Rotational (and Other) Representations
of Stochastic Matrices

Steve Alpern[1] and V. S. Prasad[2]

[1]Department of Mathematics, London School of Economics,
London WC2A 2AE, United Kingdom. *email: s.alpern@lse.ac.uk*

[2]Department of Mathematics, University of Massachusetts Lowell,
Lowell, MA. *email: vidhu_prasad@uml.edu*

May 27, 2005

Abstract

Joel E. Cohen (1981) conjectured that any stochastic matrix $P = \{p_{i,j}\}$ could be represented by some circle rotation f in the following sense: For some partition $\{S_i\}$ of the circle into sets consisting of finite unions of arcs, we have (*) $p_{i,j} = \mu\left(f\left(S_i\right) \cap S_j\right)/\mu\left(S_i\right)$, where μ denotes arc length. In this paper we show how cycle decomposition techniques originally used (Alpern, 1983) to establish Cohen's conjecture can be extended to give a short simple proof of the Coding Theorem, that any mixing (that is, $P^N > 0$ for some N) stochastic matrix P can be represented (in the sense of * but with S_i merely measurable) by *any* aperiodic measure preserving bijection (automorphism) of a Lesbesgue probability space. Representations by pointwise and setwise periodic automorphisms are also established. While this paper is largely expository, all the proofs, and some of the results, are new.

Keywords: rotational representation, stochastic matrix, cycle decomposition

MSC 2000 subject classifications. Primary: 60J10. Secondary: 15A51

1   Introduction

An *automorphism* of a Lebesgue probability space (X, Σ, μ) is a bimeasurable bijection $f : X \to X$ which preserves the measure μ. If $S = \{S_i\}_{i=1}^n$ is a non-trivial (all $\mu(S_i) > 0$) measurable partition of X, we can generate a stochastic matrix $P = \{p_{i,j}\}_{i,j=1}^n$ by the definition

$$p_{i,j} = \frac{\mu(f(S_i) \cap S_j)}{\mu(S_i)}, \quad i, j = 1, \ldots, n. \tag{1}$$

Since the partition S is non-trivial, the matrix P has a positive invariant (stationary) distribution $v = (v_1, \ldots, v_n) = (\mu(S_1), \ldots, \mu(S_n))$, and hence (by definition) is *recurrent*. If (1) holds, we say that the stochastic matrix P *is represented by* the automorphism f. Equivalently, we can say that the *mass-flow* matrix $R = \{r_{i,j}\}_{i,j=1}^n$ associated with P, v by $r_{i,j} = v_i\, p_{i,j}$ is represented by f, S if

$$r_{i,j} = \mu(f(S_i) \cap S_j), \quad i, j = 1, \ldots, n. \tag{2}$$

(R is a mass-flow matrix if its entries are non-negative and sum to 1 and for all $k = 1, \ldots, n,\ \sum_{i=1}^n r_{i,k} = \sum_{j=1}^n r_{k,j}$.)

Joel E. Cohen [7] proposed representing a stochastic matrix P by a rotation $rot_t(x) = x + t \pmod 1$ of the 'circle' [0,1), using partitions S in which each set S_i is a finite union of arcs (intervals). He called a solution f, S of (1) of this type a *rotational representation* of P. Cohen showed that such representations always exist for 2×2 irreducible stochastic matrices and conjectured that this result could be extended to $n \times n$ matrices. The subsequent results of Alpern [3], Haigh [9], Rodrigues del Tio and Valsero Blanco [20], and Kalpazidou [11][12][15], established and extended Cohen's conjecture in various ways. The purpose of this paper is to show that the cycle decomposition techniques used in [3] to establish Cohen's conjecture, together with the *multitower* constructions of [1] [3], can be modified to give short elementary proofs of results on representations by automorphisms of various types, including the well known Coding Theorem (Theorem 9). The cycle decompositions we use (the Lemmas of this paper) all are established by very elementary algorithms. We believe that further results can be obtained in this area using our multitower construction and the deeper cycle decomposition theorems for Markov chains presented in [13].

2   Definitions

The main results presented in this paper link a given stochastic matrix P with an automorphism and partition f, S of a measure space. So we need to provide some elementary definitions regarding on the one hand, automorphisms and

partitions into towers and multitowers; and on the other hand, decompositions of stochastic or mass-flow matrices into circuits, cycles and tours.

2.1 Automorphisms and multitowers

An automorphism f is said to be aperiodic if $\mu\left(\left\{x : f^k x = x \text{ for some } k > 1\right\}\right) = 0$. It is called pointwise h-periodic if f^h is the identity. It is called setwise h-periodic if there is a set B of measure $1/h$ such that $f^l(B)$ are pairwise disjoint for $l = 0, \ldots, h-1$, in which case $f^h(B)$ must be B.

