Ergod. Th. & Dynam. Sys. (1984), **4**, 283–300
Printed in Great Britain

The entropies of topological Markov shifts and a related class of algebraic integers

D. A. LIND

Department of Mathematics, University of Washington, Seattle, Washington 98195, *USA*

(*Received* 6 *October* 1983)

Abstract. We give an algebraic characterization of the class $\mathbb{P}$ of spectral radii of aperiodic non-negative integral matrices, and describe a method of constructing all such matrices with given spectral radius. The logarithms of the numbers in $\mathbb{P}$ are the entropies of mixing topological Markov shifts. There is an arithmetic structure to $\mathbb{P}$, including factorization into irreducibles in only finitely many ways. This arithmetic structure has dynamical consequences, such as the impossibility of factoring the p-shift into a direct product of nontrivial homeomorphisms for prime p.

1. *Introduction*

Topological Markov shifts play a central role in ergodic theory and topological dynamics. Their use in analysing the dynamics of geodesic flows goes back to Hadamard, and was developed by Morse [**15**]. Shannon [**17**] used such shifts as models for discrete communication systems, and proved a forerunner of the variational principle for topological entropy. Later Parry [**16**] independently established the full variational principle for topological Markov shifts, ultimately leading to its formulation and proof for general continuous mappings. The discovery by Adler and Weiss [**2**] of Markov partitions, and hence Markov shift covers, for automorphisms of the two-dimensional torus led to their classification. This idea was developed by Bowen [**4**] and Sinai [**18**] into a powerful general method for analysing Anosov and Axiom A diffeomorphisms.

The most significant numerical invariant for a topological Markov shift is its topological entropy. For aperiodic shifts this was first computed by Shanon, under the name channel capacity, to be $\log \lambda$, where λ is the positive dominant eigenvalue of the non-negative integral matrix defining the shift. For aperiodic shifts such a number λ must have two properties: λ must be an algebraic integer greater than or equal to 1, and λ must be strictly greater than the absolute value of its other conjugates. Since the second property arises from the Perron–Frobenius theory, call the class $\mathbb{P}$ of such algebraic integers Perron numbers.

Our principal result is that these two properties characterize the spectral radii of aperiodic nonnegative integral matrices, hence also the entropies of aperiodic

Markov shifts. By a result of Bowen [3], this characterizes the topological entropies of Axiom A diffeomorphisms as well.

The proof of this theorem contains an effective algorithm for using the companion matrix of the minimal polynomial of $\lambda \in \mathbb{P}$ to produce an aperiodic non-negative integral matrix with spectral radius λ. In § 3 we show that this technique in fact produces every such matrix.

The Perron-Frobenius theory shows that a non-negative integral matrix that is irreducible, but not necessarily aperiodic, has spectral radius some power of which is a Perron number. In § 4 we show that if λ^n is Perron, then λ is the spectral radius of an irreducible non-negative integral matrix. Using the decomposition of a general matrix into irreducible components, we deduce that the set of spectral radii of non-negative integral matrices, irreducible or not, is exactly the set of all positive roots of Perron numbers together with 0.

Perron numbers possess an arithmetic much like the natural numbers. Each can be factored into a finite product of irreducible Perron numbers, each incapable of further factorization in $\mathbb{P}$. In § 5 we prove that the number of such factorizations in finite, but non-uniqueness can occur. We also show that in every field extension of the rationals that contains a non-rational Perron number there are examples of non-unique factorization.

In § 6 we apply these results to factoring Markov shifts into direct products of homeomorphisms. A result of Bowen shows that such direct factors must themselves be Markov shifts. Since a rational prime p is irreducible in $\mathbb{P}$, it will follow that the full p-shift is not the direct product of non-trivial homeomorphisms. Also, using the isomorphism theorem of Adler and Marcus [1] we show that up to almost topological conjugacy there is a bijection between Perron factorizations of the spectral radius of the defining matrix of a Markov shift and direct product factorizations of the shift itself.

Non-uniqueness of Perron factorizations suggests looking for non-uniqueness at the zeta function or Markov shift levels. In § 7 we give one example of this, two pairs of Markov shifts whose products are shown to be shift equivalent by using a theorem of Krieger on dimension groups. From this follows an example of non-unique direct product factorization of Markov shifts.

This paper contains details and applications of work announced in [14]. It has benefitted substantially throughout by numerous stimulating conversations with Mike Boyle. His contributions are particularly important in § 5, where the discovery of the crucial proposition 5 on Perron factorizations, a method for producing non-unique factorizations used in the proof of theorem 5, and other ideas are due to him. The author also gratefully acknowledges the support of NSF grant MCS 8201542.

2. *Markov shift entropies*

Let A be a non-negative integral matrix. A well-known extension [19] of the zero-one matrix construction associates to A a homeomorphism σ_A of a Cantor set called a topological Markov shift. If $A^n > 0$ for some positive integer n, then A is called

aperiodic. This condition on A is equivalent to σ_A being topologically mixing, in which case we also say σ_A is aperiodic.

Throughout the rest of this paper, 'Markov shift' will mean a topological Markov shift that is assumed to be aperiodic unless otherwise stated, and 'matrix' will mean a non-negative integral matrix.

The Perron–Frobenius theory for non-negative matrices [**10**] implies that an aperiodic matrix A has a largest positive eigenvalue λ_A which strictly exceeds the absolute value for all other eigenvalues of A. Thus λ_A is also the spectral radius of A. Furthermore λ_A satisfies the characteristic polynomial $\chi_A(t)$ of A, which is monic with integer coefficients, so λ_A is an algebraic integer. The minimal polynomial of λ_A over $\mathbb{Q}$ divides $\chi_A(t)$, so the other algebraic conjugates of λ_A are eigenvalues of A and hence have absolute value strictly less than λ. Since the product of λ_A with its conjugates is an integer, it follows that $\lambda_A \geq 1$.

Let $\mathbb{P}$ denote the class of Perron numbers defined in § 1. This definition differs slightly from that in [**14**] by allowing 1 to be Perron. The above shows that if A is an aperiodic matrix, then $\lambda_A \in \mathbb{P}$. The following proves that every Perron number arises this way.

THEOREM 1. *If λ is a Perron number, then there is an aperiodic non-negative integral matrix with spectral radius λ.*

Proof. The proof begins by using the d-dimensional companion matrix B of the minimal polynomial of λ over $\mathbb{Q}$ to decompose $\mathbb{R}^d$ into invariant subspaces. These spaces are used to construct an invariant convex region Ω for B. The basic idea is to use the geometry of this region to find integral points $z_1, \ldots, z_n \in \Omega$ such that

$$Bz_j = \sum_{i=1}^{n} a_{ij} z_i \qquad \text{with } a_{ij} \in \mathbb{Z}^+.$$

The z_i and a_{ij} are chosen so that every irreducible component of $A = [a_{ij}]$ has positive trace, implying that each component is aperiodic. Replacing A by such a component if necessary, an application of the Perron–Frobenius theory and the spectral radius formula shows that $\lambda_A = \lambda$.

Let $\lambda \in \mathbb{P}$ with degree d over $\mathbb{Q}$. Let the minimal polynomial of λ over $\mathbb{Q}$ be $f(t) = t^d - c_1 t^{d-1} - \cdots - c_d$, where each $c_j \in \mathbb{Z}$. Then

$$B = \begin{bmatrix} 0 & 0 & \cdots & c_d \\ 1 & 0 & \cdots & c_{d-1} \\ 0 & 1 & \cdots & c_{d-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_1 \end{bmatrix}$$

is the companion matrix of $f(t)$, which of course can contain negative entries.

