PERTURBATIONS OF SHIFTS OF FINITE TYPE*

D. A. LIND†

Abstract. Shifts of finite type describe the infinite trips on a labeled graph, and provide theoretical models for data storage and transmission. The consequences of forbidding a fixed word to occur, which can be considered as a small change or perturbation in the system, are investigated. This situation arises in prefix synchronized codes, where a certain prefix, used to synchronize code words, is forbidden to occur in the rest of the word. If T is the adjacency matrix of the graph, and λ_T is its spectral radius, then forbidding a word of length k results in a drop in spectral radius that lies between two positive constants times λ_T^{-k}. The zeta function summarizes the number of possible periodic trips. The author gives an explicit calculation of the zeta function for the resulting subshift, which involves the characteristic polynomial of T, a cofactor of $tI - T$, and the correlation polynomial of the word. A modification of the Knuth–Morris–Pratt pattern matching algorithm shows that this calculation can be done in time that is linear in the word length, answering a question of Bowen and Lanford. The structure of this correlation polynomial is used to obtain sharp bounds on the degree of the denominator of the zeta function. The Jordan form of the higher order presentations of the shift of finite type is also computed, and that of the perturbation in many cases. Most of the results were discovered experimentally with a computer.

Key words. shift of finite type, symbolic dynamics, channel capacity, topological entropy, zeta function, perturbation, correlation polynomial

AMS(MOS) subject classifications. primary: 54H20, 58F20, 58F30; secondary: 05A15, 28D20, 34C35

1. Introduction. A *dynamical system* is a pair (X, σ), where X is a topological space and σ is a continuous map from X to itself. Shifts of finite type are combinatorially defined dynamical systems that arise naturally in the study of data storage and transmission [1]. They also play a central role in the analysis of other dynamical systems such as diffeomorphisms of manifolds [3], an idea which goes back to Hadamard's analysis in 1898 of geodesic flows. Roughly speaking, shifts of finite type are the topological analogues of the probabilistic Markov chains used in information theory.

There are two equivalent definitions of shifts of finite type. The first definition begins with a nonnegative integral matrix T, and forms the directed graph G_T having T_{ij} distinct edges from vertex i to vertex j. Let $\mathscr{E}$ denote the set of edges of G_T. The space $X_T \subset \mathscr{E}^{\mathbb{Z}}$ consists of all bi-infinite sequences from $\mathscr{E}$ that give an allowed trip on G_T. The map $\sigma_T: X_T \rightarrow X_T$ shifts a sequence one edge to the left. As an example, let $T = [\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}]$, so that G_T has two vertices, say labeled 0 and 1, and three edges, say labeled ξ_{00}, ξ_{01}, and ξ_{10}. A sequence $\cdots \xi_{i_{-1}j_{-1}} \xi_{i_0j_0} \xi_{i_1j_1} \cdots$ is in X_T exactly when, for every k, the terminal vertex of $\xi_{i_kj_k}$ agrees with the initial vertex of $\xi_{i_{k+1}j_{k+1}}$, i.e., when $j_k = i_{k+1}$.

The second definition starts with a finite alphabet A, and a finite collection $\mathscr{F}$ of "forbidden" finite words over A. The space $X_{\mathscr{F}} \subset A^{\mathbb{Z}}$ consists of all bi-infinite sequences from A that do not contain any word from $\mathscr{F}$. The shift map $\sigma_{\mathscr{F}}: X_{\mathscr{F}} \rightarrow X_{\mathscr{F}}$ acts as before. For example, if $A = \{0, 1\}$ and $\mathscr{F} = \{11\}$, then $X_{\mathscr{F}}$ is the set of all bi-infinite sequences of 0's and 1's that do not contain consecutive 1's.

In each definition there is a distance function on bi-infinite sequences defined by $d(\{x_k\}, \{y_k\}) = 2^{-n}$, where n is the largest integer such that $x_k = y_k$ for $-n \leq k \leq n$. Thus two points are close if their coordinates agree on a large symmetric interval of indices. The shift map is continuous with respect to the topology induced by the metric d.

It is usual to consider two dynamical systems to be the "same" if one can be obtained from the other by a continuous relabeling of points. More precisely, we say that (X_1, τ_1) is *topologically conjugate* to (X_2, τ_2) if there is a continuous map $\varphi\colon X_1 \to X_2$ with a continuous inverse such that $\tau_2 \varphi = \varphi \tau_1$. The two examples of shifts of finite type given above are topologically conjugate using the map $\varphi\colon X_T \to X_{\mathscr{F}}$ given by $\varphi(\{\xi_{i_k j_k}\}) = \{i_k\}$, so that φ assigns to a sequence of edges the corresponding sequence of their initial vertices.

A *topological invariant* of dynamical systems is an object (real number, group, etc.) assigned to every dynamical system that is the same for topologically conjugate systems. Examples of such invariants are entropy and the zeta function, which are defined below.

Our two definitions of shifts of finite type are equivalent in the sense that any dynamical system constructed using one definition is topologically conjugate to a dynamical system constructed by the other. If (X_T, σ_T) is a system from the first definition, let $\mathscr{E}$ be the alphabet, and let $\mathscr{F}$ be the set of pairs of edges satisfying the condition that the terminal vertex of the first edge is not the initial vertex of the second. Then $(X_{\mathscr{F}}, \sigma_{\mathscr{F}})$ is topologically conjugate to (X_T, σ_T). Conversely, if $(X_{\mathscr{F}}, \sigma_{\mathscr{F}})$ is constructed using the second definition, it is possible to determine, as in § 2, a directed graph with adjacency matrix T so that (X_T, σ_T) is conjugate to $(X_{\mathscr{F}}, \sigma_{\mathscr{F}})$.