Let $\pi = (\pi_1, \pi_2, \ldots,)$ be a denumerable probability distribution, and let K_π denote the set of coordinates k for which $\pi_k > 0$. A π-multitower for an automorphism $f : X \to X$ is a family of base sets B_k, $k \in K_\pi$, such that for each k in K_π the *column* $T_k = B_k \cup f(B_k) \cup \cdots \cup f^{k-1}B_k$ consists of disjoint sets of total measure $\mu(T_k) = \pi_k$, and the T_k partition X. Except for the final section, we will be concerned with the case where K_π has a maximum L, and write $\pi = (\pi_1, \pi_2, \ldots, \pi_L)$. If for every $k \in K_\pi$, f^k is the identity on B_k, we will call the multitower *periodic*. If for some h, $\pi_h = 1$, then we call the π-multitower an *h-tower*. Multitowers can be visualized as in Figure 1, with the sets $f^l(B_k)$ drawn as intervals, with $f^{l+1}(B_k)$ drawn directly above $f^l(B_k)$. In this figure, any point x not in the top $(\cup_{k \in K_\pi} f^{k-1}(B_k))$ of the multitower moves to the point directly above it. Points in the top move to the bottom $(\cup_{k \in K_\pi} B_k)$ of the multitower - but we cannot in general say where. However if the tower is periodic, then a point on the top moves to the point directly below it in the bottom.
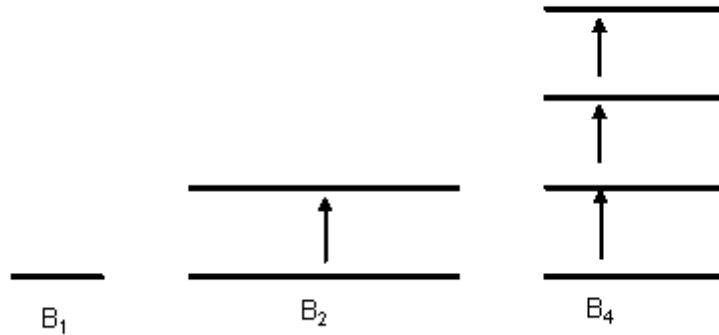

Figure 1: Multitower with $\pi = (1/15, 6/15, 0, 8/15)$

If f is setwise $h-$periodic then it follows from the definition that there is an h-tower for f. If f is pointwise h-periodic it is setwise periodic [10, Lemma

1], and the h-tower must be periodic. If f is aperiodic, the existence of a $\pi-$multitower is given by the following result of [1]. Short and elegant proofs of this finite version of the multitower theorem are given in [18] and [8]. We will show that this result is in fact sufficient to prove the Coding Theorem, which had previously been thought to require results dependent on the Infinite Multitower Theorem (see Section 7).

Theorem 1 ((Finite) Multitower Theorem) *If $\pi = (\pi_1, \pi_2, \ldots, \pi_L)$ is a finite probability distribution with K_π relatively prime, then any aperiodic automorphism f of a Lebesgue probability space (X, Σ, μ) has a π-multitower.*

For example, any aperiodic automorphism has a $(1/13, 4/13, 0, 8/13)$-multitower, because the set $\{1, 2, 4\}$ is relatively prime.

2.2 Flows and circuits

Given a mass-flow matrix R, we define a circuit $c = [c_1, c_2, \ldots, c_l]$ as a circular list of states $c_t \in \{1, \ldots, n\}$ with $r_{c_t, c_{t+1}} > 0$, $t = 1, \ldots, l-1$, and $r_{c_l, c_1} > 0$. If the elements c_t are distinct we call c a *cycle*; if the c_t include all the states $1, \ldots, n$, we call c a *tour*. For any circuit c, define the associated *circuit matrix* $C = \hat{c}$ to be the mass-flow matrix defined by setting $r_{i,j}$ to be the number of times that j follows i in c (counting the transition $c_l c_1$) divided by the length of c. For example, if $c = [1, 2, 2, 1, 1, 1, 1]$, then

$$C = \hat{c} = \frac{1}{7} \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}. \tag{3}$$

If c is a cycle (tour) we call $\hat{c}$ a cycle (tour) matrix. If c is a circuit of length l, then $l \cdot \hat{c}$ is an integer matrix. If circuits c^1 and c^2 have a state in common and respective lengths l_1 and l_2, and β_1 and β_2 are positive integers, we define $\beta_1 c^1 + \beta_2 c^2$ to be any circuit which starts at a common state, follows c^1 for β_1 circuits and then c^2 for β_2 circuits. Note that the circuit matrix corresponding to $\beta_1 c^1 + \beta_2 c^2$ is given by

$$\frac{\beta_1 l_1}{\beta_1 l_1 + \beta_2 l_2} \widehat{c^1} + \frac{\beta_2 l_2}{\beta_1 l_1 + \beta_2 l_2} \widehat{c^2}. \tag{4}$$

It is possible that different circuits define the same circuit matrix, but this will not cause any problems. Although it is more convenient for us to use matrix notation, the concepts are more easily visualized in terms of a network flow, where the (directed) flow from node i to node j is $r_{i,j}$, and circuits, cycles, and tours have the usual combinatorial interpretation.

3 Rotational Representations of 2×2 matrices

Let R be the mass flow matrix corresponding to a 2×2 stochastic matrix P with positive invariant distribution vector (v_1, v_2), $v_1 \leq v_2$. Observe that the

four entries of R are determined by any one of them, say $r_{1,1}$, which cannot exceed v_1. (Also note that $r_{1,2} = r_{2,1}$.) Let S_1 be any interval (or arc) of length v_1, and note that $\mu\left(rot_t\left(S_1\right) \cap S_1\right) = v_1 - t = r_{1,1}$ for $t = r_{1,1} - v_1$. Hence any such 2×2 stochastic matrix has a rotational representation with the interval partition $\left\{S_1, \tilde{S}_1\right\}$. This argument of Cohen ([7]) is essentially an application of the Intermediate Value Theorem to the real function $\mu\left(rot_t\left(S_1\right) \cap S_1\right)$ which takes values v_1 and 0 for t equal to 0 and $1/2$, and hence takes on any intermediate value $r_{1,1}$. For this reason Cohen's argument does not generalize to higher dimensional matrices.

