}(R_i)\Big).$$

Next we show that h is at most m^2 to one, where m is the number of partitions. Let $\mathbf{p} = h(\mathbf{s})$. As we showed above we only have to worry if $f^n(\mathbf{p})$ is on the boundary of some rectangle R_j.

We want to distinguish the boundary points of a rectangle R which are on the edge of an unstable manifold in the rectangle, $W^u(\mathbf{z}, R)$, and those which are on the edge of a stable manifold, $W^s(\mathbf{z}, R)$. Let

$$\partial^s(R) = \{\mathbf{x} \in \partial(R) : \mathbf{x} \notin \mathrm{int}(W^u(\mathbf{x}, R))\} \qquad \text{and}$$
$$\partial^u(R) = \{\mathbf{x} \in \partial(R) : \mathbf{x} \notin \mathrm{int}(W^s(\mathbf{x}, R))\}.$$

Here $\mathrm{int}(W^u(\mathbf{x}, R))$ is the interior relative to a compact part of the manifold $W^u(\mathbf{x}, R)$. Similarly for $\mathrm{int}(W^s(\mathbf{x}, R))$. Then $\partial^s(R)$ is the union of stable manifolds $W^s(\mathbf{z}, R)$, and $\partial^u(R)$ is the union of such unstable manifolds.

If $f^n(\mathbf{p}) \in \partial^s(R_{s_n})$ then $f^j(\mathbf{p}) \in \partial^s(R_{s_j})$ for $j \geq n$. There are at most m choices for s_n. (The reader can check that for a hyperbolic toral automorphism on $\mathbb{T}^2$, there are at most 4 choices.) Since the transitions of interiors are unique, a choice for s_n determines the choices of s_j for $j \geq n$. Similarly if $f^{n'}(\mathbf{p}) \in \partial^u(R_{s_{n'}})$ then a choice for $s_{n'}$ determines the choices of s_j for $j \leq n'$. Combining, there are at most m^2 choices as claimed. □

Example 5.4. As a second example of a hyperbol[illegible]ism, let $A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. As we noted above, if $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, then $A^2 = A_2$. The rectangles R_1 and R_2 from Example 5.2 are still rectangles for this matrix. However the image of R_1

by f_{A_2} crosses R_1 twice. This partition satisfies conditions (i)-(iii) and has A_2 as an adjacency matrix.

If we want to get a transition matrix with only 0's and 1's, we must subdivide the rectangles (split symbols) by taking components of $R_1 \cap f_{A_2}(R_1)$: let the rectangle

$$\begin{aligned} R_{1a} &= \text{comp}\left(\pi(\mathbf{0}), \text{cl}(\text{int}(R_1) \cap f_{A_2}(\text{int}(R_1)))\right) \\ &= \pi(\bar{R}_1 \cap L_{A_2}(\bar{R}_1)) \end{aligned}$$

where L_{A_2} is the map on $\mathbf{R}^2$, and

$$R_{1b} = \text{cl}(R_1 \setminus R_{1a}).$$

These rectangles can also be formed by extending the unstable manifold of the origin until it intersects the stable line segment $[\mathbf{0}, \mathbf{b}]_s$ at the point $\mathbf{e} = f(\mathbf{c})$. See Figures 5.3 and 5.1. The reader can check that

$$\begin{aligned} f_{A_2}(R_{1a}) &\quad \text{crosses} \quad R_{1a}, R_{1b} \text{ and } R_2, \\ f_{A_2}(R_{1b}) &\quad \text{crosses} \quad R_{1a}, R_{1b} \text{ and } R_2, \\ f_{A_2}(R_2) &\quad \text{crosses} \quad R_{1b} \text{ and } R_2. \end{aligned}$$

Thus the transition matrix is

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

This transition matrix has characteristic polynomial $p(\lambda) = -\lambda(\lambda^2 - 3\lambda + 1)$, and eigenvalues 0, $(\lambda_u)^2$, and $(\lambda_s)^2$ where λ_u, and λ_s are the eigenvalues of A. Thus the eigenvalues of B are those of A_2 together with 0. We do not prove it, but the eigenvalues of the transition matrix are always the eigenvalues of the original matrix A together with possibly 0 and/or roots of unity. See Snavely (1990).

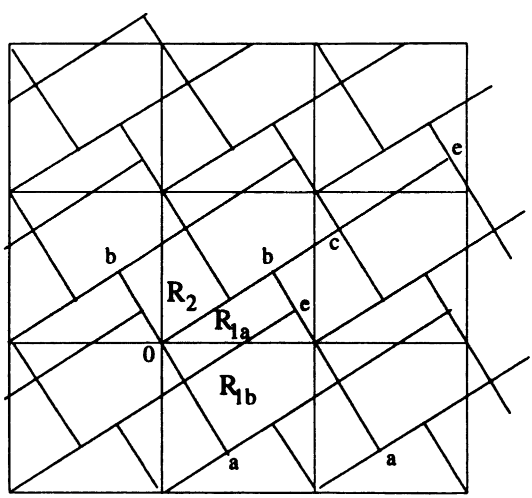

FIGURE 5.3. Markov Partition for Example 5.4