We shall use the first definition of shift of finite type throughout. When a shift arises by forbidding words, we will convert this to a conjugate shift defined by a nonnegative integer matrix using the process described in § 2. We shall also assume throughout that for each pair of nodes there is a path from the first to the second or, equivalently, that σ_T is topologically transitive. To avoid trivial exceptions, we also require $T \neq [1]$.

Prefix synchronized codes [7], [8] can be described as follows. Select a fixed block B, and choose code words of the form BA subject to the constraint that B occurs in BAB only as the first and last subblock. Thus an occurrence of B guarantees that a decoder is at the beginning of a code word. In estimating the number of code words available, we are quickly led to the study of the shift of finite type obtained by forbidding B to occur. If B is long, this shift can be considered as a small change in the original shift.

Let $B \in \mathscr{E}^k$ be an allowed path in the graph G_T of length $k = |B|$. The shift of finite type obtained from (X_T, σ_T) by forbidding B to occur can be presented by a matrix we denote by $T\langle B \rangle$ that is typically much larger than T. The details of this construction are given in § 2. For the purposes of this paper we regard passing from σ_T to $\sigma_{T\langle B \rangle}$ as a small perturbation of σ_T, with the perturbation tending to zero as $|B| \to \infty$. Our main theorems give precise information about the resulting changes in topological entropy, zeta function, and Jordan form.

To describe the change in entropy, let T have spectral radius λ_T. The topological entropy $h(\sigma_T)$ is then $\log \lambda_T$. Standard Perron–Frobenius theory [16, Thm. 1.1(e)] shows $h(\sigma_T) - h(\sigma_{T\langle B \rangle}) > 0$. During the investigation [5] of the automorphism group of σ_T, it was necessary to prove this difference tends to zero as $|B| \to \infty$ [5, Thm. 5.1]. The analysis of the speed of convergence was one motivation for this paper. We show in Theorem 3 that there are computable constants c_T, $d_T > 0$ so that for all blocks B with sufficiently large length $|B|$ we have

$$c_T \lambda_T^{-|B|} < h(\sigma_T) - h(\sigma_{T\langle B \rangle}) < d_T \lambda_T^{-|B|}.$$

The zeta function of a map was introduced by Artin and Mazur [2] to conveniently summarize periodic point data. If N_n denotes the number of points in X_T fixed by σ_T^n, then the zeta function of σ_T is defined as

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

Bowen and Lanford [4] showed $\zeta_T(t) = 1/\det(I - tT)$, which we reformulate differently as follows. If $f(t) \in \mathbb{Z}[t, t^{-1}]$ is a Laurent polynomial, choose the unique $d \in \mathbb{Z}$ so that $t^d f(t) \in \mathbb{Z}[t]$ with nonzero constant term. Let $f(t)^{\times}$ denote this $t^d f(t)$. If $\chi_T(t) = \det(tI - T)$ is the characteristic polynomial of T, then clearly

$$\zeta_T(t) = \frac{1}{\chi_T(t^{-1})^{\times}}.$$

To describe the zeta function of the perturbed matrix, we first recall the notion of correlation polynomial introduced by J. H. Conway and exploited in a series of papers by Guibas and Odlyzko [8]–[10]. If $B \in \mathscr{E}^k$, the *correlation polynomial* of B is $\phi_B(t) = \sum_{j=1}^{k} c_j t^{j-1}$, where $c_j = 1$ if B overlaps itself in j symbols, and $c_j = 0$ otherwise. A recursive description of the set of possible correlation polynomials is given in [9], having the surprising consequence that this set is independent of $\mathscr{E}$ for $|\mathscr{E}| \geq 2$. These polynomials arise in a wide range of applications, including the computation of the number of strings of a given length omitting a fixed block [8], the analysis of pattern matching algorithms [9], the study of nerve impulses in crayfish [6], and some astonishing probability calculations concerning "paradoxical" nontransitive games [10].

Suppose B is a path from node i to node j. Call B *reduced* if no proper subblock of B determines B. Let $\text{cof}_{ij}(tI - T)$ denote $(-1)^{i+j}$ times the determinant of the matrix $tI - T$ with its ith row and jth column removed. We show in Theorem 1 that

$$(1.1) \qquad \zeta_{T\langle B \rangle}(t) = \frac{1}{[\chi_T(t^{-1})\phi_B(t^{-1}) + \text{cof}_{ij}(t^{-1}I - T)]^{\times}}.$$

When $T = [n]$, so σ_T is the full n-shift, every block is reduced, and our result simplifies to

$$\zeta_{T\langle B \rangle}(t) = \frac{1}{(1 - nt)\phi_B(t^{-1})t^{|B|-1} + t^{|B|}}.$$

