ISRAEL JOURNAL OF MATHEMATICS, Vol. 27, No. 1, 1977

EQUIVALENCE OF TOPOLOGICAL MARKOV SHIFTS

BY

ROY L. ADLER, L. WAYNE GOODWYN
AND BENJAMIN WEISS

ABSTRACT

We show that any two topological Markov shifts both of whose topological entropy equals $\log n$ (for some n) are equivalent by a finitistic coding.

1. Introduction

How are two dynamical systems with the same topological entropy related? Can a theory be developed for topological entropy analogous to that of Ornstein's for Bernoulli shifts? Examples abound of systems having equal entropy but different number of fixed points. Therefore topological conjugacy is too strong an equivalence to yield anything interesting. We shall establish a weaker type of relation for the class of topological Markov shifts (shifts of finite type). We first deal with the aperiodic case and a restriction on the value of the topological entropy. Removal of these restrictions will be left to subsequent work.

The present notion of equivalence involves the construction of continuous maps which are not invertible in the topological sense. However, they are in the measurable sense and consequently provide measure theoretic isomorphisms between Markov shifts with the same maximal measure entropy. This type of isomorphism has a special property which might be described as *finitistic coding*, which means loosely speaking that a component of a sequence in the range of the isomorphism is determined by a finite, though not necessarily uniformly finite, number of components of the preimage. An example of this appeared in [1] but the idea was not developed there. A precise definition along with a brief discussion was presented in [2].

The authors wish to express their gratitude to Wm. Parry, the University of

Received April 4, 1976

Warwick, and the British Science Research Council for the opportunity of attending the 1975 symposium on ergodic theory during which some ideas of this paper crystallized, especially in discussions with Wm. Parry.

2. Preliminaries

Consider a set $\mathcal{S} = \{1, \cdots, N\}$ of N symbols and an $N \times N$ matrix $T = (t_{ij})$ of zeros and ones. An n-tuple, sometimes called an n-block, $(\xi_1, \cdots, \xi_n)$ of symbols $\xi_i \in \mathcal{S}$, $i = 1, \cdots, n$, is said to be *T-admissible* if $t_{\xi_i, \xi_{i+1}} = 1$, $i = 1, 2, \cdots, n-1$. Left infinite sequences $(\cdots \xi_{-2}, \xi_{-1})$, right infinite sequences $(\xi_0, \xi_1, \cdots)$ and two sided infinite sequences $(\cdots \xi_{-1}, \xi_0, \xi_1, \cdots)$ of symbols from $\mathcal{S}$ are likewise called *T-admissible* if $t_{\xi_n, \xi_{n+1}} = 1$ for $n \leqq -2$, $n \geqq 0$, and $n \in Z$ respectively. Instead of adopting terminology from graph theory in this work we shall use terms more suitable for symbolic dynamics. For this reason we call T a *transition matrix*. It is a one step rule which governs the allowable immediate successors of a symbol in an admissible n-block or sequence. T defines admissible 2-blocks and conversely. Regarding notation we shall occasionally write $i \to j$ for $t_{ij} > 0$ or for the fact that (i, j) is admissible. Also we shall find it convenient to write $\xi \,|_i^j$ for the $(j - i + 1)$-block $(\xi_i, \xi_{i+1}, \cdots, \xi_j)$ where ξ_k is the kth component of an n-block or sequence ξ, and sometimes $\xi \,|_i = \xi \,|_i^i = \xi_i$ for the ith component of ξ.

Let (T) denote the subspace of $\mathcal{S}^Z$ consisting of two-sided T-admissible sequences; and σ the shift transformation on (T), i.e., $\sigma\xi \,|_n = \xi \,|_{n+1}$, $\xi \in (T)$. In the subspace topology of the usual product topology on $\mathcal{S}^Z$ originating from the discrete topology on $\mathcal{S}$ the space (T) is compact metric and σ a homeomorphism. The pair $((T), \sigma)$ is an abstract dynamical system called by various names, *intrinsic Markov chain, topological Markov shift*, and *subshift of finite type.*

T is said to be *irreducible* if for $i, j \in \mathcal{S}$ there is an integer n depending on i, j such that $t_{ij}^{(n)} > 0$, i.e., there exists an admissible n-block beginning with i and ending with j. A *cycle* is defined as an admissible n-block with the same initial and final symbol. The greatest common divisor of cycle lengths minus one is called the *period* of T. In this paper T will be called *aperiodic* if there exists an n such that $T^n > 0$, i.e., $t_{ij}^{(n)} > 0$ for $i, j \in \mathcal{S}$. It can be shown that T is aperiodic if and only if T is irreducible and has period 1.

3. Main theorem

Let T_1 and T_2 be two aperiodic transition matrices of dimension $N_1 \times N_1$ and $N_2 \times N_2$ respectively having λ as common maximal characteristic value. In this paper we shall assume λ is an integer. Then there exists an aperiodic T_3 with the same maximal characteristic value λ and mappings π_1 and π_2 of (T_3) into (T_1) and (T_2) respectively such that π_1 and π_2 are

i) onto,

ii) finite-to-one,

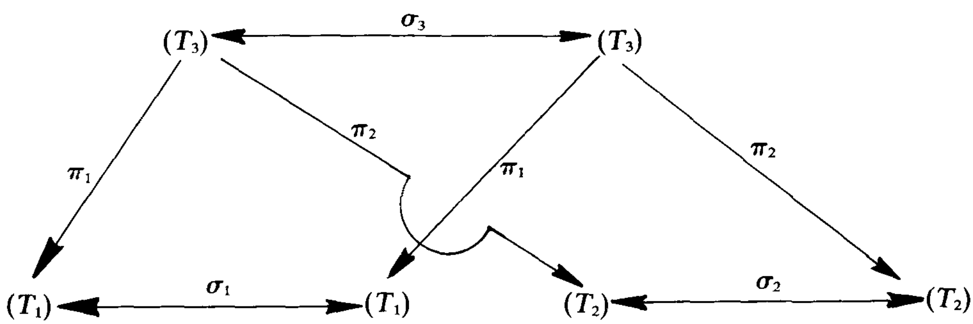
iii) continuous,

satisfy

iv) $\pi\sigma_3 = \sigma_1\pi_1$, $\pi_2\sigma_3 = \sigma_2\pi_2$,

and are

v) one-to-one onto after removing from (T_1), (T_2) and (T_3) shift invariant sets having measure zero with respect to any of the existent shift invariant ergodic probability measures which are positive on open sets.