Since $f(t)$ is irreducible, it has no repeated roots. The real Jordan form [**11**] for B shows that $\mathbb{R}^d$ splits into the direct sum of three classes of B-invariant subspaces. The first consists of the single 1-dimensional dominant eigenspace D for λ. Fix a non-zero $w \in D$. We sometimes identify D with $\mathbb{R}$ by identifying rw with r. The

second class $\mathscr{E}$ contains eigenspaces E corresponding to conjugates of λ strictly outside the unit circle. Then $\dim E$ is 1 if the conjugate is real, and 2 otherwise. The third class $\mathscr{F}$ contains those 1- or 2-dimensional eigenspaces F of conjugates of λ with absolute value ≤ 1.

There are norms on these subspaces so that

$$\|Bx\|_D = \lambda \|x\|_D, \qquad x \in D,$$
$$\|Bx\|_E = \tau_E \|x\|_E, \qquad x \in E,\ 1 < \tau_E < \lambda,$$
$$\|Bx\|_F = \tau_F \|x\|_F, \qquad x \in F,\ \tau_F \leq 1.$$

It is convenient from now on to norm $\mathbb{R}^d$ with the maximum of these norms.

If G represents one of the subspaces above, let $\pi_G: \mathbb{R}^d \to G$ be the projection to G along the complementary direct sum. Let $\pi_C = I - \pi_D$ be projection to the invariant complement C of D. The identification of D with $\mathbb{R}$ mentioned above means we shall sometimes use π_D as mapping to $\mathbb{R}$ and speak of $\pi_D(x)$ as the D-coordinate of x.

We next construct a B-invariant convex region. An element $x \in \oplus_{\mathscr{E}} E$ will be denoted by $\Sigma_E x_E$, where $x_E = \pi_E x$. Let $p(E) = \log \lambda / \log \tau_E$ for $E \in \mathscr{E}$. Clearly $p(E) > 1$. Define $\Phi: \oplus \mathscr{E} E \to \mathbb{R}$ by

$$\Phi\left(\sum_E x_E\right) = \sum_E \|x_E\|^{p(E)}.$$

For fixed ξ, $\eta > 0$, consider the region

$$\Omega = \Omega_{\xi,\eta} = \left\{ x \in \mathbb{R}^d : \max_F \|\pi_F x\| \leq \xi, \quad \Phi\left(\sum_E x_E\right) \leq \eta \pi_D x \right\}.$$

Since $p(E) > 1$ for every E, the region Ω is bowl-shaped, tangent to C at the origin, and curved towards D.

We claim Ω is B-invariant. The basic reason for this is that Φ has an invariant graph. Let $x \in \Omega$. Then

$$\max_F \|\pi_F Bx\| = \max_F \tau_F \|\pi_F x\| \leq (\max_F \tau_F)\xi \leq \xi,$$

$$\Phi\left(\sum_E \pi_E Bx\right) = \sum_E \|\pi_E Bx\|^{p(E)} = \sum_E (\tau_E)^{p(E)} \|\pi_E x\|^{p(E)}$$

$$= \lambda \Phi\left(\sum_E \pi_E x\right) \leq \lambda \eta \pi_D(x) = \eta \pi_D(Bx),$$

verifying $B\Omega \subset \Omega$.

We shall be concerned with representing integral points as non-negative integral combinations of a fixed set of integral points. If $S \subset \mathbb{R}^d$, let $\mathrm{sg}(S)$ denote the additive semigroup generated by S.

For $\theta > 0$, define the cone K_θ in $\mathbb{R}^d$ about D by $K_\theta = \{x: \pi_D x \geq \theta \|\pi_C x\|\}$. For r, $s > 0$ define

$$K_\theta(r) = \{x \in K_\theta: \pi_D \leq r\}, \qquad K_\theta(r, s) = \{x \in K_\theta: r \leq \pi_D x \leq s\}.$$

The following lemma shows that the semigroup generated by the integral points in a truncated cone contains all the integral points in a slimmer cone.

LEMMA 1. *Fix $\theta > 0$. For all sufficiently large r,*

$$K_{2\theta} \cap \mathbb{Z}^d \subset \mathrm{sg}(K_\theta(r) \cap \mathbb{Z}^d).$$

Proof of the lemma. We first claim that if $\delta = (2\theta+2)^{-1}$ and $x \in K_{2\theta}$ with $\pi_D x > 4$, then $[x - K_\theta(1,3)] \cap K_{2\theta}$ contains a ball of radius δ. For suppose $\pi_D x = s > 4$ and put

$$y = \frac{2}{s}(x + \tfrac{1}{2}\pi_C x) = \frac{2}{s}\pi_D x + \frac{3}{s}\pi_C x.$$

We show the ball of radius δ centred at $x - y$ satisfies our claim. Suppose $\|u\| < \delta$. Then

$$\begin{aligned} 2\theta\|\pi_C(x-y+u)\| &\le 2\theta\left[\left(1-\frac{3}{s}\right)\|\pi_C x\| + \delta\right] \\ &\le \left[1-\frac{3}{s}\right]\pi_D x + \frac{2\theta}{2\theta+2} = \left[1-\frac{2}{s}\right]\pi_D x - \frac{2}{2\theta+2} \\ &\le \pi_D(x-y+u), \end{aligned}$$

proving $x - y + u \in K_{2\theta}$. Since

$$2\theta\|\pi_C y\| = \frac{3}{s}(2\theta)\|\pi_C x\| \le \frac{3}{s}\pi_D x = 3,$$

we have

$$\begin{aligned} \theta\|\pi_C(y-u)\| \le \theta(\|\pi_C y\| + \delta) &\le \tfrac{3}{2} + \frac{\theta}{2\theta+2} \\ &= 2 - \frac{1}{2\theta+2} \le \pi_D(y-u), \end{aligned}$$

so $y - u \in K_\theta$. Also, since $\delta < 1$, it follows that $1 \le \pi_D(y-u) \le 3$. These inequalities establish our claim.

The lemma is proved inductively. First choose ρ such that any ball in $\mathbb{R}^d$ of radius ρ intersects $\mathbb{Z}^d$. Choose r so that $r\delta > \rho$. The first claim implies that if $x \in K_{2\theta}$ with $\pi_D x > 4r$, then $[x - K_\theta(r, 3r)] \cap K_{2\theta}$ contains a ball of radius $r\delta > \rho$, hence intersects $\mathbb{Z}^d$.

Let $\Gamma = \mathrm{sg}(K_\theta(4r) \cap \mathbb{Z}^d)$. We show $K_{2\theta} \cap \mathbb{Z}^d \subset \Gamma$. Clearly $K_{2\theta}(4r) \cap \mathbb{Z}^d \subset \Gamma$. Suppose $K_{2\theta}(t) \cap \mathbb{Z}^d \subset \Gamma$ for some $t \ge 4r$; we show this forces $K_{2\theta}(t+r) \cap \mathbb{Z}^d \subset \Gamma$, which suffices by induction.

Let $z \in [K_{2\theta}(t+r) \setminus K_{2\theta}(r)] \cap \mathbb{Z}^d$. By the above, there is an element $y \in \mathbb{Z}^d$ contained in $[z - K_\theta(r, 3r)] \cap K_{2\theta}(t)$. Hence $y \in \Gamma$ by hypothesis, and $y = z - x$ for some $x \in K_\theta(r, 3r) \cap \mathbb{Z}^d \subset \Gamma$. Therefore $z = x + y \in \Gamma + \Gamma \subset \Gamma$, concluding the proof of the lemma. □

We now possess the pieces necessary to prove theorem 1. Begin by fixing $\theta > 0$. By the lemma, there is an $r > 0$ so that $K_{2\theta} \cap \mathbb{Z}^d \subset \mathrm{sg}(K_\theta(r) \cap \mathbb{Z}^d)$.