Since $T\langle B \rangle$ has size that is exponential in $|B|$, it is not clear whether there is an algorithm to compute $\zeta_{T\langle B \rangle}(t)$ in time polynomial in $|B|$. But since there is a trivial algorithm to compute $\phi_B(t)$ in time that is quadratic in $|B|$, one consequence of Theorem 1 is the existence of a quadratic-time algorithm to compute $\zeta_{T\langle B \rangle}(t)$. This is sharpened in Corollary 2 to a linear algorithm based on the Knuth–Morris–Pratt pattern matching algorithm [13], answering a question raised by Bowen and Lanford [4, p. 45]. Also, although there are roughly λ_T^k blocks of length k, Guibas and Odlyzko show that there are only $O(k^{c \log k})$ distinct correlation polynomials for these blocks. Thus another consequence of our calculation, stated in Corollary 3, is that the number of distinct zeta functions produced by omitting blocks of length k is only $O(k^{c \log k})$. In Corollary 4 we show that the number of distinct zeta functions produced by deleting blocks of length k is the same for all full shifts.

Because of possible divisibility by powers of t in $\chi_T(t)$, $\phi_B(t)$, and $\text{cof}_{ij}(tI - T)$, together with possible cancellations, the formula (1.1) does not immediately yield the degree of $\zeta_{T\langle B \rangle}^{-1}(t)$. Using machinery in §§ 2–3, combined with the recursive characterization of $\phi_B(t)$ from [9], we show in Theorem 2 that

$$|B| - 4r \leq \deg \zeta_{T\langle B \rangle}^{-1} \leq |B| + r - 1,$$

where r is the size of T, and that the upper bound is attained precisely when

$$(1.2) \qquad\qquad (\det T)\phi_B(0) + \text{cof}_{ij}(-T) \neq 0.$$

In particular, the upper bound obtains for all blocks in a full shift.

The Jordan form of T is not a topological invariant, but our analysis allows the determination in Theorem 4 of the Jordan form of all higher block presentations of T. Also, if (1.2) holds, then the Jordan form of $T\langle B \rangle$ is obtained from that of the k-block presentation of T by deleting a single Jordan nilpotent block of size $k - 1$. Again, every block in a full shift satisfies this condition.

The paper is organized as follows. We begin in § 2 by setting up the machinery for higher block presentations of T. In § 3 we determine a low-dimensional subspace containing the "interesting part" of $T\langle B \rangle$ and compute the matrix of $T\langle B \rangle$ with respect to a certain natural basis. This matrix contains the companion matrix of $\phi_B(t)$ as a principal submatrix and is used in § 4 to compute $\zeta_{T\langle B \rangle}(t)$ and to estimate deg $\zeta_{T\langle B \rangle}^{-1}(t)$. In § 5 we first motivate the estimate on decrease in entropy using an elementary but apparently little-known fact on the rate of change in an eigenvalue with respect to the matrix entries. We then prove the estimate by using the calculation of $\chi_{T\langle B \rangle}(t)$ together with the recursive structure of $\phi_B(t)$. Finally, we apply this machinery in § 6 to derive the results on Jordan forms mentioned above.

Most of our results were discovered experimentally using the Matlab interactive linear algebra computer program. Jim Scherer (private communication) has indicated to us an alternative approach to an upper bound on deg $\zeta_{T\langle B \rangle}^{-1}$, but his results are not as sharp as Theorem 2. Razmik Karabed (private communication) has described an interesting method for obtaining an exponential upper bound in Theorem 3, but his method does not yield either the best exponent or the lower bound.

2. Higher order presentations. In this section we introduce the linear algebra machinery we will use to analyze $T\langle B \rangle$. Our treatment of higher block presentations varies slightly from the standard one, since we use a consistent treatment for all nonnegative integer matrices rather than just 0-1 matrices.

Let $T = [T_{ij}]$ be an $r \times r$ matrix over the nonnegative integers. Let $\mathscr{S} = \{1, \cdots, r\}$ be the indexing set of T, and call the elements of $\mathscr{S}$ the 1-*states*. For $1 \leq p \leq T_{ij}$, introduce 1-*symbols* ξ_{ij}^p, and let $\mathscr{E} = \{\xi_{ij}^p : 1 \leq p \leq T_{ij}, 1 \leq i, j \leq r\}$. Thus $\mathscr{E}$ is a labeling of the edges of the directed graph determined by T. Define beginning and ending maps $\beta, \eta : \mathscr{E} \to \mathscr{S}$ by $\beta(\xi_{ij}^p) = i$ and $\eta(\xi_{ij}^p) = j$. Thus

$$T_{ij} = |\{\xi \in \mathscr{E} : \beta(\xi) = i \text{ and } \eta(\xi) = j\}|.$$

Put

$$X_T = \{x = (x_n) \in \mathscr{E}^{\mathbb{Z}} : \eta(x_n) = \beta(x_{n+1}) \text{ for } n \in \mathbb{Z}\},$$

and define the shift $\sigma_T : X_T \to X_T$ by $(\sigma_T x)_n = x_{n+1}$. Then σ_T is a homeomorphism of the compact totally disconnected metric space X_T. The pair (X_T, σ_T) is the *shift of finite type* determined by T.