Lemma 1. *Let T be an irreducible $N \times N$ transition matrix with maximal positive characteristic value λ, an integer. Then there exists an irreducible $\tilde{N} \times \tilde{N}$ transition matrix $\tilde{T}$ with row sum λ and a mapping π of $(\tilde{T})$ into (T) such that π is*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfies

iv) $\pi\sigma = \tilde{\sigma}\pi$

and is

v) *a one-to-one map of $(\tilde{T}) - \tilde{\mathcal{N}}$ onto $(T) - \mathcal{N}$ where $\tilde{\mathcal{N}}$ and $\mathcal{N}$ are shift invariant sets of measure zero with respect to any of the existent shift invariant ergodic probability measures which are positive on open sets.*

Proof. Since T is irreducible there exist cycles. Relabeling if necessary we can assume that $c = (1, 2, \cdots, i_0, 1)$ is a cycle of minimum length. This means that the upper left $i_0 \times i_0$ block of T has the following form:

$$\begin{matrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \end{matrix}$$

From the Perron–Frobenius theorem on nonnegative matrices there exist positive numbers and in this case integers ν_i, $1 \leqq i \leqq N$ such that

$$(*) \qquad \sum_{j=1}^{N} t_{ij}\,\nu_j = \lambda \nu_i.$$

The symbol set $\tilde{\mathscr{S}}$ for $\tilde{T}$ is defined by $\tilde{\mathscr{S}} = \{(i, \alpha) : 1 \leqq \alpha \leqq \nu_i,\ 1 \leqq i \leqq N\}$. For a given i let $j_1 < j_2 < \cdots < j_{K_i}$ be the indices such that $t_{i,j_k} = 1$, $1 \leqq k \leqq K_i$. Consider the following subset $\{(j_k, \beta) : 1 \leqq \beta \leqq \nu_k,\ 1 \leqq k \leqq K_i\}$ of $\tilde{\mathscr{S}}$ in lexicographic order. Because of (*) we can relabel the elements of this subset by $\{b_{(\alpha-1)\lambda+l} : 1 \leqq l \leqq \lambda,\ 1 \leqq \alpha \leqq \nu_i\}$ and order lexicographically. We construct $\tilde{T}$ by defining the admissible transitions for (i, α), namely

$$(i, \alpha) \rightarrow b_{(\alpha-1)\lambda+1}, b_{(\alpha-1)\lambda+2}, \cdots, b_{\alpha\lambda}\ ; 1 \leqq \alpha \leqq \nu_i.$$

$\tilde{T}$ has row sum λ.

To prove $\tilde{T}$ is irreducible we must show any $(j, \beta) \in \mathscr{S}$ can be "reached" from any $(i, \alpha) \in \mathscr{S}$. In one step λ symbols are reached from (i, α). In the second step the transitions spread out still further; and from the way $\tilde{T}$ is constructed we see that eventually some $(k, 1)$ can be reached from (i, α). From $(k, 1)$ the transitions spread out until they eventually include all (l, γ), $1 \leqq \gamma \leqq \nu_l$ for some l. Because T is irreducible j can be reached from l. Thus (j, β) can be reached from some (l, γ).

Let us define a projection π of $\tilde{\mathscr{S}}$ onto $\mathscr{S}$ by $\pi(i, \alpha) = i$. Since $(\pi\tilde{\xi}_1, \cdots, \pi\tilde{\xi}_n)$ is T-admissible if $(\tilde{\xi}_1, \cdots, \tilde{\xi}_n)$ is $\tilde{T}$-admissible, we can define a mapping again denoted by π of $\tilde{T}$-admissible n-blocks or sequences to T-admissible n-blocks or sequences defined by $\pi(\cdots, \tilde{\xi}_k, \cdots) = (\cdots, \pi\tilde{\xi}_k, \cdots)$, $k \in Z$. The mapping π of $(\tilde{T})$ to (T) is obviously continuous and satisfies $\pi\tilde{\sigma} = \sigma\pi$.

$\tilde{T}$ has the following feature, the proof of which is immediate from its construction: for (j, β) if $i \rightarrow j$ then there is a unique α such that $(i, \alpha) \rightarrow (j, \beta)$. Thus given a T-admissible one-sided left infinite sequence $(\cdots, \xi_{-1}, \xi_0)$ and $\tilde{\xi}_0 \in \mathscr{S}$ such that $\pi\tilde{\xi}_0 = \xi_0$ there exists a unique $\tilde{T}$-admissible one-sided sequence