We now give another proof for the 2×2 case which does generalize to higher dimensions. Take $t = 1/2$ and write the pointwise 2−periodic rotation $f = rot_{1/2}$ as a periodic 2-tower of based on the interval $B_2 = [0, 1/2)$. Then partition B_2 into three intervals $B_{2,k}$, $k = 1, 2, 3$, of respective lengths $r_{1,1}/2$, $r_{2,2}/2$, and $r_{1,2} = r_{2,1}$. Define S_1 to be the union of the disjoint intervals $B_{2,1}$, $f\left(B_{2,1}\right)$, and $B_{2,3}$, that is, the intervals labeled 1 in Figure 2. For visual clarity, we have put a space between the three columns of the tower corresponding to the labellings, and taken $r_{1,1} = 1/6$, $r_{2,2} = 1/3$, $r_{1,2} = r_{2,1} = 1/4$, $\mu\left(S_1\right) = v_1 = 5/12$, $\mu\left(S_2\right) = 7/12$. The action of f is to take any point x to the point either directly above or below it.

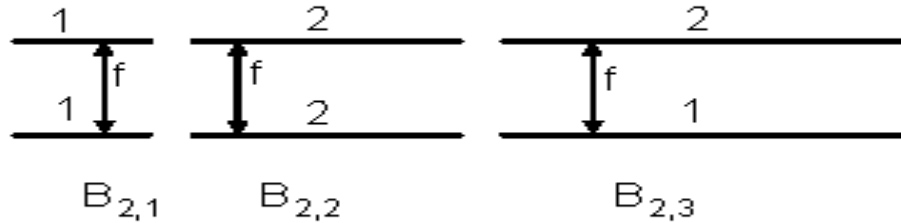Figure 2: A labeled periodic 2-tower for $f = rot_{1/2}$

Essentially what we have done here is to write R as a convex combination of cycle matrices based on the cycles $c^1 = [1]$, $c^2 = [2]$, and $c^3 = [1, 2]$. (Recall

4

that $r_{1,2} = r_{2,1}$.)

$$\begin{pmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{pmatrix} = r_{1,1}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + r_{2,2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + (r_{1,2} + r_{2,1})\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$$
$$= r_{1,1}\widehat{[1]} + r_{2,2}\widehat{[2]} + (r_{1,2} + r_{2,1})\widehat{[1,2]}$$
$$= \alpha_1 \widehat{c^1} + \alpha_2 \widehat{c^2} + \alpha_3 \widehat{c^3}$$

We then choose a common multiple m of all the cycle lengths (in this case any multiple of 2), take $t = 1/m$ and label a periodic m−tower for $f = rot_t$ according to the cycles, repeating each cycle of length l exactly m/l times. The distribution of the measures $\mu\left(f\left(S_i\right) \cap S_j\right)$ in each column k (above B_k) is $\widehat{c^k}$, so the distribution on the whole circle is R.

Using Cohen's construction one can specify in advance (as one) the number of intervals in each partition element S_i, whereas in this construction one can specify in advance the angle of rotation (as $1/m$ for any even m). In fact we can take $f = rot_{d/m}$ for any lowest term fraction d/m where m is even, taking an m-tower over $B_m = [0, 1/m]$ for the pointwise m−periodic automorphism f, and labeling the column over $B_{k,m}$ by m/l_k copies of the cycle c^k, where l_k is the length of c^k.

4 Rotational or pointwise periodic representations of $n \times n$ matrices

In this section fix a stochastic $n \times n$ matrix P, with positive invariant distribution v and associated mass-flow matrix R. We follow [3] to show how it can be rotationally represented. Then we observe that this technique works for any pointwise m−periodic automorphism with appropriate m. We illustrate our method, using the particular example

$$\begin{aligned} \bar{P} &= \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \\ \bar{v} &= (1/2, 1/4, 1/4), \\ \bar{R} &= \begin{pmatrix} 1/4 & 1/8 & 1/8 \\ 1/8 & 0 & 1/8 \\ 1/8 & 1/8 & 0 \end{pmatrix}. \end{aligned} \tag{5}$$

We can write $\bar{R}$ as a convex combination of four cycle matrices $\widehat{c^k}$ as follows:

$$\bar{R} = \frac{1}{4}\widehat{[1,2]} + \frac{1}{4}\widehat{[2,3]} + \frac{1}{4}\widehat{[3,1]} + \frac{1}{4}\widehat{[1]}.$$