Next we recode points in X_T by k-blocks of symbols. For $k \geq 1$ let

$$\mathscr{B}_k = \mathscr{B}_k(X_T) = \{\xi_1 \cdots \xi_k \in \mathscr{E}^k : \eta(\xi_m) = \beta(\xi_{m+1}) \text{ for } 1 \leq m \leq k-1\}$$

be the set of allowed k-blocks in X_T. By convention we put $\mathscr{B}_0 = \mathscr{S}$. For $k \geq 1$, let $\mathscr{S}^{[k]} = \mathscr{B}_{k-1}$ denote the k-*states* and $\mathscr{E}^{[k]} = \mathscr{B}_k$ be the k-*symbols*. Note that $\mathscr{S}^{[1]} = \mathscr{S}$ and $\mathscr{E}^{[1]} = \mathscr{E}$. For $k \geq 2$, define $\beta, \eta : \mathscr{E}^{[k]} \to \mathscr{S}^{[k]}$ by $\beta(\xi_1 \cdots \xi_k) = \xi_1 \cdots \xi_{k-1}$ and $\eta(\xi_1 \cdots \xi_k) = \xi_2 \cdots \xi_k$. Since $\mathscr{S}^{[k]} = \mathscr{E}^{[k-1]} = \mathscr{B}_{k-1}$, we may compose various β's and η's in any order, and they clearly commute.

We will now define the kth *order presentation matrix* $T^{[k]}$ of T, which is indexed by k-states $\mathscr{S}^{[k]}$. For $A, B \in \mathscr{S}^{[k]}$ define

$$(T^{[k]})_{AB} = |\{C \in \mathscr{E}^{[k]} : \beta(C) = A \text{ and } \eta(C) = B\}|.$$

Note that $T^{[1]}$ is just T. If $k \geqq 2$, then $T^{[k]}$ is a 0-1 matrix. Let $\mathbb{R}^{[k]}$ denote $\mathbb{R}^{|\mathscr{S}^{[k]}|}$, with its basis the elementary vectors $\{e_A^{[k]} : A \in \mathscr{S}^{[k]}\}$. Then $T^{[k]}$ is a linear map of $\mathbb{R}^{[k]}$. It is notationally convenient to have all matrices act on the *right*. Thus, for $k \geqq 2$,

$$e_A^{[k]} T^{[k]} = \sum_{\{B \in \mathscr{S}^{[k]} : \beta B = \eta A\}} e_B^{[k]}.$$

PROPOSITION. *The systems* (X_T, σ_T) *and* $(X_{T^{[k]}}, \sigma_{T^{[k]}})$ *are topologically conjugate.*

Proof. This holds for $k = 1$ by definition. Suppose $k \geqq 2$. Denote $\alpha: X_{T^{[k]}} \to \mathscr{E}^{\mathbb{Z}}$ by $\alpha((C_n)_{n \in \mathbb{Z}}) = (\beta^{k-1} C_n)_{n \in \mathbb{Z}}$. Clearly, α is continuous. Since $\eta C_n = \beta C_{n+1}$, we have

$$\eta(\beta^{k-1} C_n) = \beta^{k-1} \eta C_n = \beta(\beta^{k-1} C_{n+1}),$$

so the image of α is contained in X_T. Clearly, $\sigma_T \alpha = \alpha \sigma_{T^{[k]}}$. Define $\gamma: X_T \to X_{T^{[k]}}$ by $(\gamma x)_m = x_m x_{m+1} \cdots x_{m+k-1} \in \mathscr{B}_k$. Then γ is continuous, and $\alpha \gamma$ and $\gamma \alpha$ are the identity maps. This shows α is a topological conjugacy of $\sigma_{T^{[k]}}$ with σ_T.   □

3. Invariant subspaces for perturbations. Let $B \in \mathscr{B}_k(X_T)$, where we assume from now on that $k \geqq 2$. Forbidding B in X_T is the same as forbidding the transition from βB to ηB in $T^{[k]}$. Thus a matrix presentation of the resulting shift of finite type is $T^{[k]} - E_{\beta E, \eta B}$, where $E = E_{\beta B, \eta B}$ has a 1 in the $(\beta B, \eta B)$th place and is zero elsewhere. Let $T\langle B \rangle = T^{[k]} - E$ denote this perturbed matrix. In this section we will find a $T^{[k]}$-invariant subspace $V_k \subset \mathbb{R}^{[k]}$ on which $T^{[k]}$ mimics T. Then we extend V_k to a $T\langle B \rangle$-invariant subspace $W_{k,B}$ containing the "interesting part" of $T\langle B \rangle$. Finally, we compute the matrix of $T\langle B \rangle$ with respect to a certain basis for $W_{k,B}$ and show how the companion matrix for the correlation polynomial of B appears as a principal submatrix.

Define $\psi: \mathbb{R}^{[1]} \to \mathbb{R}^{[k]}$ by

$$(3.1) \qquad \psi(e_i^{[1]}) = \sum_{\{A \in \mathscr{S}^{[k]} : \beta^{k-1} A = i\}} e_A^{[k]}, \qquad 1 \leqq i \leqq r.$$

Note that the sum is over k-states A whose initial *state* (not initial symbol) is i. Let V_k be the image of ψ in $\mathbb{R}^{[k]}$. The *eventual range* of a linear map on a vector space is the intersection of the images of its nonnegative powers.

LEMMA 1. *The map* ψ *in* (3.1) *is an isomorphism from* $\mathbb{R}^{[1]}$ *to* V_k, *its range* V_k *is* $T^{[k]}$-*invariant, and* $\psi T = T^{[k]} \psi$. *Furthermore,* $T^{[k]}$ *is nilpotent on* $\mathbb{R}^{[k]}/V_k$, *so* V_k *contains the eventual range of* $T^{[k]}$.