Next find $\xi, \eta > 0$ so that $K_\theta(r) \subset \Omega_{\xi,\eta} = \Omega$. The following estimates show this is possible. If $x \in K_\theta(r)$, then

$$r \ge \pi_D x \ge \theta\|\pi_C x\| \ge \theta\|\pi_F x\|, \qquad F \in \mathscr{F}.$$

Hence choose $\xi = r/\theta$. Since $r \geq \pi_D x \geq \theta \|\pi_E x\|$ for $E \in \mathscr{E}$,

$$\sum_E \|\pi_E x\|^{p(E)} \leq \frac{\pi_D(x)}{\theta} \sum_E \|\pi_E x\|^{p(E)-1} \leq \left[\frac{1}{\theta} \sum_E \left(\frac{r}{\theta}\right)^{p(E)-1}\right] \pi_D x,$$

so let η be the bracketed expression on the right. Note the role $p(E) > 1$ plays.

For $s > 0$ define

$$\Omega(s) = \{x \in \Omega: \pi_D x \leq s\}, \qquad \Omega(s, \infty) = \{x \in \Omega: s \leq \pi_D x\}.$$

We next show that if s is sufficiently large, then $(B - I)\Omega(s, \infty) \subset K_{2\theta}$. Suppose $x \in \Omega(s, \infty)$. Then $\pi_D(Bx - x) = (\lambda - 1)\pi_D x$ while

$$\|\pi_E(Bx - x)\| \leq (\tau_E + 1)\|\pi_E x\|.$$

Since

$$\|\pi_E x\|^{p(E)} \leq \Phi\left(\sum_E x_E\right) \leq \eta \pi_D x,$$

we have since $p(E) > 1$ that

$$\frac{\pi_D(Bx - x)}{\|\pi_E(Bx - x)\|} \geq \frac{(\lambda - 1)\pi_D x}{(\tau_E + 1)(\eta \pi_D x)^{1/p(E)}} \geq \left[\frac{\lambda - 1}{(\tau_E + 1)\eta^{1/p(E)}}\right] s^{1 - 1/p(E)} \geq 2\theta,$$

for large enough s. Similarly, $\|\pi_F(Bx - x)\| \leq (\tau_F + 1)\|\pi_F x\| \leq (\tau_F + 1)\xi$, so

$$\frac{\pi_D(Bx - x)}{\|\pi_F(Bx - x)\|} \geq \frac{(\lambda - 1)\pi_D x}{(\tau_F + 1)\xi} \geq \left[\frac{\lambda - 1}{(\tau_F + 1)\xi}\right] s \geq 2\theta$$

for sufficiently large s, establishing the claim.

Now fix $s > r$ so that $(B - I)\Omega(s/\lambda, \infty) \subset K_{2\theta}$. Let $\Gamma = \Omega(s) \cap \mathbb{Z}^d$, and order the elements of Γ as $z_1, \ldots, z_n$. The following procedure specifies a rule for writing each Bz_j as a non-negative integral combination

$$Bz_j = \sum_{i=1}^{n} a_{ij} z_j. \tag{1}$$

If $\pi_D(z_j) \leq s/\lambda$, then $Bz_j = z_k \in \Gamma$, and put $a_{ij} = \delta_{ik}$. If $s/\lambda < \pi_D(z_j) \leq s$, then

$$Bz_j - z_j \in K_{2\theta} \subset \operatorname{sg}[K_\theta(r) \cap \mathbb{Z}^d] \subset \operatorname{sg}(\Gamma),$$

so choose the a_{ij} in (1) with $a_{jj} \geq 1$. This process yields a matrix $A = [a_{ij}]$.

If A is reducible, it has an irreducible component [**10**]. This amounts to selecting a minimal subset Γ_0 of Γ for which (1) holds with $z_i, z_j \in \Gamma_0$. Replace A by this component, so A is now irreducible and indexed by Γ_0. We claim $\operatorname{tr} A \geq 1$ persists. Let $z_j \in \Gamma_0$. There is a unique m so that $s/\lambda < \pi_D(B^m z_j) \leq s$. By the procedure above, we must have $z_j, Bz_j, \ldots, B^m z_j \in \Gamma_0$, and if $B^m z_j = z_k$ then $a_{kk} = 1$, proving $\operatorname{tr} A \geq 1$. Thus A is aperiodic since it is irreducible and has positive trace.

Finally, let $\lambda_A = \mu$. We show $\mu = \lambda$. Let A be n-dimensional. If e_i is the ith unit vector in $\mathbb{R}^n$, define $P: \mathbb{R}^n \to \mathbb{R}^d$ by $P(e_i) = z_i$. Then (1) shows that $PA = BP$.

Since A is non-negative and aperiodic, the Perron–Frobenius theory shows it has a positive eigenvector v for μ. Now Pv is a positive combination of the z_j, and $\pi_D(z_j) > 0$, so $\pi_D(Pv) > 0$ and therefore $Pv \neq 0$. Also,

$$B(Pv) = P(Av) = \mu(Pv),$$

so μ is an eigenvalue of B. Thus $\mu \leq \lambda$.

To prove $\mu \geq \lambda$, we first prove P is surjective. For this it suffices to show that a B-invariant subspace of $\mathbb{R}^d$ containing a non-zero integral point z must be all of $\mathbb{R}^d$. If not, there is a polynomial $g(t) \in \mathbb{Q}[t]$ with $\deg g < d$ and $g(B)z = 0$. Since the characteristic polynomial $f(t)$ of B is irreducible of degree d, it is relatively prime to $g(t)$. Thus there are $a(t)$, $b(t) \in \mathbb{Q}[t]$ with $a(t)f(t) + b(t)g(t) \equiv 1$, so that

$$0 = [a(B)f(B) + b(B)g(B)]z = z.$$

This contradiction establishes surjectivity of P. Now choose $u \in \mathbb{R}^n$ with $Pu = w$, where w is an eigenvector for B corresponding to λ. Then

$$\begin{aligned} \lambda^n \|w\| &= \|B^n w\| = \|B^n P u\| \\ &= \|PA^n u\| \leq \|P\| \, \|A^n\| \, \|u\|. \end{aligned}$$

The spectral radius formula for A shows that

$$\mu = \limsup_{n \to \infty} \|A^n\|^{1/n} \geq \lambda,$$

completing the proof. □

3. *A general algorithm*

The proof of theorem 1 gives a practical procedure for finding an aperiodic matrix with a specified Perron number as its spectral radius. This can be roughly described as follows. Given $\lambda \in \mathbb{P}$, decompose $\mathbb{R}^d$ under the action of the companion matrix B of λ into $C \oplus D$, where D is the 1-dimensional dominant eigendirection and C its invariant complement. Then find integral points $z_1, \ldots, z_n$ with positive D-coordinate so that

$$Bz_j = \sum_{i=1}^{n} a_{ij} z_i, \qquad a_{ij} \in \mathbb{Z}^+, \tag{2}$$

and take an irreducible component of the resulting matrix $[a_{ij}]$. The proof specifies one way to choose the z_i and a_{ij}.

As an example, consider the Perron root $\lambda \approx 3.8916$ of

$$f(t) = t^3 + 3t^2 - 15t - 46.$$

The procedure above was carried out to obtain the 10-dimensional aperiodic matrix with spectral radius λ given in [**14**]. Notice that since $f(t)$ is irreducible, a 3-dimensional matrix with spectral radius λ has trace -3, and hence could not be non-negative. Thus in general the dimension of the non-negative matrix obtained by theorem 1 must strictly exceed the degree of λ.

The following shows that *every* aperiodic non-negative integral matrix with specified spectral radius arises this way by a suitable choice of the z_i.

THEOREM 2. *Let $\lambda \in \mathbb{P}$ have degree d, and let B and D have the meaning above. If $A = [a_{ij}]$ is an n-dimensional aperiodic non-negative integral matrix with spectral radius λ, then there are $z_i \in \mathbb{Z}^d$ with positive D-coordinate such that $Bz_j = \sum_{i=1}^{n} a_{ij} z_i$.*

Proof. Since vectors are column vectors, it is notationally convenient to replace A by its transpose, and find z_i with $Bz_i = \sum_{j=1}^{n} a_{ij} z_j$.