Let h be any integer multiple of these cycle lengths (in this example, any even number). In our illustrations we take $h = 6$. Let $f = rot_{d/h}$, where d/h is in

lowest terms, and consider the $h-$tower for f over the interval $B_6 = [0, 1/6)$. See the left side of Figure 3. Assume that, as for $\bar{R}$, we can write

$$R = \alpha_1 \widehat{c^1} + \cdots + \alpha_k \widehat{c^k} + \cdots + \alpha_K \widehat{c^K} \qquad (6)$$

as a convex combination of cycle matrices. Divide the h-tower into k columns with relative distribution α, where the α_k are the coefficients. That is, the interval bases $B_{h,k}$, $k = 1, \ldots, K$ have lengths $\mu(B_{h,k}) = \alpha_k/h$. Then label the k'th column of the h-tower (the one with base $B_{h,k}$), from the bottom up, with repetitions of the cycle c_k. This process, analogous to that in the previous section, is illustrated on the right side of Figure 3. The action of f in the tower is to move every point x to the point directly above it; if there is no such point (i.e. if x is in the top level) then $f(x)$ is the point on the bottom level below x.
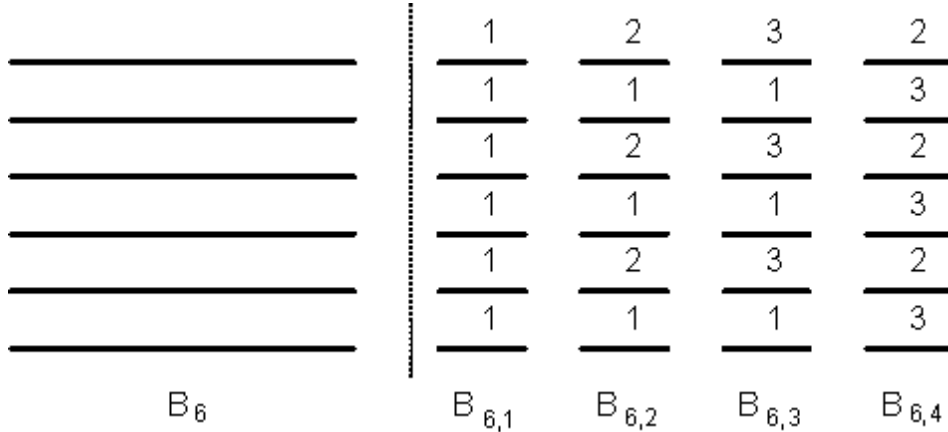

Figure 3: A periodic 6-tower over $B_6 = [0, 1/6)$ - split into four labeled columns

Setting S_i to be the union of all column levels labeled i gives the required representation (2). So all that remains is to establish that a cycle decomposition like (6) can always be found. This was done in [3] by observing that the mass-flow matrices form a convex compact subset of Euclidean n^2 space, and that the cycle matrices are its extreme points (also a combinatorial algorithm for the decomposition is given).

Lemma 2 *Every mass-flow matrix R is a convex combination of cycle matrices.*

Together with the labeling algorithm described above, this gives the positive answer to Cohen's conjecture obtained in [3].

Theorem 3 *Every finite recurrent stochastic matrix can be represented by a circle rotation via a partition into sets consisting of finite unions of intervals.*

By analyzing the proof, we can give a sufficient condition for P to be representable by a given rotation f.

Theorem 4 *Let P be an $n \times n$ stochastic matrix with an invariant distribution v. Let $f = rot_{d/h}$ be the circle rotation by d/h, where h is a common multiple of all the cycle lengths in P (e.g. $n!$) and d/h is in lowest form. Then P can be represented by f via a partition $\{S_i\}_{i=1}^n$ consisting of finite unions of intervals, with $(\mu(S_1), \dots, \mu(S_n)) = v$.*

Suppose that f is not a circle rotation, but merely some other automorphism of some space (X, Σ, μ) which is pointwise $h-$periodic. As observed in Section 2, there is a periodic h-tower for f, based on some periodic set B_h which has analogous properties to the set $[0, 1/h)$ for the rotation. The proof is identical to that for the rotation $rot_{d/h}$, except we can no longer assert that each S_i is a union of intervals. So we have simply,

Corollary 5 *Let P be an $n \times n$ recurrent stochastic matrix, and let f be pointwise h-periodic automorphism of (X, Σ, μ). If h is a common multiple of all cycle lengths of P, then P can be represented by f.*

5 Representations by setwise periodic automorphisms

Suppose now that, unlike Corollary 5, the automorphism f no longer has *pointwise* period h, but merely *setwise* period h. Can we still represent P by f? Setwise periodicity for f means there is a measurable set B_h which forms the base of an h-tower for f. So given a cycle decomposition (6) for the associated mass-flow matrix R, we can still partition B_h into subsets $B_{h,k}$ and label the column levels $f^q(B_{h,k})$ as before. So for our example $\bar{P}, \bar{R}$, given in (5) we can still obtain the labeled tower of Figure 3 - *but it will no longer be a periodic tower*. That is, for a point x at the top of the tower, we can no longer specify the bottom level containing $f(x)$. For example if x is in the top left column level $f^5(B_{6,1})$, it will still be labeled 1, but we cannot determine from the figure the label of $f(x)$. It might be 1 or 3. So the transitions $\mu(f(S_i) \cap (S_j))$ will no longer necessarily match those in the cycle matrices. The main observation is that this problem (not knowing the label of $f(x)$ when x is in the top of the tower) disappears if the labeling has *all the base levels given a common label*. Since we can start the cycles at any state (e.g. [3,1] is the same as [1,3]) all we need is that the cycles in the decomposition (6) all contain some common state, which we can use to label the entire bottom B_h of the tower.

