Mh. Math. 88, 239—247 (1979)

Monatshefte für
Mathematik

Factors and Extensions of Full Shifts

By

Brian Marcus*, Chapel Hill, N. C.

(Received 17 November 1978)

Abstract. Let Σ_A be an irreducible shift of finite type with entropy $\log n$. Then Σ_A is a continuous extension of the full n-shift. Also, if Σ_A is a continuous factor of the full n-shift, then it is shift equivalent to the full n-shift.

1. Introduction

Let Σ_A be an irreducible shift of finite type of entropy $\log n$ where n is a positive integer (see Section 2 for definitions). In this paper we show (Theorem 5) that Σ_A is topologically conjugate to an irreducible shift of finite type all of whose row sums and column sums equal n. From this and a result of ADLER, GOODWYN and WEISS ([1]) it follows (Theorem 6) that Σ_A is a finite-to-one, 1-to-1 a.e., continuous extension of the full n-shift. We also show (Theorem 8) that if Σ_A is a continuous factor of the full n-shift, then A is shift-equivalent to the full n-shift in the sense of WILLIAMS ([8]). It is a well-known conjecture and at least in this case we believe that shift equivalence implies topological conjugacy (it is implied by topological conjugacy). In any event, it does at least imply that Σ_A has the same number of periodic points of all orders (namely n^k) as the full n-shift so that the zeta functions of Σ_A and the full n-shift are the same. And it does imply that sufficiently high powers of Σ_A are topologically conjugate to the corresponding powers of the full n-shift. Thus, in a sense, the full n-shift is the "minimal" shift of finite type with entropy $\log n$.

We hope that our view point will be useful in the problem of deciding whether shift equivalence is indeed equivalent to topological conjugacy. In particular, see Theorem 7.

* Partially supported by NSF grants MCS 76-05969 and MCS 78-01244.

0026—9255/79/0088/0239/$01.80

We thank ROY ADLER, RUFUS BOWEN, and ETHAN COVEN for useful discussions regarding these problems.

2. Background

Let A be an $l \times l$ matrix of 0's and 1's. Let

$$\Sigma_A = \{x \in \{1, \ldots, l\}^Z : A_{x_i x_{i+1}} = 1 \text{ for each } i \in Z\}.$$

Let $\sigma : \Sigma_A \to \Sigma_A$ denote the left shift:

$$\sigma(x) = \mathfrak{y} \text{ where } y_i = x_{i+1}.$$

The map, σ, is known as a shift of finite type often referred to simply as Σ_A. The set $\{1, \ldots, l\}$ is the alphabet or "symbols" of Σ_A.

One says that i is a predecessor of j (or that j is a successor of i) if $A_{ij} = 1$ and we denote this $i \to j$. Let $S(i)$ denote the set of successors of i. A p-block (or allowable word) is a word $a_1 \ldots a_p$, $a_i \in \{1, \ldots, l\}$, $A_{a_i a_{i+1}} = 1$ for $i = 1, \ldots, p-1$. Sometimes a p-block $a_1 \ldots a_p$ also refers to the cylinder set $\{x : x_i = a_i, 1 \leqslant i \leqslant p\}$.

The usual notion of conjugacy is topological conjugacy: Σ_A and Σ_B are topologically conjugate if there is a homeomorphism from Σ_A onto Σ_B which commutes with the shifts. Two important invariants of topological conjugacy are

a) Transitivity – i. e., there exists a point x whose forward orbit $\{\sigma^n(x)\}_{n \geqslant 1}$ is dense. In such a case it is well-known that we may assume that A is a irreducible matrix i. e., $\forall i, j \in \{1, \ldots, l\} \exists n$ s. t. $A^n_{ij} > 0$. From this it follows from the beautiful Perron-Frobenius theorem ([4]) that among the largest (in modulus) eigenvalues there is one (called $\lambda = \lambda(A)$) which is simple and positive. Moreover, λ has a positive eigenvector and is uniquely determined as the only positive eigenvalue with a positive eigenvector.

b) Topological Entropy. The topological entropy of Σ_A is simply $\log(\lambda)$. (The topological entropy of an arbitrary homeomorphism has an abstract definition and turns out to be $\log \lambda$ in this case — see [6], [9].)