Consider $A:\mathbb{R}^n \to \mathbb{R}^n$ and its restriction to $\mathbb{Q}(\lambda)^n$. The Perron–Frobenius theory shows there is a positive eigenvector $v \in \mathbb{R}^n$ for A corresponding to λ. By working over $\mathbb{Q}(\lambda)$ instead of $\mathbb{R}$, we can guarantee that $v \in \mathbb{Q}(\lambda)^n$, and multiplying by a positive integer, that $v \in \mathbb{Z}[\lambda]^n$. Thus there are integers z_{ij} such that $v_i = z_{i1} + z_{i2}\lambda + \cdots + z_{id}\lambda^{d-1} > 0$.

Define $\Psi:\mathbb{Q}(\lambda) \to \mathbb{Q}^d$ by

$$\Psi(r_1 + r_2\lambda + \cdots + r_d\lambda^{d-1}) = [r_1\ r_2\ \ldots\ r_d]^T.$$

Put $z_i = \Psi(v_i) \in \mathbb{Z}^d$. Since v is an eigenvector,

$$\lambda v_i = (Av)_i = \sum_{j=1}^{n} a_{ij} v_j.$$

Multiplication by λ on $\mathbb{Q}(\lambda)$ has matrix B with respect to the basis $\{1, \lambda, \ldots, \lambda^{d-1}\}$, so applying Ψ yields

$$B\Psi(v_i) = \Psi(\lambda v_i) = \sum_{j=1}^{n} a_{ij} \Psi(v_j)$$

or

$$Bz_i = \sum_{j=1}^{n} a_{ij} z_j.$$

It remains to verify that each z_i has positive D-coordinate $\pi_D(z_i)$. First note that

$$w^* = [1\ \lambda\ \cdots\ \lambda^{d-1}]$$

is a left eigenvector for B with eigenvalue λ. Linear algebra shows that

$$C = \{x \in \mathbb{R}^d : w^* x = 0\},$$

so π_D is a multiple of w^*. But $w^* z_i = v_i > 0$, so replacing every z_i by $-z_i$ if necessary, it follows that $\pi_D(z_i) > 0$ for every i. □

To illustrate theorem 2, consider $\lambda = 2$. Then to produce all matrices with spectral radius 2, it suffices to examine finite collections $\{z_1, \ldots, z_n\}$ of positive integers and the possible ways of writing $2z_j = \sum_{i=1}^{n} a_{ij} z_i$. The proof of theorem 1, together with the normal form of a reducible matrix [**10**], shows that every matrix $[a_{ij}]$ produced this way has spectral radius 2, although it may be reducible. Conversely, theorem 2 shows that every aperiodic matrix with spectral radius 2 arises this way.

Finally, we wish to point out the role of positivity of the D-coordinate of the z_j. Since all that seems to be required of the z_j is that (2) holds, it seems plausible to use the choice $z_i = e_i$, $z_{d+i} = -e_i$ for $1 \le i \le d$, where e_i is the ith unit vector of $\mathbb{R}^d$. The signs can be adjusted so that (2) holds, yielding a $2d$-dimensional matrix A. The proof in theorem 1 that $\lambda_A = \lambda$ holds until the last paragraph. The essential point is that if v is a positive eigenvector for λ_A, then positivity of $\pi_D(z_i)$ guarantees that

$$\pi_D(Pv) = \sum_{i=1}^{n} v_i \pi_D(z_i) > 0$$

and hence $Pv \neq 0$. In the method just suggested, it is possible that $Pv = 0$, and thus that $\lambda_A > \lambda$. Indeed this occurs with the cubic example above. If A denotes the

6-dimensional matrix produced, then

$$\chi_A(t) = (t^3+3t^2-15t-46)(t^3-3t^2-15t-46).$$

The first factor has maximum root $\lambda \approx 3.8916$ while for the second it is $\lambda_A \approx 6.4390$.

Indeed, this method *never* works unless the companion matrix is already non-negative itself. For suppose λ has minimal polynomial

$$f(t) = t^d - c_1 t^{d-1} - \cdots - c_d.$$

It can be shown that the $2d$-dimensional A produced as above has charactristic polynomial

$$(t^d - c_1 t^{d-1} - \cdots - c_d)(t^d - |c_1| t^{d-1} - \cdots - |c_d|).$$

An elementary argument shows that the maximum root of the second factor strictly exceeds λ unless each $c_j \geq 0$, proving our assertion.

4. *Periodic matrices*

The focus so far has been on aperiodic matrices. A weaker condition is irreducibility of A, that for every pair i, j there is an $n > 0$ with $(A^n)_{ij} > 0$. Irreducibility of A corresponds to topological transitivity of σ_A.

Perron's original paper concerned positive, hence aperiodic, matrices. Frobenius extended the theory to non-negative irreducible matrices, and analysed possible periodic phenomena. For such an A he showed there is a positive dominant eigenvalue λ_A corresponding to a positive eigenvector. Furthermore, there is a period k for A such that the spectrum of A is invariant under multiplication under the kth roots of unity, and the other eigenvalues of A with absolute value λ are the kth roots of unity times λ. The aperiodic case corresponds to $k = 1$. Since the conjugates of λ occur among the roots of $\chi_A(t)$, it follows that λ^k must be Perron.

Thus the spectral radii of irreducible matrices are positive numbers some positive power of which is Perron. An example is $\lambda = \sqrt{2}$, which is not Perron, but whose square is. The next result shows that the converse holds.

THEOREM 3. *A positive number is the spectral radius of an irreducible non-negative integral matrix if and only if some positive integral power of it is Perron.*

Proof. Although it is likely the geometric ideas from the proof of theorem 1 can be adapated here, the following proof uses only the statement of that theorem.

Let $\lambda > 0$ and $\lambda^k \in \mathbb{P}$. By theorem 1, there is an aperiodic matrix A with $\lambda_A = \lambda^k$. Let P be a cyclic permutation matrix of dimension k. Then $A \otimes P$ is irreducible, and

$$(A \otimes P)^k = A^k \otimes I = \operatorname{diag}(A, \ldots, A)$$

has spectral radius $\lambda_A = \lambda^k$. Hence $\lambda_{A \otimes P} = \lambda$. □

Recall that the matrix A produced in theorem 1 has $\operatorname{tr} A \geq 1$. Hence this gives a different proof of a result of Adler and Marcus [1] that if the greatest common divisor of the periods of an irreducible matrix is k, there is another such matrix with the same spectral radius containing an actual k-cycle.

We remark that the proof of theorem 1 breaks down if $\lambda \notin \mathbb{P}$ but $\lambda^k \in \mathbb{P}$ because the region Ω is no longer strictly curved towards the dominant eigendirection.

Specifically, the inclusion

$$(B-I)\Omega(s,\infty)\subset K_{2\theta}$$

always fails, regardless of the size of s.

Our results can be used to characterize the spectral radii of general non-negative integral matrices.

COROLLARY. *The set of spectral radii of non-negative integral matrices equals*

$$\{\lambda^{1/k}:\lambda\in\mathbb{P},\ k\geq 1\}\cup\{0\}.$$

Proof. Let A be non-negative integral. By a permutation of coordinates, A can be given the form ([**10**])

$$\begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$

where the A_{ii} are irreducible and square. Since $\lambda_A=\max\lambda_{A_{ii}}$, either λ_A is zero or some power of it is Perron. The reverse inclusion is a consequence of theorem 3. □

5. *Perron arithmetic*

In this section we study the arithmetic structure of the class $\mathbb{P}$ of Perron numbers. The principal results are that each Perron number can be decomposed into a finite product of irreducibles incapable of further decomposition, that this decomposition is not always unique although the number of decompositions is finite, and that non-unique Perron factorization occurs in every field containing a non-rational Perron number. These results are applied to Markov shifts in § 6, 7. They have also recently been applied by Boyle and Tuncel [**7**] to study infinite-to-one codings of Markov shifts.