For example, if the setwise period is 6, we can decompose the mass-flow matrix $\bar{R}$ into three cycles, all containing the state 1, as follows.

$$\bar{R} = \frac{1}{2}\widehat{[1]} + \frac{1}{4}[1, \widehat{2, 3}] + \frac{1}{4}[1, \widehat{3, 2}]. \tag{7}$$

We can then label the (non periodic) f−tower of height 6 by partitioning the base B_6 into three sets with relative distribution $(1/2, 1/4, 1/4)$ and labeling them as in Figure 4.
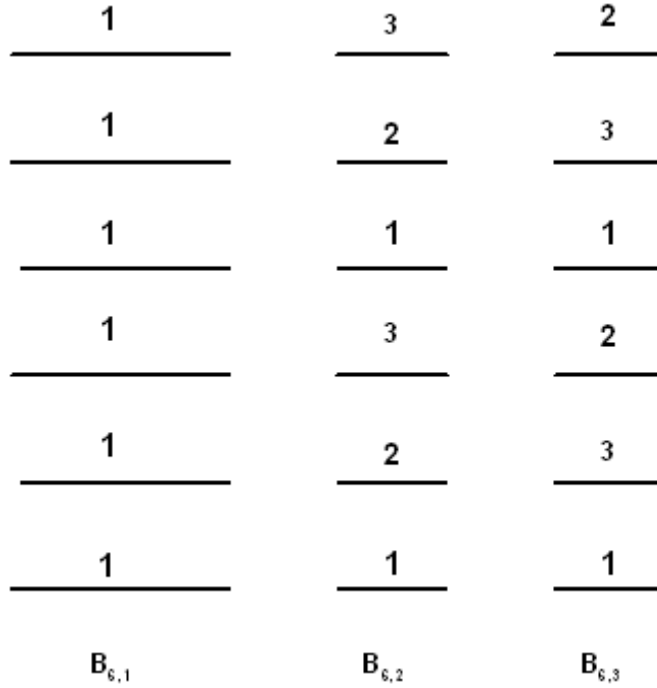
1	3	2
1	2	3
1	1	1
1	3	2
1	2	3
1	1	1
$B_{6,1}$	$B_{6,2}$	$B_{6,3}$

Figure 4: Labeling of a 6−tower for f with base in S_1.

But notice that we didn't really need cycles in our previous constructions-*circuits would have been just as good*. With this relaxation, we can indeed get a decomposition of R into circuits with a common state. In fact we do this by getting the circuits to all contain all states, that is, to be tours (in fact with common lengths). But for a tour to exist, the matrix R (or P) must of course be irreducible. Thus to code R onto the B_h tower for f, it will suffice to have a the following decomposition result.

Lemma 6 *Every mass-flow matrix R corresponding to an irreducible stochastic matrix P is a convex combination of tour matrices with a common length.*

Proof. Since R is irreducible, there exists a tour c^0 of some length l_0. Let $C^0 = \widehat{c^0}$ be the corresponding tour matrix. Observe that for any sufficiently small rational number $p/q < 1$, the matrix

$$\frac{q}{q-p}\left(R - \frac{p}{q}C^0\right) \tag{8}$$

8

has all positive entries and therefore is a mass-flow matrix. Hence by Lemma 2, there is a probability $K-$vector α and cycle matrices C^k of length l_k with

$$\frac{q}{q-p}\left(R - \frac{p}{q}C^0\right) = \alpha_1 C^1 + \cdots + \alpha_K C^K, \text{ or}$$

$$\frac{q}{q-p}R = \frac{p}{(q-p)}C^0 + \alpha_1 C^1 + \cdots + \alpha_K C^K, \text{ or}$$

$$R = \frac{q-p}{q}\left[\frac{p}{(q-p)}C^0 + \alpha_1 C^1 + \cdots + \alpha_K C^K\right]$$

Since $\alpha_1 + \cdots + \alpha_K = 1$, we may rewrite this as

$$R = \frac{q-p}{q}\left[\alpha_1\left(\frac{pC^0}{(q-p)} + C^1\right) + \cdots + \alpha_K\left(\frac{pC^0}{(q-p)} + C^K\right)\right], \text{ or} \quad (9)$$

$$R = \alpha_1\left(\frac{p}{q}C^0 + \frac{(q-p)}{q}C^1\right) + \cdots + \alpha_K\left(\frac{p}{q}C^0 + \frac{(q-p)}{q}C^K\right). \quad (10)$$

The result will now follow if we can show that the mass-flow matrices

$$M^k = \frac{p}{q}C^0 + \frac{(q-p)}{q}C^k = \widehat{t^k}$$

are tour matrices corresponding to tours t^k of some common length s.