There is also the notion of measure theoretic entropy [9] of a map relative to a shift invariant Borel probability measure. PARRY ([6]) proved that there is a unique such measure μ_A whose entropy is $\log \lambda$, and μ_A is constructed as follows: Let $r > 0$ be a right (column) eigenvector and $s > 0$ be a left (row) eigenvector corresponding to λ.

Normalize r such that $\sum_{i=1}^{l} r_i s_i = 1$. If $a_1 \dots a_p$ is a p-block, then

$$\mu_A(a_1 \dots a_p) = \frac{s_{a_1} r_{a_p}}{\lambda^{p-1}}.$$

A continuous factor map $\pi: \Sigma_B \to \Sigma_A$ is a continuous map which commutes with the shifts. One says that Σ_A is a continuous factor of Σ_B (or Σ_B is a continuous extension of Σ_A) if there exists such a map which is also onto. The well-known theorem of Curtis—Hedlund—Lyndon ([5]) asserts that (modulo composition with a power of the shift) every continuous factor map $\pi: \Sigma_A \to \Sigma_B$ is of the following form:

$\exists$ a positive integer k and a map π^*: $\{k$-blocks of $\Sigma_B\} \to \{1$-blocks of $\Sigma_A\}$ such that

$$\pi\left((b_i)_{i=-\infty}^{\infty}\right) = \left(\pi^*(b_{i+1}, \dots, b_{i+k})\right)_{i=-\infty}^{\infty}.$$

Note that π^* in a natural way defines a map (also called π^*) from the $(k+p-1)$-blocks of Σ_B to the p-blocks of Σ_A.

Also, we shall use $(1 \Rightarrow 3)$ of the following

Theorem (Coven-Paul [3]): *Let Σ_B and Σ_A be irreducible shifts of finite type with $\lambda(A) = \lambda(B)$. Let $\pi: \Sigma_B \to \Sigma_A$ be a continuous factor map. Then the following are equivalent*

1) *π is onto,*

2) *π is boundedly finite-to-one,*

3) *π is measure preserving (i.e., $\mu_A(C) = \mu_B(\pi^{-1}(C))$ for all measurable sets $C \subset \Sigma_A$).*

Finally we mention the notion of shift equivalence ([8]). Let A and B be square matrices with non-negative integer entries. One says that A is shift equivalent to B if there exists a positive integer k and non-negative integer valued matrices R, S such that

$$A^k = RS,\ B^k = SR,\ AR = RB,\ SA = BS.$$

Shift equivalence is an equivalence relation [8]. Also, if A and B are 0-1 matrices and Σ_A is topologically conjugate to Σ_B, then A and B are shift equivalent ([8]). The converse is conjectured to be true.

Let J_n be the $n \times n$ matrix consisting of all 1's. Then Σ_{J_n} is simply the full shift on n symbols. Note that J_n is shift equivalent to the 1×1 matrix (n).

By a constant matrix (or vector) we simply mean an integer-valued matrix (or vector) all of whose entries are the same.

In this paper we are interested in the case when $\lambda(A)$ is an integer. It is easy to see that in this case the corresponding eigenvectors can be chosen to be positive and integral. Since the corresponding eigenspace is 1-dimensional, the notion of a smallest positive integer-valued right eigenvector makes sense. We will call this vector: the smallest right eigenvector. Note that the row sums of A are all equal to $\lambda(A)$ if and only if the smallest right eigenvector is $(1, \ldots, 1)$.

We shall occasionally refer to (forward or backward) one-sided shifts of finite type (see [8] for more about this notion).

3. A Doubly Stochastic Model

Proposition 1: *Let A be an irreducible 0-1 matrix with $\lambda(A) = n \in Z^+$. Then $\exists$ an irreducible 0-1 matrix B such that Σ_A is topologically conjugate to Σ_B and*

a) each row-sum of B is n,

b) the set of column sums of B is the same as the set of column sums of A.

Remarks: 1) The conjugacy is in fact a conjugacy on the forward one-sided level.

2) In general the dimension of B will be much larger than that of A (in fact the dimension of B will be the sum of the components in the smallest right eigenvector corresponding to n).

3) It was previously known ([1]) that one could get such a Σ_B which can be seen to be an at most 2-to-1 continuous extension of Σ_A. Their method involves "filling in a tableau" and our method (although different) essentially shows that one can "fill in the tableau" so that their factor map is actually a homeomorphism.