PROPOSITION 1. $\mathbb{P}$ *is closed under addition and multiplication.*

Proof. Let $\lambda\in\mathbb{P}$ with conjugates $\lambda_1=\lambda,\lambda_2,\ldots,\lambda_m$, and $\mu\in\mathbb{P}$ with conjugates $\mu_1=\mu,\ldots,\mu_n$. Then $\lambda+\mu>1$, and its conjugates are among the numbers $\lambda_i+\mu_j$. Now

$$|\lambda_i+\mu_j|\leq|\lambda_i|+|\mu_j|\leq\lambda+\mu,$$

with equality only when $\lambda_i=\lambda$ and $\mu_i=\mu$. This proves $\lambda+\mu\in\mathbb{P}$. A similar argument shows $\lambda\mu\in\mathbb{P}$. □

PROPOSITION 2. $\mathbb{P}$ *is dense in* $[1,\infty)$.

Proof. This follows from proposition 1 if we can show that 1 is a limit point of $\mathbb{P}$. Consider the largest root λ_n of $p_n(t)=t^n-t-1$. Since the companion matrix of $p_n(t)$ is aperiodic and non-negative, $\lambda_n\in\mathbb{P}$. Now $p_n(1)=-1$, and $p_n'(t)=nt^{n-1}-1\geq n-1$ for $t\geq 1$. Thus $\lambda_n\in(1,n/n-1)$, so $\lambda_n\to 1$. □

In contrast to proposition 2, the next results show that certain algebraically defined subsets of $\mathbb{P}$ are discrete.

PROPOSITION 3. *For every* $d>1$, $\{\lambda\in\mathbb{P}:\deg\lambda\leq d\}$ *is discrete in* $[1,\infty)$.

Proof. Fix a bound $M \geq 1$. Suppose $\lambda \in \mathbb{P}$ with $\deg \lambda = r \leq d$ and $\lambda \leq M$. Let λ have conjugates $\lambda_1 = \lambda, \lambda_2, \ldots, \lambda_r$. The minimal polynomial for λ is

$$\prod_i (t - \lambda_i) = t^r - \sigma_1 t^{r-1} t + \sigma_2 t^{r-2} - \cdots + (-1)^r \sigma_r,$$

where the σ_j are the elementary symmetric functions in the λ_i. The σ_j are integers, and since $|\lambda_j| \leq \lambda$ and $r \leq d$,

$$|\sigma_j| \leq \binom{r}{j} \lambda^r \leq 2^d M^d, \qquad 1 \leq j \leq r.$$

There are only a finite number of possible polynomials, so

$$\{\lambda \in \mathbb{P}: \deg \lambda \leq d\} \cap [1, M]$$

is finite. □

PROPOSITION 4. *If J is a finite extension of $\mathbb{Q}$ then $K \cap P$ is discrete.*

Proof. If $[K:\mathbb{Q}] = d$, then $K \cap \mathbb{P} \subset \{\lambda \in \mathbb{P}: \deg \lambda \leq d\}$. □

The next result is crucial to establishing factorizations of Perron numbers into irreducibles.

PROPOSITION 5. *If $\lambda = \alpha\beta$ with $\alpha, \beta, \lambda \in \mathbb{P}$, then $\alpha, \beta \in \mathbb{Q}(\lambda)$.*

Proof. Suppose $\alpha \notin \mathbb{Q}(\lambda)$. Then the minimal polynomial $f(t)$ of α over $\mathbb{Q}(\lambda)$ has degree $d \geq 2$. Let $f(t) = \prod_{i=1}^{d} (t - \alpha_i)$ with $\alpha_1 = \alpha$. Since $f(t)$ is irreducible over $\mathbb{Q}(\lambda)$, so is

$$g(t) = t^d \prod_{i=1}^{d} \left(\frac{\lambda}{t} - \alpha_i\right).$$

Now $g(\beta) = 0$, so the conjugates of β over $\mathbb{Q}(\lambda)$ are $\beta_i = \lambda/\alpha_i$, $1 \leq i \leq d$. Since α_i and β_i are also rational conjugates of α and β, for $i \geq 2$ we have by definition of $\mathbb{P}$ that

$$\beta = \frac{\lambda}{\alpha} < \frac{\lambda}{|\alpha_i|} = |\beta_i| < \beta.$$

This contradiction proves $\alpha \in \mathbb{Q}(\lambda)$. Then $\beta = \lambda/\alpha \in \mathbb{Q}(\lambda)$ as well. □

Call $\lambda \in \mathbb{P}$ *irreducible* if $\lambda > 1$ and λ cannot be written as $\alpha\beta$ with $\alpha, \beta \in \mathbb{P}$ and $\alpha, \beta > 1$. Proposition 5 proves that a rational integer is irreducible if and only if it is prime. As a further example, we show that $\lambda = (1+\sqrt{5})/2$ is irreducible. For suppose $\lambda = \alpha\beta$ with $\alpha, \beta \in \mathbb{P}$. By proposition 5, $\alpha, \beta \in \mathbb{Q}(\sqrt{5})$. Since λ is a unit, α and β must also be units. The unit group of $\mathbb{Q}(\sqrt{5})$ is $\{\pm\lambda^n: n \in \mathbb{Z}\}$. Thus $\alpha = \lambda^m$ and $\beta = \lambda^n$. Since $\alpha, \beta \in \mathbb{P}$, it follows $m, n \geq 1$. But $\lambda = \alpha\beta$ implies $1 = m + n$, a contradiction. A similar argument shows $\lambda + 2$ is irreducible by using primality of its norm. This will be used shortly.

In the following result, Perron factorizations that differ only in the order of terms are counted as being the same.

THEOREM 4. *Every Perron number greater than 1 can be factored into a finite number of irreducibles. There are only a finite number of such factorizations, but factorization into irreducibles is not always unique.*

Proof. If $\lambda \in \mathbb{P}$ with $\lambda \neq 1$, and $\lambda = \alpha_1 \cdots \alpha_n$ with $1 \neq \alpha_j \in \mathbb{P}$, then proposition 5 shows that each $\alpha_j \in \mathbb{Q}(\lambda)$. Since $\mathbb{Q}(\lambda)$ is a finite extension of $\mathbb{Q}$, proposition 4 shows that $S = \mathbb{Q}(\lambda) \cap \mathbb{P} \cap [1, \lambda]$ is finite. Since $\alpha_j \in [1, \lambda]$, all Perron factorizations of λ use only terms from the finite set S. Thus there are only a finite number of possibilities, some of which must be irreducible factorizations. Finally, if $\lambda = (1+\sqrt{5})/2$, then 5, λ, and $\lambda+2$ are irreducible and so $5 \cdot \lambda \cdot \lambda = (\lambda+2) \cdot (\lambda+2)$ is a case of non-unique irreducible factorization. □

Not every number field has non-unique Perron factorizations since the field may not contain any non-trivial real subfields. For example, let α be a non-real root of $t^p - t - 1$, where p is prime, and let $K = \mathbb{Q}(\alpha)$. Since $t^p - t - 1$ is irreducible over $\mathbb{Q}$ (see [**13**, p. 215]), we have $[K:\mathbb{Q}] = p$. Let $L = K \cap \mathbb{R}$, and suppose $L \neq \mathbb{Q}$. Since $p = [K:L][L:\mathbb{Q}]$ and $[L:\mathbb{Q}] \geq 1$, it follows $[L:\mathbb{Q}] = p$, so $L = K$. But K is not real. This forces $L = \mathbb{Q}$. Here $K \cap \mathbb{Q} = \{2, 3, \ldots\}$ has unique Perron factorizations. If the intersection is any larger, then non-uniqueness occurs.