Let s be any common multiple of the circuit lengths $l_0, \ldots, l_K$ and define the integers $s_j = s/l_j$, $j = 0, \ldots, K$. Then the tour $t^k = ps_0 \cdot c^0 + (q-p)\, s_k \cdot c^k$ has length qs and by (4) its associated tour matrix $\widehat{t^k}$ is given by

$$
\begin{aligned}
\widehat{t^k} &= \frac{p\, s_0\, l_0}{p\, s_0\, l_0 + (q-p)\, s_k\, l_k}C^0 + \frac{(q-p)\, s_k\, l_k}{p\, s_0\, l_0 + (q-p)\, s_k\, l_k}C^k \\
&= \frac{ps}{qs}C^0 + \frac{(q-p)\, s}{qs}C^k = M^k, \text{ as claimed.}
\end{aligned}
$$

∎

If we apply this algorithm to the mass-flow matrix $\bar{R}$ of (5) we don't get the decomposition (7) but the following: Take $p/q = 3/8$ and $C_0 = [\widehat{1,2,3}]$. Then we have, taking $s = 3$, the following decomposition of $\bar{R}$:

$$\bar{R} - \frac{3}{8}[\widehat{1,2,3}] = \begin{pmatrix} 2/8 & 1/8 & 1/8 \\ 1/8 & 0 & 1/8 \\ 1/8 & 1/8 & 0 \end{pmatrix} - \frac{3}{8}\begin{pmatrix} 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \\ 1/3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2/8 & 0 & 1/8 \\ 1/8 & 0 & 0 \\ 0 & 1/8 & 0 \end{pmatrix},$$

$$\frac{8}{5}\left(\bar{R} - \frac{3}{8}[\widehat{1,2,3}]\right) = \frac{1}{5}\begin{pmatrix} 2 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \frac{2}{5}[\widehat{1}] + \frac{3}{5}[\widehat{1,3,2}], \text{ or}$$

$$\bar{R} = \frac{5}{8}\left(\frac{3}{5}[\widehat{1,2,3}] + \left(\frac{2}{5}[\widehat{1}] + \frac{3}{5}[\widehat{1,3,2}]\right)\right)$$

$$= \frac{2}{5}\cdot(1/24)\begin{pmatrix} 15 & 3 & 0 \\ 0 & 0 & 3 \\ 3 & 0 & 0 \end{pmatrix} + \frac{3}{5}\cdot(1/24)\begin{pmatrix} 0 & 3 & 5 \\ 5 & 0 & 3 \\ 3 & 5 & 0 \end{pmatrix}$$

9

The two tours are

$$
\begin{aligned}
3\,[1,2,3] + 15\,[1] &= [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,3,1,2,3,1,2,3]\,, \\
3\,[1,2,3] + 5\,[1,3,2] &= [1,3,2,1,3,2,1,3,2,1,3,2,1,3,2,1,2,3,1,2,3,1,2,3]\,,
\end{aligned}
$$

where the common length $qs = 8 \cdot 3 = 24$.

Now we know that we can label the f-tower of height h over the periodic set B_h with all of the base set B_h having a common label. Thus the same arguments as before allow us to represent P by h. In particular

Theorem 7 *Let P be an $n \times n$ irreducible stochastic matrix. Let l_0 be the length of a minimal tour, let r be a minimum nonzero entry of the associated mass-flow matrix R, and let q be any integer with $ql_0 > \dfrac{n}{r}$. Let f be a setwise periodic automorphism of a Lebesgue probability with period h. If h is an integer multiple of $q\,l_0\,\operatorname{lcm}[2,\ldots,n]$ then P can be represented by f.*

Proof. In the proof of Lemma 6, let $p = 1$ and take M^0 to correspond to a tour c^0 of minimum length l_0. It is easy to see that this implies that no state appears in the tour c^0 more than n times and consequently that no entry $c_{i,j}^0$ of the integer matrix $l_0 M^0$ can exceed n. So if $ql_0 > n/r$, we have

$$
r_{i,j} - \frac{c_{i,j}}{ql_0} > r - \frac{n}{(n/r)} = 0,
$$

implying that this condition indeed makes q 'sufficiently large' in the sense of the proof of Lemma 6. Next take s in the proof of the previous Lemma to be $l_0 \cdot \operatorname{lcm}[2,\ldots,n]$. This ensures that all the tours t^k have length $qs = q\,l_0\,\operatorname{lcm}[2,\ldots,n]$, so the corollary holds if m is equal to this length. If for some integer u we have $m = u\,q\,l_0\,\operatorname{lcm}[2,\ldots,n]$, then simply replace the t^k in the cycle decomposition by their u-fold repetitions. ∎

6 Representations by aperiodic automorphisms.

If we want to represent a matrix P by an aperiodic automorphism, we must obtain the multitower to label by the multitower theorem. However, the circuit decomposition given by Lemma 6 has all the circuit lengths equal. To apply the multitower theorem, it will be necessary to have a decomposition in which the circuit lengths are relatively prime. This is not always possible, even for irreducible matrices. For example, if we have

$$
P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix},
$$

then all R-circuits have even length. To avoid this we require that P be a mixing matrix, that is, P^m has all positive entries, for some m (and all larger m). A matrix is mixing if and only if it is irreducible and aperiodic. In the example we are using, the matrices $\bar{P}$ and $\bar{R}$, are indeed mixing (take $m = 2$).