Proof. Since the sets $\beta^{-(k-1)}(i) \subset \mathscr{S}^{[k]}$ are disjoint for $i \in \mathscr{S}$, it follows that ψ is injective. Recalling our convention that matrices act on the right, we compute the action of $T^{[k]} \psi$ on a basis vector $e_i^{[1]}$ for $i \in \mathscr{S}$ by

$$T^{[k]} \psi(e_i^{[1]}) = T^{[k]} \left(\sum_{\{A : \beta^{k-1} A = i\}} e_A^{[k]} \right) = \sum_{\{A : \beta^{k-1} A = i\}} e_A^{[k]} T^{[k]}$$

$$= \sum_{\{A : \beta^{k-1} A = i\}} \sum_{\{B : \beta B = \eta A\}} e_B^{[k]} = \sum_{\{C \in \mathscr{E}^{[k]} : \beta^k C = i\}} e_{\eta C}^{[k]}.$$

Similarly,

$$\psi T(e_i^{[1]}) = \psi(e_i^{[1]} T) = \psi \left(\sum_{j \in \mathscr{S}} T_{ij} e_j^{[1]} \right)$$

$$= \sum_{j \in \mathscr{S}} \sum_{\{A : \beta^{k-1} A = j\}} T_{ij} e_A^{[k]} = \sum_{\{C \in \mathscr{E}^{[k]} : \beta^k C = i\}} e_{\eta C}^{[k]}.$$

Thus $\psi T = T^{[k]}\psi$, which also shows V_k is $T^{[k]}$-invariant. To show that $T^{[k]}$ is nilpotent on $\mathbb{R}^{[k]}/V_k$, let $A \in \mathscr{S}^{[k]}$ have terminal state $j = \eta^{k-1}A$. Then

$$(3.2) \qquad e_A^{[k]}(T^{[k]})^{k-1} = \sum_{\{B \in \mathscr{S}^{[k]} : \beta^{k-1}B = \eta^{k-1}A\}} e_B^{[k]} = \psi(e_j^{[1]}) \in V_k. \qquad \square$$

Suppose $B = \xi_1 \cdots \xi_k$, and that $j = \eta(\xi_{k-1})$ has the property that $\sum_{l=1}^r T_{jl} = 1$. Then the subblock $B' = \xi_1 \cdots \xi_{k-1}$ determines B, and forbidding B' is the same as forbidding B. Call a block *reduced* if no proper subblock determines it. We shall prove our results for reduced blocks. By irreducibility of T, every block of length $k > 2r$ contains a reduced subblock of length greater than $k - 2r$ that determines it. Hence most of our results hold for all blocks of length k, perhaps with different constants, by applying them to this reduced subblock.

Let $B \in \mathscr{B}_k(X_T)$ be reduced. We will enlarge V_k by $k - 1$ dimensions to obtain a $T\langle B \rangle$-invariant subspace $W_k = W_{k,B}$ containing the nonnilpotent part of $T\langle B \rangle$ and calculate the matrix of $T\langle B \rangle$ with respect to a natural basis of W_k.

For notational simplicity, let U denote $T^{[k]}$, and $E = E_{\beta B, \eta B}$, so $T\langle B \rangle = U - E$. Also, let $e = e_{\eta B}^{[k]}$, so the range of E is $\mathbb{R}e$. Since B is reduced, it follows that $e \notin V_k$, since otherwise the proper subblock $\beta^{k-1}B$ would determine B.

The next result shows that the absorption time of e into V_k under U is $k - 1$.

LEMMA 2. *With the above notations*, $\min \{m > 0 : eU^m \in V_k\} = k - 1$.

Proof. By (3.2) the minimum is at most $k - 1$.

Equality is proven by observing that the next to last state i in B must be followed by at least two symbols since B is reduced. Hence, the set of blocks in $\mathscr{S}^{[k]}$ whose first symbol is the last symbol of B is a proper subset of the set of those whose initial state is i.

Let $\xi = \eta^{k-1}B$ be the terminal symbol of B, and $i = \beta\eta^{k-1}B$ its initial state. Since B is reduced, $\sum_{j \in \mathscr{S}} T_{ij} \geqq 2$. Hence

$$eU^{k-2} = e_{\eta B}^{[k]}(T^{[k]})^{k-2} = \sum_{\{C : \beta^{k-2}C = \xi\}} e_C^{[k]}$$

is summed over a proper subset of $\beta^{-k+1}(i)$. For fixed i, every vector in V_k must have the same coordinate on each $e_C^{[k]}$ for $C \in \beta^{-k+1}(i)$. Hence $eU^{k-2} \notin V_k$. $\qquad \square$

LEMMA 3. *With the above notations, the vectors* $e, eU, \cdots, eU^{k-2}$ *are linearly independent of each other and of* V_k.