THEOREM 5. *Let K be a finite extension of $\mathbb{Q}$. Then the following are equivalent.*

(a) *$K \cap \mathbb{P}$ has non-unique Perron factorizations;*

(b) *$K \cap \mathbb{P}$ contains non-rational Perron numbers;*

(c) $K \cap \mathbb{R} \neq \mathbb{Q}$.

Proof. Since the rational integers have unique factorization in primes, (a) implies (b). If $\lambda \in K \cap \mathbb{P}$ and $\lambda \notin \mathbb{Q}$, then $\mathbb{Q} \neq \mathbb{Q}(\lambda) \subset K \cap \mathbb{R}$, proving that (b) implies (c).

To prove (c) implies (a), assume K is real by replacing it with $K \cap \mathbb{R}$. Let U be the multiplicative group of units of K. There will be three cases to consider: $[K:\mathbb{Q}] = 2$, rank $U \geq 2$, and $[K:\mathbb{Q}] = 3$. The first and third are handled by special arguments, while the second uses the Dirichlet unit theorem.

First suppose $[K:\mathbb{Q}] = 2$. Then $K = \mathbb{Q}(\sqrt{d})$ for a square-free integer d. Since $K \cap \mathbb{R} \neq \mathbb{Q}$, we have $d > 1$. Let u be a fundamental unit for K, so $U = \{\pm u^n : n \in \mathbb{Z}\}$. We may assume $u > 1$. The proof that $(1+\sqrt{5})/2$ is irreducible extends directly to show that u is irreducible. Consider $\lambda = u\sqrt{d}$. Clearly $\lambda \in \mathbb{P}$. Factor λ into a product $\lambda_1 \cdots \lambda_m$ of irreducibles. Since d is square-free, it is a product $p_1 \cdots p_n$ of distinct primes, which are also irreducible. Thus

$$\lambda^2 = \lambda_1^2 \cdots \lambda_m^2 = uup_1 \cdots p_n.$$

But every irreducible on the left occurs to an even power, while each p_i occurs once. Hence λ^2 has distinct irreducible factorizations.

To consider the next case we use information about U supplied by the Dirichlet unit theorem. Recall that K has s real embeddings $\theta_1, \ldots, \theta_s$ and t conjugate pairs of complex embeddings $\theta_{s+1}, \overline{\theta_{s+1}}, \ldots, \theta_{s+t}, \overline{\theta_{s+t}}$, where $[K:\mathbb{Q}] = s+2t$. Assume without loss that θ_1 is the identity. The unit group of K is best studied using the logarithmic embedding

$$g(\alpha) = (\log|\theta_1(\alpha)|, \ldots, \log|\theta_s(\alpha)|, \log|\theta_{s+1}(\alpha)|^2, \ldots, \log|\theta_{s+t}(\alpha)|^2) \in \mathbb{R}^{s+t}.$$

If $V = \{x \in \mathbb{R}^{s+t} : \Sigma_i x_i = 0\}$, then for $\alpha \in U$ we have $g(\alpha) \in V$. The content of Dirichlet's theorem is that $g(U)$ is a cocompact lattice in V, so $g(U)$ is a free abelian group on $s+t-1$ generators. Since here $g(U) \cong U/\{\pm 1\}$, it follows that rank $U = s+t-1$.

Since K is real, $s \geq 1$. Then if $[K:\mathbb{Q}] \geq 4$ or $s = 3$ and $t = 0$, it is easily deduced that rank $U \geq 2$. Thus the only case left is $s = t = 1$, where $[K:\mathbb{Q}] = 3$. We deal with this case shortly.

The condition $\theta_1(\lambda) > |\theta_j(\lambda)|$, $j \geq 2$ defines an open cone in V. Since $g(U)$ is a cocompact lattice in V, it follows that $U \cap \mathbb{P}$ is non-empty and discrete. Hence $U \cap \mathbb{P}$ has a smallest element α, which is clearly irreducible. Let $L = \mathbb{Q}(\alpha)$. If $[L:\mathbb{Q}] = 2$ we are done by the first case. We handle the possibility $[L:\mathbb{Q}] = 3$ and L has non-real embeddings below. Thus replace K by L and assume rank $U \geq 2$. Hence there is a $\beta \in U \cap \mathbb{P}$ such that $\beta \neq \alpha^n$ for $n \geq 1$. Then $\lambda = \alpha^n/\beta \in U$, and clearly $\lambda \in \mathbb{P}$ if n is sufficiently large. Consider $\alpha^n = \lambda\beta$. Since β is a unit that is not a power of α, the irreducible factorization of β must contain units different from α. Since α is irreducible, this yields distinct irreducible factorizations of α^n, concluding this case.

Finally, consider the case $[K:\mathbb{Q}] = 3$ with $s = t = 1$. Here K has one real embedding, the identity θ_1, and one conjugate pair of complex embeddings $\theta_2, \overline{\theta_2}$. Thus rank $U = 1$, so there is a fundamental unit u, and we may assume $u > 1$. It follows that $u \in \mathbb{P}$, and that u is irreducible from a previous argument.

Let $\mathbb{Z}U$ denote the set of rational integers times units. We next find an irreducible $\alpha \in K \cap \mathbb{P}$ that is not in $\mathbb{Z}U$. Define $\theta: K \to \mathbb{R} \times \mathbb{C}$ by $\theta(\alpha) = (\theta_1(\alpha), \theta_2(\alpha))$. Put $\theta(u) = (u, v)$ and define $T(x, z) = (ux, vz)$. For $a > 0$ put

$$W_a = \left\{(x, z) \in \mathbb{R} \times \mathbb{C}: \left|\frac{xv}{u}\right| < |z| \leq |x|, |xz^2| \leq a\right\}.$$

Then W_a is a fundamental wedge for T in that $\{T^n W_a: n \in \mathbb{Z}\}$ is a partition of $\{(x, z): |xz^2| \leq a\}$. Thus if $nu^k \in \theta(\mathbb{Z}U) \cap W_a$, then $k = 0$ and $n \leq a^{1/3}$. Hence

$$|\theta(\mathbb{Z}U) \cap W_a| = O(a^{1/3}).$$

On the other hand, if $\mathcal{O}$ denotes the ring of integers in K, then $\theta(\mathcal{O})$ is a cocompact lattice in $\mathbb{R} \times \mathbb{C}$. Now $\text{vol}(W_{ta}) = t^3 \text{vol}(W_a)$, so $|\theta(\mathcal{O}) \cap W_a| / a$ converges to a positive limit as $a \to \infty$. This disparity shows there are quantitatively many more integers in K than in $\mathbb{Z}U$.

Take $\gamma \in \mathcal{O}$, $\gamma \notin \mathbb{Z}U$, and assume $\gamma > 0$. Note that $|\theta_2(u)| < u$, so $\alpha = \gamma u^k \in \mathbb{P}$ for large enough k. Decompose $\alpha = \alpha_1 \cdots \alpha_n$ into irreducibles. Since $\alpha \notin \mathbb{Z}U$, some $\alpha_j \notin \mathbb{Z}U$. Replace α by α_j, so now α is an irreducible not in $\mathbb{Z}U$. Since $\alpha \notin \mathbb{Q}$ and $[K:\mathbb{Q}] = 3$, we have $K = \mathbb{Q}(\alpha)$, so $|\theta_2(\alpha)| < \alpha$. Thus $\lambda = \alpha^n / u \in \mathbb{P}$ for sufficiently large n. Then $\alpha^n = u\lambda$, and since both α and u are irreducible, $K \cap \mathbb{P}$ has non-unique Perron factorizations. □

In the case rank $U \geq 2$, a slightly more elaborate proof, using the geometry of the lattice $g(U)$ in V, shows there is a pairwise distinct set

$$\{\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n\}$$

of irreducibles in K with $\lambda_1 \cdots \lambda_m = \mu_1 \cdots \mu_n$.