The decomposition of $\bar{R}$ given in (7) already has cycle lengths of $1, 3, 3$, which are relatively prime. So if $f : X \to X$ is any aperiodic automorphism, we set $\pi = (1/4, 0, 3/4)$ and apply the multitower theorem to obtain sets B_1 and B_3 with $\mu(B_1) = 1/4$ and $3\mu(B_3) = 3/4$, such that the four sets $B_1, B_3, f(B_3), f^2(B_3)$ partition the underlying space X. We then partition B_3 into two sets $B_{3,1}$ and $B_{3,2}$ of measure $1/4$ each, and label the towers over $B_1, B_{3,1}$, and $B_{3,2}$ according to the cycles $[1], [1, 2, 3], [1, 3, 2]$, as in Figure 6.
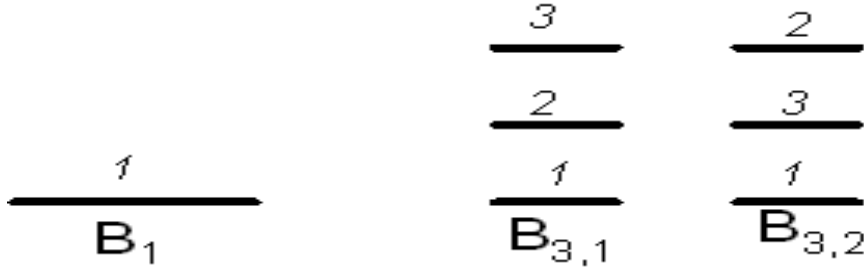

Figure 5: A $(1/4, 0, 3/4)$ multitower for f, labeled to represent $\bar{R}$

This represents $\bar{P}$ by any given aperiodic automorphism f. To show that this can be accomplished for any mixing matrix P, we must establish the following.

Lemma 8 *Let R be the mass-flow matrix corresponding to a mixing stochastic matrix P. Then R is a convex combination of circuit matrices of relatively prime lengths and with a common state.*

Proof. Fix any state, say $i = 1$. Since P is mixing, we have $p_{1,1}^m > 0$ and $p_{1,1}^{m+1} > 0$. Hence there are circuits of length m and $m + 1$ containing the state 1. Hence there are circuit matrices M^{-1} and M^0 of respective lengths m and $m + 1$, which both contain the state 1 (this means that the sum of entries in their first row is positive). For sufficiently small scalars α_{-1} and α_0, the matrix

$$R - \alpha_{-1}M^{-1} - \alpha_0 M^0$$

has all positive entries, and hence

$$\frac{1}{1 - \alpha_{-1} - \alpha_0} \left(R - \alpha_{-1}M^{-1} - \alpha_0 M^0 \right)$$

is a mass flow matrix. Hence by Lemma 2, we have

$$\frac{1}{1 - \alpha_{-1} - \alpha_0} \left(R - \alpha_{-1}M^{-1} - \alpha_0 M^0 \right) = \alpha_1 M^1 + \cdots + \alpha_K M^K,$$

where $(\alpha_1, \ldots, \alpha_K)$ is a probability vector and $M^1, \ldots, M^K$ are tour matrices (in particular they contain the state 1). Hence we have

$$R = \alpha_{-1}M^{-1} + \alpha_0 M^0 + \left[\alpha_1 (1 - \alpha_{-1} - \alpha_0) M^1 + \cdots + \alpha_K (1 - \alpha_{-1} - \alpha_0) M^K \right].$$

Since the $K+2$ matrices on the right hand side all contain the state 1 and their lengths include m and $m+1$, we are done. ∎

Theorem 9 *Let $P = \{p_{i,j}\}_{i,j=1}^n$ be a mixing stochastic matrix and let $f : X \to X$ be an aperiodic automorphism of a Lebesgue probability space (X, Σ, μ). Then there is a non-trivial measurable partition $X = S_1 \cup S_2 \cup \cdots \cup S_n$ of X such that*

$$p_{i,j} = \frac{\mu\left(f\left(S_i\right) \cap S_j\right)}{\mu\left(S_i\right)}, \quad for\ all\ i, j = 1, \ldots, n.$$

Proof. Let R be the mass-flow matrix associated with P and, using Lemma 3, write it as
$$R = \alpha_1 M^1 + \alpha_2 M^2 + \cdots + \alpha_K M^K,$$
where all the M^k contain the state 1 and their lengths are relatively prime. Define a finite probability vector $\pi = (\pi_1, \ldots, \pi_L)$ by $\pi_i = \sum\limits_{k:\text{length}(M^k)=i} \alpha_k$, so that by the previous sentence the i with $\pi_i > 0$ are relatively prime, and L is the largest such i. Let B_i denote the base sets given by the multitower theorem applied to f and π. If M^k corresponds to a circuit c^k of length i, define $B_{i,k} \subset B_i$ to be a partition of B_i with $\mu(B_{i,k}) = \alpha_k/\pi_i$, and attach the label c_{h-1}^k to the set $f^h(B_{i,k})$. As usual, set S_i to be the union of all sets in the multitower partition which are labeled with i. ∎

It is interesting to note that the logical structure of this section, where the Coding Theorem is obtained as a corollary of the Finite Multitower Theorem, is exactly the reverse of that of [1]. In that paper, the Coding Theorem was proved first, by a detailed limiting argument, and then the Finite Multitower Theorem was obtained as an easy consequence.

7 Representing infinite stochastic matrices

This paper is primarily concerned with demonstrating how *finite* cycle decompositions and the *finite* version (Theorem 1) of the multitower theorem can be combined to represent finite stochastic matrices by various types of automorphisms. However if we use the (harder) *infinite* version of the multitower theorem [2], where the word *finite* in the hypotheses of Theorem 1 is replaced by *denumerable,* then we can fairly easily obtain the infinite version of the Coding Theorem. (An irreducible matrix is said to be aperiodic if the set of return times to a given state are relatively prime.)