$(\cdots \tilde{\xi}_{-1}, \tilde{\xi}_0)$ such that $\pi\tilde{\xi}_m = \xi_m$ $m \leqq 0$. We shall call this feature the *conditional left resolving property* of π. The term "conditional" describes the fact that $(\cdots \tilde{\xi}_{-1}, \tilde{\xi}_0)$ is not determined by $(\cdots, \xi_{-1}, \xi_0)$ until $\tilde{\xi}_0$ is specified. One consequence of the resolving property is that π is onto: for given any two-sided T-admissible $\xi = (\cdots, \xi_{n-1}, \xi_n, \cdots)$ we can find for each n a one-sided $\tilde{T}$-admissible sequence $(\cdots \tilde{\xi}_{n-1}, \tilde{\xi}_n)$ such that $\tilde{\xi}_n \in \pi^{-1}\xi_n$. By a diagonal argument we can find $\tilde{\xi} \in (\tilde{T})$ such that $\pi\tilde{\xi} = \xi$. Also, from the resolving property it is easy to see that π is everywhere finite-to-one. We shall now prove in addition that it has the asserted almost everywhere one-to-one property. We shall call any T-admissible n-block $b = (b_1, \cdots, b_n)$ *resolving for* π^{-1} if at least one component $\tilde{b}_i$ of $\tilde{b} = (\tilde{b}_1, \cdots \tilde{b}_n) \in \pi^{-1}b$ is uniquely determined. By the left resolving property of π the component $\tilde{b}_1$ will always be one of the uniquely determined ones. Let $b = (1, 2, \cdots, i_0, 1 \ldots 1, \cdots, i_0, 1)$ be a T-admissible block consisting of m repetitions of the cycle c. We shall show that b is a resolving block if m is chosen large enough. Consider the possible $\tilde{T}$-admissible blocks $\tilde{b} \in \pi^{-1}b$, i.e., $\tilde{b} = ((1, \alpha_1), (2, \alpha_2) \cdots (i_0, \alpha_{mi_0}), (1, \alpha_{mi_0+1}))$. There are ν_1 choices for α_{mi_0+1}. If $\nu_1 = 1$ we are done. If $\nu_1 > 1$, from the way $\tilde{T}$ was constructed there are fewer choices for α_{mi_0} than for α_{mi_0+1}: actually there are ν_1/λ or $[\nu_1/\lambda] + 1$ choices depending on whether λ divides ν, or not. Continuing in this manner each step to the left cuts down the number of alternatives for the second coordinate of the components of $\tilde{b}$ until we are down to a single possibility. Therefore if m is large enough ($m > [\nu_1/\lambda] + 1$ will more than suffice) then $(1, \alpha_1) = (1, 1)$. By the left resolving property of π for every sequence $\xi \in (T)$ in which b occurs infinitely often to the right there exists a unique $\tilde{\xi} \in (\tilde{T})$ such that $\pi\tilde{\xi} = \xi$. The cylinder set $B = \{\xi : \xi\,|_1^{mi_0+1} = b\}$ is open hence will have positive measure. From the ergodic theorem the subset $\mathcal{N}$ of (T) of ξ for which b does not appear infinitely often to the right, i.e., $\mathcal{N} = \{\xi \mid \sigma^n\xi \in B$ for only finite number of $n \geqq 0\}$, has measure zero with respect to any ergodic shift invariant measure which is positive on open sets. The same is true for $\tilde{\mathcal{N}} = \pi^{-1}\mathcal{N}$. The irreducibility of (T) and $(\tilde{T})$ insures the existence of ergodic shift invariant probability measures which are positive on open sets.

4. The road coloring problem

Let $\mathcal{S} = \{1, \cdots, N\}$ be a symbol set and T an $N \times N$ transition matrix with row sum λ (an integer). We shall call a set of maps ρ_i, $i = 1, 2, \cdots, \lambda$ of $\mathcal{S}$ into $\mathcal{S}$ a *road coloring for* T if $(a, \rho_1(a)), \cdots, (a, \rho_\lambda(a))$ are distinct T-admissible two-blocks for each $a \in \mathcal{S}$. Let $\mathcal{S}^* = \{1, 2, \cdots, \lambda\}$ and T^* be the $\lambda \times \lambda$ transition matrix having all entires equal to 1. Associated with a road coloring is a mapping ρ of (T) to (T^*) defined by $\rho(\cdots,\xi_n,\xi_{n+1}\cdots)=(\cdots,\xi^*_n,\xi^*_{n+1}\cdots)$ where $\rho_{\xi^*_n}(\xi_n)=\xi_{n+1}$. It is easy to see that ρ has what can be called a *conditional right resolving property*; i.e., given a T^*-admissible one-sided right infinite sequence $(\xi^*_1,\xi^*_2,\cdots)$ and ξ_1 there is a unique one-sided sequence $(\xi_1,\xi_2,\cdots)$ such that $\rho(\xi_1,\xi_2,\cdots)=(\xi^*_1,\xi^*_2,\cdots)$. We say a road coloring $\{\rho_1,\cdots,\rho_\lambda\}$ is *resolving* if there exists an n-tuple $(\rho_{i_1},\cdots,\rho_{i_n})$ of maps from $\{\rho_1,\cdots,\rho_\lambda\}$ such that $\rho_{i_n}\circ\cdots\circ\rho_{i_2}\circ\rho_{i_1}\mathcal{S}$ consists of a single element of $\mathcal{S}$.

LEMMA 2. *If T is an irreducible transition matrix with row sum λ and has a resolving road coloring $\{\rho_1,\cdots,\rho_\lambda\}$ then the associated map ρ of (T) to (T^*) is*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfies

iv) $\rho\sigma=\sigma^*\rho$

and

v) *maps $(T)-\mathcal{N}$ one-to-one onto $(T^*)-\mathcal{N}^*$ where $\mathcal{N}$ and $\mathcal{N}^*$ are shift invariant sets of measure zero with respect to any of the existent shift invariant ergodic probability measures which are positive on open sets.*

PROOF. The argument for (i), (ii), (iii) and (iv) is the same as for π in Lemma 1. To establish (v) we can take $(\rho_{i_1},\cdots\rho_{i_n})$ to be an n-tuple of maps such that $\rho_{i_n}\cdots\rho_{i_1}\mathcal{S}=\{1\}$. Let $b^*=(i_1,\cdots,i_n)$. Then b^* is resolving for ρ^{-1} because $b_{n+1}=1$ for any T-admissible $b=(b_1\cdots b_{n+1})$ where $\rho_{i_j}(b_j)=b_{j+1}$, $1\leqq j\leqq n$.

By the right resolving property of ρ for every $\xi^*\in(T^*)$ for which b^* occurs infinitely often to the left there is a unique $\xi\in(T)$ such that $\rho(\xi)=\xi^*$. The remainder of the proof is a repetition of the final argument of Lemma 1.