6. *Topological factorizations of Markov shifts*

In this section are some applications of Perron arithmetic to Markov shifts. We

show that the inability to factor a Markov shift into a direct product of homeomorphisms can sometimes be deduced from the arithmetic nature of its entropy. One consequence is that if p is a prime number, then the full p-shift has no such factorizations. Since this fact does not seem to be in the literature, we give two other proofs. Another consequence is that the full n-shift has a kth root exactly when n is a perfect kth power. We also use the Adler–Marcus theorem to deduce that, up to almost topological conjugacy, there is a bijection between direct product factorizations of σ_A and Perron factorizations of λ_A.

We begin by showing that direct factors of Markov shifts are again Markov shifts. Recall our convention that Markov shifts are assumed to be aperiodic. For brevity, say that a homeomorphism is Markov if it is topologically conjugate to a Markov shift.

PROPOSITION 6. *If φ and ψ are homeomorphisms such that $\varphi \times \psi$ is Markov, then both φ and ψ are Markov.*

Proof. Since $\varphi \times \psi$ is Markov, it follows that both φ and ψ are expansive homeomorphisms of zero-dimensional compact spaces and have canonical coordinates. Clearly both are topologically mixing. A result of Bowen [3] shows that φ and ψ are Markov. □

Call a homeomorphism *topologically prime* if it is not topologically conjugate to the direct product of two non-trivial homeomorphisms. The following shows that an arithmetic condition on entropy alone is enough to imply topological primality.

THEOREM 6. *If A is an aperiodic matrix with λ_A irreducible, then σ_A is topologically prime.*

Proof. Suppose $\sigma_A \cong \varphi \times \psi$. By proposition 6, there are aperiodic matrices B, C with $\varphi \cong \sigma_B$ and $\psi \cong \sigma_C$. Hence $\lambda_B, \lambda_C \in \mathbb{P}$ and $\lambda_A = \lambda_B \lambda_C$, contradicting irreducibility of λ_A. □

THEOREM 7. *The full n-shift is topologically prime if and only if n is prime.*

First proof. If n is prime, it is irreducible, so the n-shift is topologically prime by theorem 6. If $n = ab$, then $\sigma_n \cong \sigma_a \times \sigma_b$. □

This can be obtained more directly. The second proof, recently communicated by G. Hansel, uses only a periodic point count and does not rely on proposition 6.

Second proof (Hansel). Suppose p is prime, and $\sigma_p \cong \varphi \times \psi$, where $\varphi: X \to X$ and $\psi: Y \to Y$. Let $N_n(\varphi)$ be the number of fixed points for φ^n, and $L_n(\varphi)$ be the number of points of least period n under φ. Then $N_n(\varphi) = \sum_{d|n} L_d(\varphi)$. Clearly

$$N_n(\varphi \times \psi) = N_n(\varphi) N_n(\psi).$$

Now $p = N_1(\sigma_p) = N_1(\varphi) N_1(\psi)$, so assume without loss that $N_1(\psi) = 1$. Let y_0 be the unique fixed point of ψ. We show $Y = \{y_0\}$, showing ψ is trivial. Note that

$$p^{p^n} = N_{p^n}(\sigma_p) = N_{p^n}(\varphi) N_{p^n}(\psi).$$

Since $p|L_{p^k}(\psi)$ for $k \geq 1$, there is an integer r such that

$$N_{p^n}(\psi) = 1 + \sum_{k=1}^{n} L_{p^k}(\psi) = 1 + pr.$$

Then $(1+pr)|p^{p^n}$, so $r=0$. Therefore $N_{p^n}(\psi)=1$ for $n \geq 0$. Thus for all $n \geq 0$ the points of period p^n under σ_p have Y-coordinate y_0. But these are dense, so $Y=\{y_0\}$. □

This approach is elementary and elegant, but it appears difficult to extend it to a proof of theorem 6. The following proof is more analytic.

Third proof. We anticipate material on zeta functions from the next section. Refer there for unexplained notation and terminology.

Let p be prime. Suppose as in the first proof that $\sigma_p \cong \sigma_B \times \sigma_C$. Let

$$\zeta_B(t) = \prod_i (1-\alpha_i t)^{-1}, \qquad \zeta_C(t) = \prod_j (1-\beta_j t)^{-1}.$$

Since $(1-pt)^{-1} = \zeta_p(t) = \zeta_B(t) \otimes \zeta_C(t)$, taking inverses gives

$$1 - pt = \prod_{i,j} (1 - \alpha_i \beta_j t).$$

Unique factorization in $\mathbb{C}[t]$ shows that both $\zeta_B(t)^{-1}$ and $\zeta_C(t)^{-1}$ are linear, say $1-n_B t$ and $1-n_C t$ with $n_B, n_C \in \mathbb{Z}$. Then $p = n_B n_C$, a contradiction unless one of the factors is trivial. □

Another result of number theoretic considerations is a necessary condition for the existence of roots of Markov shifts.

THEOREM 8. *A necessary condition for a Markov shift σ_A to have a k'th root is $\lambda_A^{1/k} \in \mathbb{P}$. In particular, σ_n has a k'th root if and only if n is a perfect k'th power.*

Proof. If φ is a homeomorphism such that $\varphi^k = \sigma_A$, an argument similar to that in proposition 6 shows that φ is a mixing Markov shift as well, say $\varphi \cong \sigma_B$. Then $\lambda_A^{1/k} = \lambda_B \in \mathbb{P}$. The second statement then follows since $n^{1/k} \in \mathbb{P}$ exactly when $n^{1/k}$ is an integer. □

In [1] Adler and Marcus introduced a natural relationship between Markov shifts which they called almost topological conjugacy. They showed that topological entropy classified aperiodic Markov shifts up to almost topological conjugacy. Combining this with proposition 6 and theorem 1 yields the following.

THEOREM 9. *Let σ be a Markov shift with entropy* $\log \lambda$. *Then up to almost topological conjugacy there is a bijection between factorizations of σ into a direct product of non-trivial homeomorphisms and Perron factorizations of λ. In particular, the number of such factorizations is finite.*

7. *Non-unique factorizations*

It is a consequence of the finiteness of the number of Perron factorizations together with proposition 6 that a Markov shift is the direct product of a finite number of topologically prime Markov shifts. Is such a factorization unique up to order? Direct

factorization of a Markov shift induces a kind of factorization of its zeta function, providing a necessary condition for the former. The non-uniqueness of Perron factorizations suggests corresponding non-uniqueness for factorizations of zeta functions and of Markov shifts. We illustrate this here with an example, leading to a case of a non-unique topological factorization.

The periodic point information about a Markov shift σ_A is summarized in its zeta function

$$\zeta_A(t) = \exp\left(\sum_{n=1}^{\infty} \frac{N_n(\sigma_A)t^n}{n}\right).$$

The zeta function has a product formula [**9**]

$$\zeta_A(t) = \prod_P (1 - t^{|P|})^{-1},$$

where the product is over all periodic orbits P of σ_A. Bowen and Lanford [**5**] showed that if $\chi_A(t) = \prod_i (t - \lambda_i)$ then $\zeta_A(t) = \prod_i (1 - \lambda_i t)^{-1}$.

The zeta function of a product is computed as follows. There is a natural conjugacy between $\sigma_B \times \sigma_C$ and $\sigma_{B \otimes C}$. If $\chi_B(t) = \prod_i (t - \lambda_i)$ and $\chi_C(t) = \prod_j (t - \mu_j)$, then

$$\chi_{B \otimes C}(t) = \prod_{i,j} (t - \lambda_i \mu_j).$$

Thus

$$\zeta_{B \otimes C}(t) = \prod_{i,j} (1 - \lambda_i \mu_j t)^{-1}.$$

If we define the $\otimes$-product of zeta functions of Markov shifts by

$$\prod_i (1 - \lambda_i t)^{-1} \otimes \prod_j (1 - \mu_j t)^{-1} = \prod_{i,j} (1 - \lambda_i \mu_j t)^{-1},$$

then $\zeta_{B \otimes C} = \zeta_B \otimes \zeta_C$. Thus a factorization $\sigma_A = \sigma_B \times \sigma_C$ implies a $\otimes$-factorization $\zeta_A = \zeta_B \otimes \zeta_C$ of the corresponding zeta functions, which in turn implies a Perron factorization $\lambda_A = \lambda_B \lambda_C$.