As we shall see, Proposition 1 will follow from

Proposition 2: *Let A as in proposition 1. Let $(r_1, \ldots, r_l)$ be the smallest right eigenvector corresponding to n and assume (by a permutation) $r_l \geqslant r_i$, $i = 1, \ldots, l$. If $(r_1, \ldots, r_l) \neq (1, \ldots, 1)$ then there is an irreducible $(l+1) \times (l+1)$ 0-1 matrix B such that Σ_A is topologically conjugate to Σ_B and*

a) the smallest right eigenvector of B corresponding to n is $(r_1, \ldots, r_{l-1}, e_1, e_2)$ where $e_1, e_2 < r_l$;

b) the set of column sums of B is the same as the set of column sums of A.

Note that proposition 2 improves the smallest right eigenvector; it produces another matrix whose smallest right eigenvector has either a smaller maximal component or has the same maximal component but with one less repetition. Thus, by repeated application proposition 2 will yield proposition 1.

Proof of Proposition 2: The alphabet for Σ_B will consist of equivalence classes (soon to be specified) of the 2-blocks of Σ_A satisfying $ij \sim i'j' \Rightarrow i = i'$. The equivalence class of ij is denoted $[ij]$. Given such equivalence classes, one has the natural transitions

$$[ij] \to [i'j'] \text{ iff } ij \sim i\,i'.$$

The associated transition matrix B is easily seen to be irreducible and the 1-block map defined by $\pi^*([ij]) = i$ is a topological conjugacy from Σ_B onto Σ_A. This is an example of an elementary equivalence in the sense of Williams ([8]). By a straight forward computation one has

Lemma 3: *The vector v whose components are*

$$v_{[ij]} = \sum_{j' \in M} r_{j'}\,,\ M = \{j' \in S(i) : ij \sim ij'\}$$

is a right eigenvector of B corresponding to n.

Note that by definition each predecessor of an element $[i'j'] \in B$ is of the form $[i\,i']$. Hence the B-predecessors of $[i'j']$ are in 1-to-1 correspondence with the A-predecessors of i'. This means that the set of column sums of B will be the same as the set of column sums of A.

After a little discussion, we shall specify the equivalence classes.

We first show that $\exists\, i^*, j^* \in \{1, \ldots, l\}$ such that $i^* \to j^*$, $r_{i^*} = r_l$, and $r_{j^*} < r_l$. If not, then one cannot escape from $\{i \in \{1, \ldots, l\} : r_i = r_l\}$ whence by irreducibility of A we would have each $r_i = r_l$. But since $(r_1, \ldots, r_l)$ is the smallest right eigenvector corresponding to n, we would have $(r_1, \ldots, r_l) = (1, \ldots, 1)$ contrary to assumption.

So, such i^*, j^* must exist and (by a permutation) we may assume $i^* = l$.

Now, note

$$\sum_{j \in S(l)} r_j = n\, r_l.$$

Also for each $j \in S(l)$ $r_j \leqslant r_l$, $j^* \in S(l)$ and $r_{j^*} < r_l$. Thus, $\# S(l) > n$.
Now we claim

Lemma 4: *There is a nonempty subset $E \subset S(l)$ such that*

i) $\# E \leqslant n$ *and*
ii) $\sum_{j \in E} r_j \equiv 0 \bmod n$.

Proof: To see this, first pick n arbitrary elements $j_1, \ldots, j_n$ in $S(l)$. Now, if $\sum_{k=1}^{p} r_{j_k}, p = 1, \ldots, n$ are all distinct $\bmod n$ then one must be congruent to 0 $\bmod n$. Otherwise, two of the sums must coincide $(\bmod n)$ and so $\exists\, 1 \leqslant q < p \leqslant n$ such that $\sum_{k=q}^{p} r_{j_k} \equiv 0 \bmod n$. Thus, in either case we can find such (in fact many) a subset E. □

Now, the equivalence classes are as follows: For $i \neq l$ $\{ij\}_{j \in S(l)}$ forms one entire class, denoted simply $[i]$.