LEMMA 3. *If T and $\bar{T}$ are two irreducible transition matrices having row sum λ (an integer) and resolving road colorings $\{\rho_1,\cdots,\rho_\lambda\}$ and $\{\bar{\rho},\cdots,\bar{\rho}_\lambda\}$ respectively then there exists a $\hat{T}$ also with row sum λ and mappings π and $\bar{\pi}$ of $(\hat{T})$ into (T) and $(\bar{T})$ respectively such that π, $\bar{\pi}$ are*

i) *onto,*

ii) *finite-to-one,*

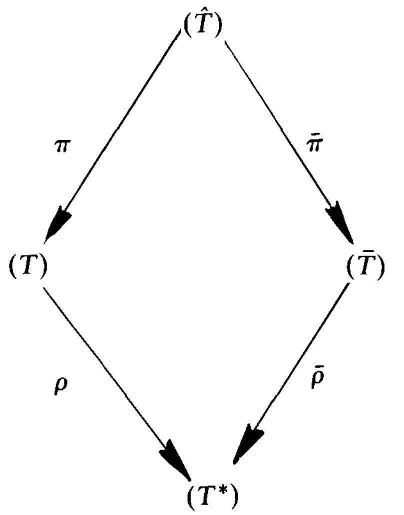
iii) *continuous,*

satisfy

iv) $\pi\hat{\sigma}=\sigma\pi$, $\bar{\pi}\hat{\sigma}=\bar{\sigma}\bar{\pi}$

and are

v) *one-to-one onto after removing from the three spaces shift invariant sets of measure zero with respect to any of the existent shift invariant ergodic probability measures which are positive on open sets.*



PROOF. Let $\hat{\rho}_1, \cdots, \hat{\rho}_\lambda$ denote mappings of $\mathcal{S} \times \bar{\mathcal{S}}$ into $\mathcal{S} \times \bar{\mathcal{S}}$ defined by $\hat{\rho}_i(u, \bar{u}) = (\rho_i(u), \bar{\rho}_i(\bar{u}))$, $1 \leqq i \leqq \lambda$. Let T' denote a transition matrix on $\mathcal{S}' = \mathcal{S} \times \bar{\mathcal{S}}$ defined by the transitions

$$\omega \rightarrow \hat{\rho}_1(\omega), \cdots, \hat{\rho}_\lambda(\omega)$$

where $\omega = (u, \bar{u}) \in \mathcal{S} \times \bar{\mathcal{S}}$. T' has row sum λ. Because the road colorings are resolving and the transition matrices are irreducible we can assume that there exist an m-tuple $(i_1, \cdots, i_m)$ and an n-tuple $(j_1, \cdots, j_n)$ such that $\rho_{i_m} \cdots \rho_{i_1} \mathcal{S} = \bar{\rho}_{j_n} \cdots \bar{\rho}_{j_1} \bar{\mathcal{S}} = \{1\}$. Let $a = \rho_{j_n} \cdots \rho_{j_1}(1)$. Then $\hat{\rho}_{j_n} \cdots \hat{\rho}_{j_1} \hat{\rho}_{i_m} \cdots \hat{\rho}_{i_1}(\omega) = (a, 1)$ for any $\omega \in \mathcal{S} \times \bar{\mathcal{S}}$. There exists a T'-cycle: namely $(a, 1) \rightarrow \hat{\rho}_{i_1}(a, 1) \rightarrow \cdots \rightarrow \hat{\rho}_{j_n} \cdots \hat{\rho}_{i_1}(a, 1) = (a, 1)$.

Let $\hat{\mathcal{S}}$ denote the nonempty irreducible component of $\mathcal{S}'$ containing $(a, 1)$ and $\hat{T}$ the transition matrix defined on $\hat{\mathcal{S}}$ with transitions inherited from T'. $\hat{T}$ also has row sum λ. In addition $\hat{\rho}_1, \cdots, \hat{\rho}_\lambda$ defines a resolving road coloring for $\hat{T}$. We define the natural projections π, $\bar{\pi}$ by $\pi(u, \bar{u}) = u$ and $\bar{\pi}(u, \bar{u}) = \bar{u}$. The proof that these mappings are finite-to-one, onto, continuous and commute with the shifts is the same as in the previous lemmas. The proof of the one-to-one property is also similar. Let us just mention that π and $\bar{\pi}$ are right resolving. So are the maps ρ and $\bar{\rho}$ of (T) and $(\bar{T})$ onto (T^*) associated with the road colorings as described in the previous lemma. The T^*-admissible $(m+n)$-block $b^* =$

$(i_1, \cdots, i_m j_1, \cdots, j_n)$ is resolving for both ρ^{-1} and $\bar{\rho}^{-1}$ since $\rho^{-1}b^*|_{m+n} = a$ and $\bar{\rho}^{-1}b^*|_{m+n} = 1$. In addition $\rho^{-1}b^*$ and $\bar{\rho}^{-1}b^*$ are resolving for π and π^{-1}, because

$$\pi^{-1}\rho^{-1}b^*|_{m+n} = \bar{\pi}^{-1}\bar{\rho}^{-1}b^*|_{m+n} = (a, 1).$$

PROPOSITION. *If T is irreducible and there exists a resolving road coloring then T is aperiodic.*

PROOF. We can assume that $\rho_{i_n} \cdots \rho_{i_1} \mathcal{S} = \{1\}$. Thus

$$1 \to \rho_{i_1}(1) \to \rho_{i_2}\rho_{i_1}(1) \to \cdots \to \rho_{i_n} \cdots \rho_{i_1}(1) = 1$$

is a cycle of length $n+1$ and

$$1 \to \rho_{i_1}(1) \to \rho^2_{i_1}(1) \to \cdots \to \rho_{i_n} \cdots \rho_{i_1}(\rho_{i_1}(1)) = 1$$

is a cycle of length $n+2$. Hence 1 is the greatest common divisor of cycle lengths minus one. Conversely we state the following.