Proof. Suppose $a_1 e + a_2 eU + \cdots + a_{k-1} eU^{k-2} + v = 0$ for some $v \in V_k$. If all $a_j = 0$, we are done. If not, choose j minimal so $a_j \neq 0$. Applying U^{k-j-1} on the right gives

$$a_j eU^{k-2} = v - (a_{j+1} eU^{k-1} + \cdots + a_{k-1} eU^{2k-3-j}) \in V_k,$$

contradicting Lemma 2. $\qquad \square$

In view of Lemma 3, we can introduce the subspace

$$W_k = W_{k,B} = V_k \oplus \mathbb{R}e \oplus \mathbb{R}eU \oplus \cdots \oplus \mathbb{R}eU^{k-2} \subset \mathbb{R}^{[k]},$$

having dimension $r + k - 1$, where $r = \dim V_k = \dim \mathbb{R}^{[1]}$ is the size of T. Although V_k depends only on k, note that W_k also depends on B.

LEMMA 4. *The subspace W_k is invariant under both U and $U - E$, and it contains the eventual range of both transformations.*

Proof. The invariance of W_k under U is clear. Since the range of E is $\mathbb{R}e$, it follows that W_k is also invariant under $U - E$. Since by Lemma 1 we have U nilpotent on

$\mathbb{R}^{[k]}/W_k$, and E vanishes there, $U - E$ is nilpotent on $\mathbb{R}^{[k]}/W_k$, so the eventual range of $U - E$ is contained in W_k. □

If ψ is the map defined by (3.1), by Lemma 1 the set $\{\psi(e_i^{[1]}) : i \in \mathscr{S}\}$ forms a basis for V_k. Hence

$$\mathbb{B} = \{\psi(e_1^{[1]}), \cdots, \psi(e_r^{[1]}), e, eU, \cdots, eU^{k-2}\}$$

is a basis for W_k.

To describe the matrix of $U - E$ with respect to $\mathbb{B}$, let $\phi_B(t) = \sum_{j=1}^{k} c_j t^{j-1}$ be the correlation polynomial of $B = \xi_1 \cdots \xi_k$, where $c_j = 1$ if $\xi_1 \cdots \xi_j = \xi_{k-j+1} \cdots \xi_k$ and $c_j = 0$ otherwise. Note that $c_k = 1$, so deg $\phi_B(t) = k - 1$. Define

$$C(\phi_B) = \begin{bmatrix} -c_{k-1} & 1 & 0 & \cdots & 0 \\ -c_{k-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -c_2 & 0 & 0 & \cdots & 1 \\ -c_1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

to be its companion matrix. Our definition differs from the usual one by permuting the rows and columns with the permutation $i \to k - i$, but the characteristic polynomial is still $\phi_B(t)$. Let e_j be the $1 \times r$ row vector $[0 \cdots 0\ 1\ 0 \cdots 0]$, and e_i^T be the transpose of e_i. Denote the $m \times n$ zero matrix by $0_{m,n}$.

LEMMA 5. *Let* $B \in \mathscr{B}_k(X_T)$ *be a reduced block with initial state* $i = \beta^k B$ *and terminal state* $j = \eta^k B$. *Then the matrix of* $T\langle B \rangle = U - E$ *with respect to* $\mathbb{B}$ *is*

$$M_{\mathbb{B}}(U - E) = \left[\begin{array}{c|c} T & \begin{array}{cc} -e_i^T & 0_{r,k-2} \end{array} \\ \hline \begin{array}{c} 0_{k-2,r} \\ e_j \end{array} & C(\phi_B) \end{array}\right].$$

Proof. We first compute $M_{\mathbb{B}}(U)$. By Lemma 1, on V_k the matrix of U with respect to the $\psi(e_j^{[1]})$ is just T. Also, by (3.1) and (3.2), $(eU^{k-2})U = \psi(e_j^{[1]})$. Hence if

(3.3) $$J_0(k-1) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

denotes the Jordan nilpotent block of size $k - 1$, we have

$$M_{\mathbb{B}}(U) = \left[\begin{array}{c|c} T & 0_{r,k-1} \\ \hline \begin{array}{c} 0_{k-2,r} \\ e_j \end{array} & J_0(k - 1) \end{array}\right].$$

Next we compute $M_{\mathbb{B}}(E)$. Using (3.1), we see for $m \in \mathscr{S}$ that

$$\psi(e_m^{[1]}) E_{\beta B, \eta B} = \begin{cases} e_{\eta B} & \text{if } m = i = \beta^k B, \\ 0 & \text{otherwise,} \end{cases}$$

and for $0 \leqq m \leqq k - 2$ that

$$(eU^m)E_{\beta B,\eta B} = \begin{cases} e_{\eta B}^{[k]} & \text{if } \eta B \text{ and } \beta B \text{ overlap in } k - m - 1 \text{ symbols,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$M_{\mathbb{B}}(E) = \left[\begin{array}{c|c} 0_{r,r} & \begin{array}{ccc} e_i^T & & 0_{r,k-2} \end{array} \\ \hline & \begin{array}{c} c_{k-1} \\ c_{k-2} \\ \vdots \\ c_1 \end{array} \\ 0_{k-1,r} & \quad\quad 0_{k-1,k-2} \end{array} \right].$$

Subtraction gives the result. $\square$

It is perhaps interesting to note that although $U - E$ is nonnegative integral, we have represented its "interesting part" by a far smaller matrix $M_{\mathbb{B}}(U - E)$ containing negative entries.

4. Zeta functions. We will use the results of the previous section to compute the zeta function of $T\langle B \rangle$ in terms of $\chi_T(t)$, $\phi_B(t)$, and $\text{cof}_{ij}(tI - T)$ described in § 1. Recall our notation from § 1 that $f(t)^\times$ denotes $f(t)$ multiplied by the unique power of t making the product a polynomial with nonzero constant term.