For $i = l$, we have two classes,

$$[l_1] = \{ij\}_{j \in E} \text{ (as in lemma 4)}, \; [l_2] = \{ij\}_{j \in S(l) \sim E}.$$

Note that $[l_2]$ is indeed nonempty since $\# S(l) > n$ and $\# E \leqslant n$. Also, note that we have simply split the symbol l into two pieces.

Now by lemma (3) we have

$$v_{[i]} = \sum_{j \in S(i)} r_j = n r_i \text{ for } i \neq l,$$

$$v_{[l_1]} = \sum_{j \in E} r_j,$$

$$v_{[l_2]} = \sum_{j \in S(l) \sim E} r_j.$$

But this means that each component of the vector v is divisible by n: this is clear for $i \neq l$; true for $[l_1]$ by the choice of E (lemma 4); and true for $[l_2]$ since

$$v_{[l_2]} = \Big(\sum_{j \in S(l)} r_j\Big) - v_{[l_1]} = n r_l - v_{[l_1]}. \tag{*}$$

Thus, dividing each component of v by n we obtain a smaller right eigenvector corresponding to n whose components are

$$\left\{ r_1, r_2, \ldots, r_{l-1}, \frac{v_{[l_1]}}{n}, \frac{v_{[l_2]}}{n} \right\}$$

Since by (*) $\frac{v_{[l_1]}}{n} + \frac{v_{[l_2]}}{n} = r_l$ and both are positive, the proof is complete.

We next strengthen proposition 1.

Theorem 5: *Let A be an irreducible 0-1 matrix with $\lambda(A) = n \in Z^+$. Then $\exists$ an irreducible 0-1 matrix B such that Σ_A is topologically conjugate to Σ_B, each row sum of B is n and each column sum of B is n.*

Proof: First apply Proposition 1 to A^t to obtain a matrix B_1 with Σ_{A^t} top conjugate to Σ_{B_1} and each row sum of B_1 is n. Since a conjugacy between two maps is a conjugacy between their inverses we have Σ_A conjugate to $\Sigma_{B_1^t}$ and each column sum of B_1^t is n. Now, apply Proposition 1 to B_1^t.

4. Factors and Extensions

Theorem 6: *The full n-shift is a continuous, boundedly finite-to-one, 1-to-1 a.e., factor of every irreducible shift of finite type, Σ_A whose entropy is* $\log n$.

Note: This is also valid on either one-sided level.

Proof: (This is an idea of ADLER and WEISS.) By Theorem 5, we may assume that each row sum of A is n. So, we can define a block map π^* from the 2-blocks of A to the symbols $\{1, \ldots, n\}$ in such a way that π^* is 1-to-1 when restricted to each set of the form $\{ij : j \in S(i)\}$. One easily sees that this defines a continuous factor map π from Σ_A onto Σ_{J_n}. It is also easy to see that is boundedly finite-to-one (alternatively one can refer to the result of COVEN and PAUL mentioned in section 2). In general, it will not be 1-to-1 a. e. But in [1] it is shown that such a π can be chosen to be 1-to-1 a. e.

In our discussion of factors of the full n-shift we shall need (3 $\Rightarrow$ 2) of the following.

Theorem 7: *Let A be an irreducible shift of finite type with $\lambda(A) = n$. The following are equivalent:*

1) *n is the only non-zero eigenvalue of A,*

2) *A is shift equivalent to J_n,*

3) *Σ_A is topologically conjugate to some Σ_B where B^k is a constant matrix for some k.*

Proof: $1 \Rightarrow 2$. See [8, section 8].

$2 \Rightarrow 3$. By Theorem 5 we can choose B with all row and column sums equal to n (and Σ_B topologically conjugate to Σ_A). Then B will be shift equivalent to the 1×1 matrix n. This means that there exists a right eigenvector R and left eigenvector S for B (corresponding to the eigenvalue n) and also $k > 0$ such that

$$B^k = RS, \quad n^k = SR.$$

But S and R are constant vectors — so B^k is a constant matrix.

$3 \Rightarrow 1$. This is obviously true for B and therefore true for A since the non-zero eigenvalues determine the numbers of periodic points of all orders ([2]) and therefore is an invariant of topological conjugacy.