CONJECTURE. *If T is aperiodic then there exists a resolving road coloring.*

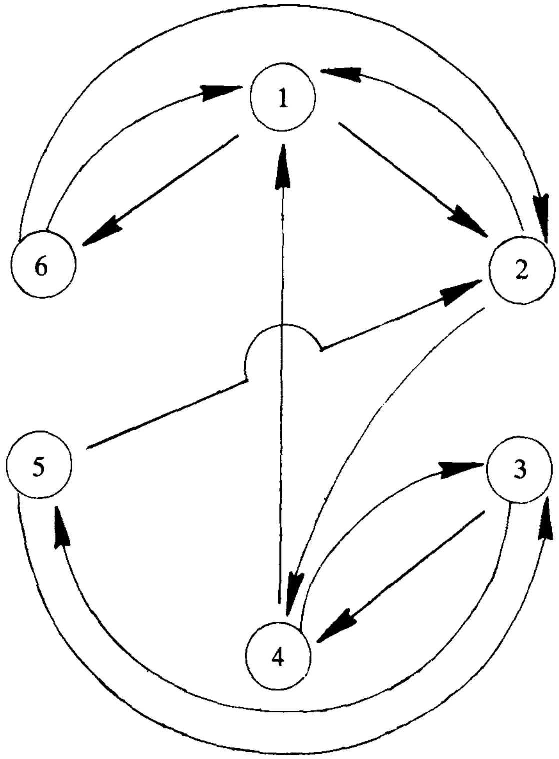
Many examples have been tested but a counter example for the conjecture as well as a proof has so far failed to show up.

We shall now explain the origin of some of the above terminology. We imagine $\mathcal{S}$ to be a set of cities and T to define a system of one-way roads connecting them. Each city has λ exit roads (a city having a road leading to itself is not excluded). The highway department has λ colors with which to paint the λ different roads leading from each city. The road from city i to j is colored K if $\rho_K(i) = j$. If a resolving road coloring exists then the roads can be painted so that a motorist upon calling the National Automobile Club to inquire how to get to city 1 is not asked where he is presently at but merely told to follow a certain finite sequence of colors. Motorists from each city will simultaneously arrive at city 1 folowing this sequence. True, some may have passed through city 1 several times but that is besides the point.

Perhaps working an example at this point would be helpful. Let T be given by the following transitions:

$$\begin{aligned} 1 &\to 2, 6 \\ 2 &\to 1, 4 \\ 3 &\to 4, 5 \\ 4 &\to 1, 3 \\ 5 &\to 2, 3 \\ 6 &\to 1, 2 \end{aligned}$$

Here $\lambda = 2$ and it is easy to verify that T is aperiodic. Picture this as 6 cities connected by one-way roads



Let the roads

$$1 \rightarrow 2$$
$$2 \rightarrow 1$$
$$3 \rightarrow 4$$
$$4 \rightarrow 1$$
$$5 \rightarrow 2$$
$$6 \rightarrow 1$$

be colored red and the others blue. Then

	Red		Red		Blue		Red	
1	→	2	→	1	→	6	→	1
2	→	1	→	2	→	4	→	1
3	→	4	→	1	→	6	→	1
4	→	1	→	2	→	4	→	1
5	→	2	→	1	→	6	→	1
6	→	1	→	2	→	4	→	1

5. A method bypassing the unsolved conjecture

Given transition matrix T with symbol set $\mathcal{S}$ and assuming $n > 1$ let us denote by $\mathcal{S}^{(n)}$ the set of T-admissible n-blocks and by $T^{(n)}$ the transition matrix for $\mathcal{S}^{(n)}$ defined by $u = (u_1, \cdots, u_n) \to v = (v_1, \cdots, v_n)$ if $u\,|_2^n = v\,|_1^{n-1}$. For $n = 1$ we set $\mathcal{S}^{(1)} = \mathcal{S}$ and $T^{(1)} = T$.

$(T^{(n)})$ is obviously homeomorphic to (T) and the natural homeomorphism ϑ defined by

$$\vartheta(\cdots u\,|_{-1}^{n-2}, \cdots, u\,|_0^{n-1}, u\,|_1^{n}, \cdots) = (\cdots u_{-1}, u_0, u_1, \cdots)$$

satisfies the commuting relation between the shifts on $(T^{(n)})$ and (T). If T has row sum λ then so does $T^{(n)}$. Every road color of T gives rise to one for $T^{(n)}$ but $T^{(n)}$ has many more. For this reason it becomes easier with large enough n to solve the conjecture of Section 4.

LEMMA 4. *If T has row sum λ and $T^n > 0$, then there exists a road coloring $\{\rho_1, \cdots, \rho_\lambda\}$ of $T^{(2n)}$ such that $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}\,\mathcal{S}^{(2n)}$ consists of a single element of $\mathcal{S}^{(2n)}$.*

Let C be a simple cycle, that is one containing no subcycles. Relabeling if necessary we can assume $C = (1, 2, \cdots, i_0, 1)$. It will be convenient to let (C) denote the element of (T) consisting of infinite repetitions of the cycle C with $(C)\,|_0 = 1$. In $T^{(k)}$ there is a special cycle denoted $C^{(k)}$ which corresponds to C in T, namely