THEOREM 1. *If $B \in \mathscr{B}_k(X_T)$ is reduced with initial state $i = \beta^k B$ and terminal state $j = \eta^k B$, then*

$$\zeta_{T\langle B \rangle}(t) = \frac{1}{[\chi_T(t^{-1})\phi_B(t^{-1}) + \text{cof}_{ij}(t^{-1}I - T)]^\times}.$$

COROLLARY 1. *If $T = [n]$, then for every $B \in \mathscr{B}_k(X_T)$ we have*

$$\zeta_{T\langle B \rangle}(t) = \frac{1}{(1 - nt)\phi_B(t^{-1})t^{k-1} + t^k}.$$

In particular, if B has only the trivial overlap with itself, then

$$\zeta_{T\langle B \rangle}(t) = \frac{1}{1 - nt + t^k}.$$

Proof of Corollary 1. Since $r = 1$, here $\text{cof}_{11}(tI - T) = 1$, and the constant term of

$$(t - n)\phi_B(t) + \text{cof}_{11}(tI - T)$$

is $n\phi_B(0) \pm 1 \neq 0$ since $n \geqq 2$ and $\phi_B(0) = 0$ or 1. Hence

$$[(t^{-1} - n)\phi_B(t^{-1}) + 1]^\times = t^k[(t^{-1} - n)\phi_B(t^{-1}) + 1]$$

$$= (1 - nt)\phi_B(t^{-1})t^{k-1} + t^k.$$

If B overlaps itself only in the entire block, then $\phi_B(t) = t^{k-1}$, completing the proof. $\square$

Proof of Theorem 1. Recall from § 3 our notations that $U = T^{[k]}$, $E = E_{ij}$, $T\langle B \rangle = U - E$, $e = e_{\eta B}$, and that $T\langle B \rangle$ is nilpotent on $\mathbb{R}^{[k]}/W_{k,B}$. The matrix of $T\langle B \rangle$ on $W_{k,B}$ is given in Lemma 5, and expansion by the last row shows

(4.1) $\chi_{T\langle B \rangle | W_{k,B}}(t) = \chi_T(t)\phi_B(t) + \text{cof}_{ij}(tI - T).$

Applying [4, Thm. 1] finishes the proof. $\square$

Bowen and Lanford [4, p. 45] have pointed out that although deg $\zeta_{T\langle B\rangle}^{-1}(t)$ for full shifts is linear in $|B|$, the straightforward calculation of $\chi_{T\langle B\rangle}(t)$ is exponential in $|B|$, and they asked whether there is a more efficient algorithm. Our reduction of $\mathbb{R}^{[k]}$ to W_k provides such an algorithm. The obvious method of computing $\phi_B(t)$ is quadratic in $|B|$, so Theorem 1 provides a computation of $\zeta_{T\langle B\rangle}(t)$ that is also quadratic in $|B|$. However, a slight modification of one part of the Knuth–Morris–Pratt pattern matching algorithm [13] actually gives a linear algorithm. (We are indebted to Andrew Odlyzko for suggesting this possibility.)

COROLLARY 2. *For fixed T there is an algorithm to compute $\zeta_{T\langle B\rangle}(t)$ in time that is linear in $|B|$.*

Proof. By Theorem 1, we need only find a linear algorithm to compute $\phi_B(t)$. Let $B = \xi_1 \cdots \xi_k$. Define $f: \{1, \cdots, k\} \to \{0, \cdots, k - 1\}$ by setting $f(j)$ to be the largest nonnegative $i < j$ such that $\xi_1 \cdots \xi_i = \xi_{j-i+1} \cdots \xi_j$. As in [13, § 2], the table of values of f can be computed in $O(k)$ steps. Our definition of f differs slightly from that in [13], but this has no consequence for the algorithm.

Since f is strictly decreasing, there is a first iterate $f^s(k)$ that equals zero. Let $k_0 = k$, $k_1 = f(k)$, $k_2 = f(f(k))$, $\cdots$, $k_{s-1} = f^{s-1}(k)$. We claim

$$\phi_B(t) = \sum_{j=0}^{s-1} t^{k_j - 1}.$$

For by definition, $k_1 = f(k)$ records the largest nontrivial overlap of B with itself. It follows that the next largest overlap of B with itself coincides with the largest nontrivial overlap of $\xi_1 \cdots \xi_{k_1}$ with itself, which by definition is $f(k_1) = k_2$. This continues until $f^s(k)$ becomes zero.   □

Since the number of k-blocks is roughly λ_T^k, one might expect the number of distinct $\zeta_{T\langle B\rangle}(t)$ to also have exponential growth. Surprisingly, this number turns out to be quite small. For $k = 20$ there are only 116 distinct correlations, and even for $k = 50$ there are only 2,240 [9, p. 29].

COROLLARY 3. *If $\gamma > \frac{1}{2} \log (\frac{3}{2})$, then*

$$|\{ \zeta_{T\langle B\rangle}(t) : B \in \mathscr{B}_k(X_T) \}| = O(k^{\gamma \log k}).$$

Proof. By [9, Thm. 6.1],

$$|\{ \phi_B(t) : B \in \mathscr{B}_k(X_T) \}| = O(k^{\gamma \log k}).$$

Since every block in $\mathscr{B}_k(X_T)$ contains a reduced subblock of length greater than $k - 2r$ determining it, and the number of the polynomials $\mathrm{cof}_{ij} (tI - T)$ is uniformly bounded, the result follows.   □

Another curious fact also follows.