Do distinct Perron factorizations always arise from distinct $\otimes$-factorizations of zeta functions, or from direct product factorizations of Markov shifts? We do not know the general answer, but give here one example of this phenomenon.

Let $\alpha = (1 + \sqrt{5})/2$, and $\beta = (1 - \sqrt{5})/2$ be its conjugate. Then as above α, $\alpha + 2$, and 5 are irreducible, so $(\alpha + 2)^2 = 5\alpha^2$ are distinct irreducible factorizations. We will find matrices B for $\alpha + 2$ and A for α^2 which will yield corresponding distinct zeta function factorizations. Next using a theorem of Krieger, we show that $B \otimes B$ is shift equivalent to $5A$. Taking fifth powers will give conjugate Markov shifts having distinct factorizations into topologically prime shifts. Let

$$B = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

Then $\chi_B(t) = [t - (\alpha + 2)][y - (\beta + 2)]$ so

$$\zeta_{B \otimes B}(t) = [1 - (\alpha + 2)^2 t]^{-1}[1 - (\beta + 2)^2 t]^{-1}[1 - 5t]^{-2}$$

and $\lambda_{B \otimes B} = (\alpha + 2)^2$. Next consider

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

with eigenvalues α^2 and β^2. In order to match zeta functions, we need to modify C to an aperiodic matrix A with an additional eigenvalue 1 of multiplicity two. An elegant technique from [**6**] can be used to construct a Markov cover σ_A for σ_C such that $N_n(\sigma_A) = N_n(\sigma_C)$ for $n \geq 2$, while $N_1(\sigma_A) = N_1(\sigma_C) + 2$. Specifically,

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

The product formula implies that $\zeta_A(t) = \zeta_C(t)(1-t)^{-2}$. Hence $\zeta_B \otimes \zeta_B = \zeta_5 \otimes \zeta_A$ gives distinct zeta function factorizations.

From this we conclude that $\sigma_B \times \sigma_B$ and $\sigma_5 \times \sigma_A$ have identical zeta functions. In fact, they are shift equivalent with lag 4. We sketch the use of Krieger's theorem to find a shift equivalence, and then give the explicit matrices.

The idea is to identify the dimension groups of $B \otimes B$ and $5A$, and show there is an order-preserving isomorphism between them. We use the notation and terminology of [**8**]. Let S denote the ring $\mathbb{Z}[\frac{1}{5}]$. The dimension group $G(B \otimes B)$ embeds naturally in $\mathbb{R}^4$ as S^4. Let

$$M = \tfrac{1}{5}\begin{bmatrix} 4 & 3 & -2 & 1 \\ 5 & 0 & 0 & 0 \\ 2 & 4 & -1 & 3 \\ 3 & 1 & 1 & 2 \\ 0 & 0 & 5 & 0 \end{bmatrix}.$$

The columns of M generate $G(5A)$ over S, and M gives an order isomorphism between $G(B \otimes B)$ and $G(5A)$. Using Krieger's theorem [**12**], this produces matrices

$$V = \begin{bmatrix} 8 & 6 & 1 & 2 \\ 9 & 3 & 3 & 1 \\ 6 & 7 & 2 & 4 \\ 7 & 4 & 4 & 3 \\ 3 & 1 & 6 & 2 \end{bmatrix},$$

$$W = \begin{bmatrix} 875 & 250 & 125 & 375 & 1000 \\ 625 & 0 & 125 & 250 & 625 \\ 250 & 250 & 250 & 250 & 625 \\ 125 & 125 & 250 & 125 & 375 \end{bmatrix},$$

such that $WV = (B \otimes B)^4$, $VW = (5A)^4$, $(B \otimes B)W = W(5A)$, and $V(B \otimes B) = (5A)V$. Thus V and W give a shift equivalence of lag 4 between $B \otimes B$ and $5A$.

It follows that for $n \geq 4$, $\sigma_5^n \times \sigma_A^n \cong \sigma_B^n \times \sigma_B^n$. Let $n = 5$. Since $\lambda_{B^5} = 5^2\alpha^4(\alpha+2)$ and $\alpha^4(\alpha+2)$ is not divisible by 5 in $\mathbb{P}$, it follows that any prime factorization of σ_B^5 contains at most two terms conjugate to σ_5. Fix a prime factorization of σ_B^5 and use it twice to factor $\sigma_B^5 \times \sigma_B^5$. Now $\sigma_5^5 \cong \sigma_5 \times \cdots \times \sigma_5$ (5 times) has more copies of σ_5 than occur in the above factorization of $\sigma_B^5 \times \sigma_B^5$. Take any prime factorization of σ_A^5 to complete the non-uniqueness example.

It is unknown to us whether $\sigma_5 \times \sigma_A$ and $\sigma_B \times \sigma_B$ are themselves conjugate. A positive answer would be an interesting example of non-unique topological factorization, while of course a negative answer would settle the shift equivalence problem of Williams.

REFERENCES

[1] R. L. Adler & B. Marcus. Topological entropy and equivalence of dynamical systems, *Mem. Amer. Math. Soc.* **219**, Providence, 1979.

[2] R. L. Adler & B. Weiss. Similarity of automorphisms of the torus, *Mem. Amer. Math. Soc.* **98**, Providence, 1970.

[3] R. Bowen, Topological entropy and Axiom A. In *Proc. Symp. Pure Math.*, **14**, 23–41, Amer. Math. Soc: Providence, RI., 1970.

[4] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms.* Springer Lecture Notes in Math. **470**, Springer-Verlag: New York, 1975.

[5] R. Bowen & O. E. Lanford III. Zeta functions of restrictions of the shift transformation. In *Proc. Symp. Pure Math.* **14**, 43–50, Amer. Math. Soc: Providence, R.I., 1970.

[6] M. Boyle. Lower entropy factors of sofic systems. *Ergod. Th. & Dynam. Sys.* **3** (1983), 541–557.

[7] M. Boyle & S. Tuncel. Infinite-to-one codes and Markov Measures. Preprint.

[8] E. G. Effros. *Dimensions and C*-algebras.* CBMS Conference Ser. no. 46, Amer. Math. Soc.: Providence, R.I., 1981.

[9] J. M. Franks. *Homology and Dynamical Systems*, CBMS Conference Ser. no. 49, Amer. Math. Soc.: Providence, R.I. 1982.

[10] F. R. Gantmacher. *The Theory of Matrices*, vol. 2. Chelsea: New York, 1959.

[11] M. W. Hirsch & S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press: New York, 1974.

[12] W. Krieger. On dimension functions and topological Markov chains. *Invent. Math.* **56** (1980), 239–250.

[13] S. Lang. *Algebra.* Addison-Wesley: Reading, 1965.

[14] D. A. Lind. Entropies and factorizations of topological Markov shifts. *Bull. Amer. Math. Soc.* To appear.

[15] M. Morse. Representation of geodesics, *Amer. J. Math.* 35–51.

[16] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–56.

[17] C. E. Shannon & W. Weaver. *The Mathematical Theory of Communication.* University of Ill.: Urbana, 1963.

[18] Ya. Sinai. Markov partitions and C-diffeomorphisms. *Funct. Anal. and its Appl.*, **2** (1968), No. 1, 64–89.

[19] R. F. Williams. Classification of subshifts of finite type. *Ann. of Math.* **98** (1973), 120–153; Errata, **99** (1974), 380–381.