$$C^{(k)} = ((C)\,|_0^{k-1}, (C)\,|_1^{k}, \cdots, (C)\,|_0^{k+i_0-1}).$$

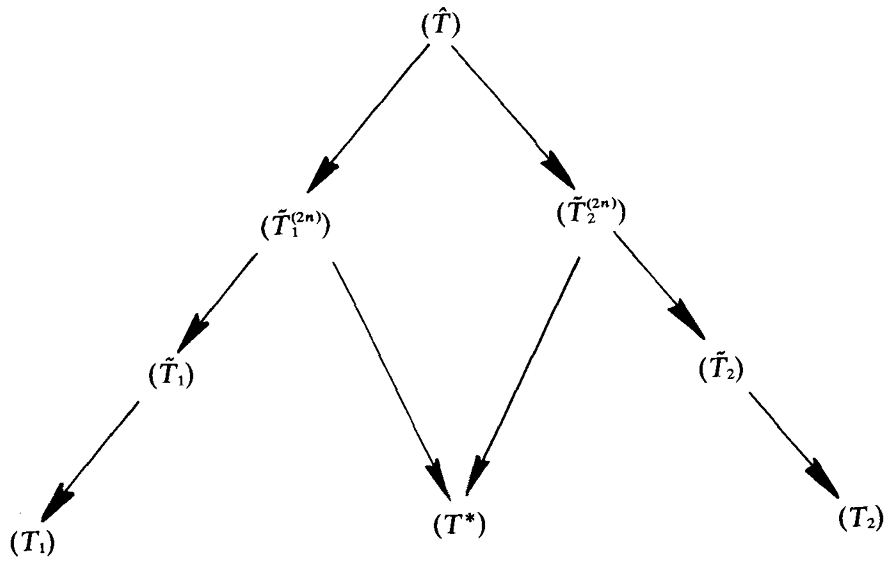
An element $u \in \mathcal{S}^{(k)}$ is said to be in $C^{(k)}$, and with a slight abuse of notation let us write $u \in C^{(k)}$, if $u = (C)\,|_j^{j+k-1}$ for some j. For $u \in \mathcal{S}^{(2n)}$ let $l(u)$ be defined as the largest l such that $u\,|_1^l \in C^{(l)}$ if $u\,|_1 \in C^{(1)}$ or $l(u) = 0$ if $u\,|_1 \notin C^{(1)}$. $l(u)$ measures how much of the tail of u lies in C. From the hypothesis $T^n > 0$ for every $a \in \mathcal{S}$ there is a $p(a) \in \mathcal{S}^{(n)}$ such that $p(a)\,|_1 = a$ and $p(a)\,|_n = 1$. $p(a)$ is the "path" from a to 1. In order to "color" $T^{(2n)}$ we give three rules for choosing $\rho_1(u)$ for each $u \in \mathcal{S}^{(2n)}$. The remaining $\rho_2(u), \cdots, \rho_\lambda(u)$ are chosen arbitrarily from the other v for which $u \to v$. I. If $l(u) = 2n$, i.e., $u = (C)\,|_j^{2n+j-1} \in C^{(2n)}$ then we define $\rho_1(u) = (C)\,|_{j+1}^{2n+j}$. II. If $l(u) \leqq n$ then either $u\,|_{2n}$ is k steps away from C and $\rho_1(u)$ is chosen so that $\rho_1(u)\,|_{2n}$ is $k-1$ steps away or $u\,|_{2n} = (C)\,|_j$ and $\rho_1(u)$ is chosen so that $\rho_1(u)\,|_{2n} = (C)\,|_{j+1}$. III. If $n < l(u) < 2n$ and $u\,|_{l(u)+1}^{2n} = p(u\,|_{l(u)+1})\,|_1^{2n-l(u)}$ then we choose $\rho_1(u)$ so that $\rho_1(u)\,|_{2n} = p(u\,|_{l(u)+1})\,|_{2n-l(u)+1}$. On the other hand if $n < l(u) < 2n$ and $u\,|_{l(u)+1}^{2n} \neq p(u\,|_{l(u)+1})\,|_1^{2n-l(u)}$ then $\rho_1(u)$ is chosen arbitrarily. Rule I keeps elements of $C^{(2n)}$ in it. Rule II sends elements not in $C^{(2n)}$ "closer" to it. Rule III continues to direct $u\,|_{2n}$ on the n-step path from $u\,|_{l(u)+1} \notin C$ to 1 if $u\,|_{l(u)+1}^{2n}$ is already on it. The mapping $\rho_1 : \mathcal{S} \to \mathcal{S}$ has the following four easily verified properties:

1) $u \in C^{(2n)} \Rightarrow \rho_1(u) \in C^{(2n)}$;
2) $u|_{n+1}^{2n} \in C^{(n)} \Rightarrow \rho_1^n(u) \in C^{(2n)}$;
3) $u|_n^{n+1} \notin C^{(2)} \Rightarrow \rho_1^{n-1}(u)|_{2n} \in C$;
4) $u|_n^{n+1} \notin C^{(2)}$ and $u|_{2n} \in C \Rightarrow \rho_1^{n-1}(u)|_{n+1}^{2n} \in C^{(n)}$.

We shall now prove $\rho_1^{4n-3}(u) \in C^{(n)}$. There are two cases. First if $u|_{n+1}^{2n} \in C^{(n)}$ we apply (2) and then (1) to get $\rho_1^{4n-3}(u) \in C^{(2n)}$. In the second case there is an i, $0 \leqq i \leqq n-1$ such that $u|_{n+i}^{n+i+1} \notin C^{(2)}$. Therefore $\rho_1^i(u)|_n^{n+1} \notin C^{(2)}$. Now there are two subcases. In the first $\rho_1^i(u)|_{2n} \in C$. We apply (4) to get $\rho_1^{n-1+i}(u)|_{n+1}^{2n} \in C^{(n)}$ and then (2) to get $\rho_1^{2n+i-1}(u) \in C^{(2n)}$. In the second case $\rho_1^i(u)|_{2n} \notin C$. We apply (3) to get $\rho_1^{n+i-1}(u)|_{2n} \in C$, then (4) to get $\rho_1^{2n+i-2}(u)|_{n+1}^{2n} \in C^{(n)}$, and finally (2) to get $\rho^{3n+i-2}(u) \in C^{(2n)}$. In either of these subcases we can use (1) to get $\rho_1^{4n-3}(u) \in C^{(2n)}$. At this stage all elements of $\rho_1^{4n-3}\mathcal{S}^{(2n)}$ are in $C^{(2n)}$ but they may be out of "phase". To get them in phase we proceed as follows. From (1) and the definition of road coloring we have that $\rho_2\rho_1^{4n-3}(u) \notin C^{(2n)}$. Here $l(\rho_2\rho_1^{4n-3}(u)) = 2n-1$ with $\rho_2\rho_1^{4n-3}(u)|_{2n-1}^{2n} \notin C^{(2)}$. Repeated applications of Rule III yields $\rho_1^{n-1}\rho_2\rho_1^{4n-3}(u)|_{2n} = 1$ with $\rho_1^{n-1}\rho_2\rho_1^{4n-3}(u)|_n^{n+1} \notin C^{(2)}$. We apply (3) and then (2) to get $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}(u) \in C^{(2n)}$ with $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}(u)|_1 = 1$. There is only one element $v \in C^{(2n)}$ with $v|_1 = 1$: namely, $v = \rho_1^{3n-2}\rho_2\rho_1^{4n-3}(u) = (1, 2, \cdots, (C)|_{2n-1})$.