Theorem 10 *Let $P = \{p_{i,j}\}_{i,j=1}^\infty$ be an irreducible aperiodic stochastic matrix with a positive invariant distribution (so that P is what is called positive-recurrent). Let $f : X \to X$ be an aperiodic automorphism of a Lebesgue probability space (X, Σ, μ). Then there is a partition $S = \{S_i\}_{i=1}^\infty$ of X such that*

$$p_{i,j} = \frac{\mu\left(f\left(S_i\right) \cap S_j\right)}{\mu\left(S_i\right)}, \quad for\ all\ i, j = 1, \ldots, \infty.$$

The same method of coding towers gives an easy proof of this result, if we use the infinite version of the multitower theorem. Fix any state, say state 1, and let $w^1, \ldots, w^M$ be all the circuits of length k containing the state 1 exactly once (other states may be repeated). Define $k\pi_k$ to be the probability that the first return time to state 1 in the Markov chain determined by P is k. Using the (infinite) multitower theorem, let B_k be the bases of the multitower, with $\mu(B_k) = \pi_k$. Partition B_k into m sets $B_{k,m}$, $m = 1, \ldots, M$, according to the relative probabilities of the circuits $w^1, \ldots, w^M$ in the Markov chain. Then label the k column levels above the base $B_{k,m}$ with the states in the circuit w^m, starting with a 1 at the bottom. Setting S_i to be the union of all the column levels labelled i gives the required transition probabilities $p_{i,j}$. In fact, for any sequence $i_1, i_2, \ldots, i_L$ in which $i_l \neq 1$, for all $1 < l \leq L$, we have

$$v(i_1) \ p_{i_1,i_2} \ p_{i_2,i_3} \ \cdots p_{i_{L-1}i_L} = \mu\left(f^{L-1}\left(S_{i_1}\right) \cap \cdots \cap f\left(S_{i_{L-1}}\right) \cap S_{i_L}\right).$$

The details can be found in [6]. For related work on rotational representations of stochastic matrices, see [14].

References

[1] Alpern, S. (1979). Generic properties of measure preserving homeomorphisms. In: Ergodic Theory, Proceedings, Oberwohlfach 1978. Lecture Notes in Mathematics 729, Springer-Verlag, Berlin.

[2] Alpern, S. (1981). Return times and conjugates of an antiperiodic transformation. Ergod. Th. & Dynam. Sys. 1, 135-143.

[3] Alpern, S. (1983). Rotational representations of stochastic matrices. Annals of Probability 11, No. 3, 789-794.

[4] Alpern, S. and Prasad,V. S. (1989), Coding a stationary process to one with prescribed marginals. Annals of Probability 17 , no. 4, 1658–1663.

[5] Alpern, S., and Prasad, V. S. (2001). Typical dynamics of volume preserving homeomorphisms. Cambridge Tracts in Mathematics, 139. Cambridge University Press, Cambridge, .xx+216 pp. ISBN 0-521-58287-3.

[6] Alpern, S., and Prasad, V. S. (2005), Multitowers, conjugacy and coding. Preprint.

[7] Cohen, Joel E. (1981). A geometric representation of stochastic matrices; theorem and conjecture. Annals of Probability 9, 899-901.

[8] Eigen, S. J., and Prasad, V. S. (1997). Multiple Rokhlin tower theorem: a simple proof. New York J. Math. 3A (1997/98), Proceedings of the New York Journal of Mathematics Conference, June 9–13, 1997, 11–14

[9] Haigh, J. (1985). Rotational representation of stochastic matrices. Annals of Probability 13, 1024-1027.

[10] Halmos, P. R. (1956). Lectures on Ergodic Theory. Chelsea Publishing Company, New York.

[11] Kalpazidou, S. (1994). Rotational representations of transition matrix functions. Annals of Probability 22, No. 2, 703-712.

[12] Kalpazidou, S. (1995). On the rotational dimension of stochastic matrices. Annals of Probability 23, No 2, 966–975.

[13] Kalpazidou, S. (1995), Cycle Representations of Markov Processes. Applications of Mathematics, Vol. 28. Springer-Verlag, New York.

[14] Kalpazidou, S. (1998). Cycle representations of denumerable stochastic matrices. Stochastic Anal. Appl. 16 , no. 5, 895–906.

[15] Kalpazidou, S. and Tzouvaras, L. (2001). The forward and backward rotational decompositions of Markov chains. Stochastic Anal. Appl. 19, 3, 399–412.

[16] Kieffer, J. C. (1980). On coding a stationary process to achieve a given marginal distribution. Ann. Probab., 8 , 131–141.

[17] Kornfeld, I. (2004). Some old and new Rokhlin towers. Contemporary Mathematics, 356 ,145–169.

[18] Pridhod'ko, A. A. (1995). Special representation of an aperiodic automorphism of a Lebesgue space. Mathematical Notes 58, No. 2.

[19] Prikhod'ko, A. A.; Ryzhikov, V. V. (1996). The Rokhlin-Halmos-Alpern maximal lemma. (Russian. Russian summary) Vestnik Moskov. Univ. Ser. I Mat. Mekh., no. 3, 92, 37–41, (1996) ; translation in Moscow Univ. Math. Bull., 51, no. 3, 31–34, .

[20] P. Rodriguez del Tio and M. C. Valsero Blanco (1991). A characterization of reversible Markov chains by rotational representation. Annals of Probability 19, No. 2, 605-608.