Theorem 8: *If Σ_B is a continuous factor of Σ_{J_n} and $\lambda(B) = n$, then B is shift equivalent to J_n.*

Proof: Since transitivity is preserved under continuous factors, we may assume that B is irreducible. Since $\lambda(B) = n$ we may (by theorem 5) assume that B has all its row and column sums equal to n; so, its right and left eigenvectors are constant vectors. By the construction in section 2, this means that the measure of maximal entropy μ_B assigns equal measure to blocks of equal length — namely the μ_B-measure of a p-block is $\frac{1}{dn^{p-1}}$ (assuming B is a $d \times d$ matrix). Also, the measure of maximal entropy μ_{J_n} assigns mass $\frac{1}{n^q}$ to each q-block in Σ_{J_n}. Note that the result of COVEN—PAUL in section 2 implies that $\mu_{J_n} = \pi^{-1}\mu_B$ where $\pi: \Sigma_{J_n} \to \Sigma_B$ is the factor map.

Now, we may assume by the Curtis—Hedlund—Lyndon theorem (section 2) that π is a k-block map for some k. So, the inverse image of a p-block in Σ_B is the disjoint union of certain $(k + p - 1)$-blocks in Σ_{J_n}.

But by the remarks in the preceding paragraph one sees that the number of such $(k + p - 1)$-blocks is

$$\frac{n^{(k+p-1)}}{dn^{p-1}} = \frac{n^k}{d}.$$

Thus, the number of $(k + p - 1)$-blocks in the inverse image of a block in Σ_B is the same for all blocks in Σ_B (and in particular is independent of the length of the block) (cf. [5, theorem 5.4]).

By virtue of theorem 7 ($3 \Rightarrow 2$) it is sufficient to show that B^k is a constant matrix. Specifically, we will show $B^k = \frac{n^k}{d} \cdot J_d$.

To see this, we must show that given $b, b' \in \{1, \ldots, d\}$ the number of words $b = b_0 \ldots b_k$ in Σ_B with $b_0 = b$ and $b_k = b'$ is $\frac{n^k}{d}$. Consider the map

$$\varphi: (\pi^*)^{-1}(b) \times (\pi^*)^{-1}(b') \to \{b_0 \ldots b_k \text{ in } \Sigma_B \text{ with } b_0 = b \text{ and } b_k = b'\}$$

defined by

$$\varphi(x_0 \ldots x_{k-1}, y_0 \ldots y_{k-1}) = \pi^*(x_0 \ldots x_{k-1} y_0 \ldots y_{k-1})$$

Now, by the remarks above we know that the cardinality of the domain of φ is $\left(\frac{n^k}{d}\right)^2$. Moreover, the mapping is onto and each element of the range has exactly $\frac{n^k}{d}$ inverse images. Thus, the cardinality of the range is $\frac{n^k}{d}$.

So, we have shown that B^k is the constant matrix $\frac{n^k}{d} \cdot J_d$ as desired.

References

[1] Adler, R., W. Goodwyn, and B. Weiss: Equivalence of topological Markov shifts. Israel J. Math. 27, 49—63 (1977).

[2] Bowen, R., and O. Lanford: Zeta functions of restrictions of the shift transformation. Proc. Symp. Pure Math. vol. 14, p. 43—50. Providence, R. I.: Amer. Math. Soc. 1970.

[3] Coven, E., and M. Paul: Endomorphisms of irreducible shifts of finite type. Math. Syst. Th. 8, 167—175 (1974).

[4] Gantmacher, F.: Theory of Matricies, vol. 2. New York: Chelsea. 1959.

[5] Hedlund, G. A.: Endomorphisms and automorphisms of the shift dynamical system. Math. Syst. Th. 3, 320—375 (1969).

[6] Parry, W.: Intrinsic Markov chains. Trans. Amer. Math. Soc. 112, 55—66 (1964).

[7] Parry, W., and R. F. Williams: Block Coding and a Zeta Function for Finite Markov Chains. Preprint.

[8] Williams, R. F.: Classification of shifts of finite type. Ann. Math. 98, 120—153 (1973); Errata. Ann. Math. 99, 380—381 (1974).

[9] Walters, P.: Ergodic Theory—Introductory Lectures. Lecture Notes Math. 458. Berlin—Heidelberg—New York: Springer. 1975.

B. Marcus
Department of Mathematics
University of North Carolina
Chapel Hill, NC 27514, U.S.A.