6. Conclusion

The proof of the main theorem is merely a combination of the four lemmas. The following diagram explains all.



Addendum (*Added in proof January* 23, 1977)

The purpose of this addendum is to remove the restriction of aperiodicity in the main theorem.

THEOREM. *Let T_1 and T_2 be two irreducible transition matrices of period p with λ = integer as common maximal characteristic value. Then there exists an irreducible T_3 with the same period and maximal characteristic value λ and mappings π_1 and π_2 of (T_3) into (T_1) and (T_2) respectively such that π_1 and π_2 are*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfy

iv) $\pi_1 \sigma_3 = \sigma_1 \pi_1$, $\pi_2 \sigma_3 = \sigma_2 \pi_2$,

and are

v) *one-to-one after removing from (T_1), (T_2) and (T_3) shift invariant sets* having measure zero with respect to any of the existent shift invariant ergodic probability measures which are positive on open sets.

We shall abbreviate the last participle phase in (v) to "*having maximal entropy measure zero*". This terminology is more than mere convenience and the reader can consult W. Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–66, for additional background material on it.

We shall now proceed with modifications needed to prove the theorem. Corresponding lemmas will be numbered as before.

PROOF OF THEOREM. An irreducible transition matrix T has period p if and only if it has the following property (P):

The symbol set $\mathcal{S} = \{1, \cdots, N\}$ can be partitioned into p disjoint sets $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{p-1}$ such that the transition $i \to j$ can only occur if $i \in \mathcal{S}_k$ and $j \in \mathcal{S}_{k+1 (\mathrm{mod}\, p)}$. And furthermore T^p acts on $\mathcal{S}_k$ aperiodically: that is, there exists an m such that for every k, $0 \leqq k \leqq p-1$, if $i, j \in \mathcal{S}_k$ then $t_{ij}^{(mp)} > 0$.

LEMMA 1. *Let T be an irreducible $N \times N$ transition matrix with maximal positive characteristic value λ = integer. Then there exists an $\tilde{N} \times \tilde{N}$ transition matrix $\tilde{T}$ with row sum λ and a mapping π of $(\tilde{T})$ into (T) such that π is*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfies

iv) $\pi\sigma = \tilde{\sigma}\pi$,

and is

v) *a one-to-one map of* $(\tilde{T}) - \tilde{\mathcal{N}}$ *onto* $(T) - \mathcal{N}$ *where* $\mathcal{N}$ *and* $\tilde{\mathcal{N}}$ *are shift invariant sets of maximal entropy measure zero.*

Furthermore if T *has period* p *so does* $\tilde{T}$.

PROOF. The first statement has already been proved in §3. The second follows from the fact that $\tilde{T}$ inherits property (P) from T by virtue of the fact that π is a one-block map satisfying (iv).

Next we let $\mathcal{S}^* = \{(i,j): 1 \leqq i \leqq \lambda, 0 \leqq j \leqq p-1\}$ and define the canonical row sum λ period p transition matrix T^* by the transitions $(i,j) \to (i',j')$ for all i, i' whenever $j' = j + 1 \pmod{p}$. The symbols of $\mathcal{S}^*$ can be ordered so that T^* takes the form of a $\lambda p \times \lambda p$ matrix

$$T^* = \begin{bmatrix} [0] & [1] & & --- & & [0] \\ [0] & [0] & [1] & --- & & [0] \\ & & & --- & & \\ [0] & & & --- & [0] & [1] \\ [1] & & & --- & & [0] \end{bmatrix}$$

where $[0]$ is a $\lambda \times \lambda$ matrix of zeros and $[1]$ a $\lambda \times \lambda$ matrix of ones. For arbitrary T with row sum λ and period p for which $\mathcal{S} = \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_{p-1}$ according to property (P) we can define a mapping $\rho: (T) \to (T^*)$ given a road coloring $(\rho_1, \cdots, \rho_\lambda)$ by $\rho(\cdots \xi_n, \xi_{n+1}, \cdots) = (\cdots \xi_n^*, \xi_{n+1}^* \cdots)$ where $\xi_n^* = (i_n^*, j_n^*) \in \mathcal{S}^*$ such that $j_n^* = k$ if $\xi_n \in \mathcal{S}_k$ and i_n^* is determined by $\rho_{i_n^*}(\xi_n) = \xi_{n+1}$. Furthermore we say $\{\rho_1, \cdots, \rho_\lambda\}$ is *resolving* if there exists an n-tuple $(\rho_{i_1}, \cdots, \rho_{i_n})$ of maps from $\{\rho_1, \cdots, \rho_\lambda\}$ such that $\rho_{i_n} \circ \cdots \circ \rho_{i_1} \mathcal{S}_k$ consists of a single element for at least one of the $\mathcal{S}_k$. Actually once this happens for one $\mathcal{S}_k$ it can be arranged to happen for all.

LEMMA 2. *If* T *is an irreducible row sum* λ *transition matrix with period* p *and has a resolving road coloring* $\{\rho_1, \cdots, \rho_\lambda\}$ *then the associated map* ρ *of* (T) *to* (T^*) *is*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfies

iv) $\rho\sigma = \sigma^*\rho$,

and

v) *maps* $(T)-\mathcal{N}$ *one-to-one onto* $(T^*)-\mathcal{N}^*$ *where* $\mathcal{N}$ *and* $\mathcal{N}^*$ *are shift invariant sets of maximal entropy measure zero.*

PROOF. Suppose $\rho_{i_n}\cdots\rho_{i_1}\mathcal{S}_k$ consists of a single element, say 1. Let $b^* = ((i_1, j_1), \cdots, (i_n, j_n))$ where $j_i = i + k - 1 \pmod p$. Then b^* is resolving for ρ^{-1} because $b_{n+1} = 1$ for any T-admissible $b = (b_1, \cdots, b_{n+1})$ where $b_1 \in \mathcal{S}_k$ and $\rho_{i_j}(b_j) = b_{j+1}, 1 \leqq j \leqq n$. The rest of the proof is the same as in §4.

LEMMA 3. *If* T *and* $\bar{T}$ *are two row sum* λ *irreducible transition matrices having period* p *and resolving road colorings* $\{\rho_1, \cdots, \rho_\lambda\}$ *and* $\{\bar{\rho}_1, \cdots, \bar{\rho}_\lambda\}$ *respectively, then there exists* $\hat{T}$ *also irreducible with row sum* λ *and period* p *and mappings* π *and* $\bar{\pi}$ *of* $(\hat{T})$ *to* (T) *and* $(\bar{T})$ *respectively such that* π, $\bar{\pi}$ *are*

i) *onto,*

ii) *finite-to-one,*

iii) *continuous,*

satisfy

iv) $\pi\hat{\sigma} = \sigma\pi$, $\pi\hat{\sigma} = \bar{\sigma}\bar{\pi}$,

and are

v) *one-to-one onto after removing from the three spaces sets of maximal entropy measure zero.*

PROOF. Suppose $\rho_{i_m}\cdots\rho_{i_1}\mathcal{S}_k$ and $\bar{\rho}_{j_n}\cdots\bar{\rho}_{j_1}\bar{\mathcal{S}}_{\bar{k}}$ consist of single elements. By irreducibility we can assume the two elements are labelled the same, say by 1. We can also assume $k = \bar{k} = 0$ and $\mathcal{S}_0$ and $\bar{\mathcal{S}}_0$ both contain 1. This means we must take $m = n = 0 \pmod p$. Let $a = \rho_{j_n}\cdots\rho_{j_1}(1) \in \mathcal{S}_0$. Then $\hat{\rho}_{j_n}\cdots\hat{\rho}_{j_1}\hat{\rho}_{i_m}\cdots\hat{\rho}_{i_1}(w) = (a, 1)$ for $w = (u, \bar{u})$, $u \in \mathcal{S}_0$, $\bar{u} \in \bar{\mathcal{S}}_0$. We see that $\{\hat{\rho}_1\cdots\hat{\rho}_\lambda\}$ is resolving for $\hat{T}$ by noting that $\hat{T}$ satisfies property (P) with $\hat{\mathcal{S}} = \hat{\mathcal{S}}_0 \cup \cdots \cup \hat{\mathcal{S}}_{p-1}$ where $\hat{\mathcal{S}}_k = \hat{\mathcal{S}} \cap (\mathcal{S}_k \times \bar{\mathcal{S}}_k)$. The remainder of the proof is the same as in §4.

LEMMA 4. *If* T *is irreducible with row sum* λ *and period* p, *then there exists a resolving road coloring* $\{\rho_1, \cdots, \rho_\lambda\}$ *of* $T^{(2n)}$ *where* $n = mp$ *with* m *defined by property* (P).

PROOF. Let C be a simple cycle as in §5. We can assume the symbols of $\mathcal{S}$ are labelled so that $C = (1, 2, \cdots, p, p+1, \cdots, i_0, 1)$ and $C|_k \in \mathcal{S}_{k-1 \pmod p}$ where $\mathcal{S}$ is partitioned by $\mathcal{S} = \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_{p-1}$ according to property (P). Because $n = mp$ we can define for each $a \in \mathcal{S}_k, 0 \leqq k \leqq p-1$, an element $p(a) \in \mathcal{S}^{(n)}$ such that $p(a)|_1 = a$ and $p(a)|_n = k + 1$. The rules for coloring $T^{(2n)}$ are the same as in §5.

We get $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}\mathcal{S}^{(2n)} \subseteq C^{(2n)}$ and $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}(u)|_1 = k+1 \;+(7n-5+1)$ $(\operatorname{mod} p)$ whenever $u|_1 \in \mathcal{S}_k$ for $u \in \mathcal{S}^{(2n)}$. For each k there is only one $v \in C^{(2n)}$ satisfying $v|_1 = k + 7n - 5 \pmod{p}$. Thus $\mathcal{S}^{(2n)}$ can be partitioned according to property (P) by $\mathcal{S}^{(2n)} = \mathcal{S}_0^{(2n)} \cup \cdots \cup \mathcal{S}_{p-1}^{(2n)}$ where $\mathcal{S}_k^{(2n)} = \{u \in \mathcal{S}^{(2n)} : u|_1 \in \mathcal{S}_k\}$ and $\rho_1^{3n-2}\rho_2\rho_1^{4n-3}\mathcal{S}_k^{(2n)}$ consists of a single element for every k.

Bibliography

1. R. Adler and B. Weiss, *Similarity of automorphisms of the torus*, Mem. Amer. Math. Soc., No. 98 (1970).

2. B. Weiss, *The structure of Bernoulli systems*, Proc. of Int. Congress of Math., Vancouver, Vol. 2, 1974, pp. 321–325.

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598 USA

University of Kentucky
Lexington, Kentucky 40506 USA

and

The Hebrew University of Jerusalem
Jerusalem, Israel.