COROLLARY 4. *Suppose $m, n \geq 2$. The number of distinct zeta functions produced by omitting the blocks of length k from the full m-shift coincides with that from the full n-shift.*

Proof. This follows from Corollary 1 together with the result [9, Cor. 5.1] that the set of correlations is independent of the size of the alphabet.   □

Because of the possible cancellation of lower powers of t in (4.1), Theorem 1 does not immediately yield deg $\zeta_{T\langle B\rangle}^{-1}(t)$. In particular, a good lower bound is not obvious.

THEOREM 2. *If $B \in \mathscr{B}_k(X_T)$ is reduced with initial state i and terminal state j, then*

$$k - 4r \leqq \deg \zeta_{T\langle B\rangle}^{-1}(t) \leqq k + r - 1.$$

The upper bound is attained if and only if

(4.2) $(\det T)\phi_B(0) + \text{cof}_{ij}(-T) \neq 0$,

and this occurs for all blocks in a full shift.

   Proof. Using the notations from the proof of Theorem 1, we see that

$$\deg \chi_{T\langle B\rangle \mid W_k}(t) = \dim W_k = k + r - 1,$$

and that $T\langle B\rangle$ is invertible on W_k if and only if the constant term in (4.1)

$$\chi_T(0)\phi_B(0) + \text{cof}_{ij}(-T) = (\det T)\phi_B(0) + \text{cof}_{ij}(-T) \neq 0.$$

For full shifts, the proof of Corollary 1 shows the constant term is nonzero for all blocks, completing the proof of the upper bound statements.

   For the lower bound, we will prove by induction on k that the highest power P of t dividing (4.1) is less than $5r$. Since the degree is $k + r - 1$, the lower bound will follow. For notational simplicity, put $c_{ij}(t) = \text{cof}_{ij}(tI - T)$. A crucial fact that we use repeatedly is that $c_{ij}(\lambda_T) \neq 0$ for all i, j [16, p. 7].

   If $k \leq 4r$, then for every $B \in \mathcal{B}_k(X_T)$ the degree of (4.1) is $< 5r$. Also, at $t = \lambda_T$ this polynomial has value $\text{cof}_{ij}(\lambda_T I - T) \neq 0$ by [16, p. 7]. Hence if $k \leq 4r$, the highest power P of t dividing (4.1) has $P < 5r$.

   Now fix $k > 4r$, and assume our inductive hypothesis for all blocks of length $< k$. Let $B \in \mathcal{B}_k(X_T)$. Choose $p \geq 1$ minimal so B overlaps itself in $k - p$ symbols. If B has only trivial overlaps, define p to be k. This p is the *fundamental* period of B as defined in [9]. We distinguish two cases, depending on the relationship of p to r.

   First suppose $p > r$. If $p = k$ the result is trivial since then $\phi_B(t) = t^{k-1}$, $\deg c_{ij}(t) \leq r - 1$, and $c_{ij}(\lambda_T) \neq 0$, so $P \leq r - 1$. If $p < k$, let $C = \eta^p B$. Then C has initial state i, terminal state j, and $\phi_B(t) = t^{k-1} + \phi_C(t)$ with $\deg \phi_C(t) = k - p - 1$. Since $\deg \chi_T(t) = r$, and $p > r$, the highest power of t dividing (4.1) coincides with the highest power dividing $\chi_T(t)\phi_C(t) + c_{ij}(t)$, which is less than $5r$ by induction.

   Next suppose $p \leq r$. We first treat the case $t^p - 1 \nmid \chi_T(t)$. By the recursive description of $\phi_B(t)$ in [9, Thm. 5.1], it follows that

(4.3) $$\phi_B(t) = \sum_{m=0}^{\lfloor (k-p-1)/p \rfloor} t^{k-1-mp} + \psi(t),$$

where $\deg \psi(t) < 2p$. We claim that for every $K \geq 1$, in the product

$$(t^{pK} + t^{p(K-1)} + \cdots + t^p + 1)\chi_T(t)$$

every set of p consecutive powers of t in $[0, Kp + r]$ has at least one nonzero coefficient. For if not, then letting $|t| < 1$ and $K \to \infty$ shows that $(t^p - 1)^{-1}\chi_T(t) \in \mathbb{Z}[t]$, contradicting $t^p - 1 \nmid \chi_T(t)$. Substituting (4.3) into (4.1) and applying our claim, it follows that (4.1) has at least one nonzero coefficient of a power of t in the range $[\max\{r, 2p\} + 1, \max\{r, 2p\} + p]$, showing that $P \leq 3r$ in this case.

   Finally, suppose $p \leq r$ and $t^p - 1 \mid \chi_T(t)$. By (4.3),

$$\phi_B(t) = t^{p+l}\left(\frac{t^{pK} - 1}{t^p - 1}\right) + \psi(t),$$

where $K = \lfloor (k - p - 1)/p \rfloor$, $l < p$, and $\deg \psi(t) < 2p$. Let $\chi_T(t) = (t^p - 1)\tilde{\chi}(t)$, so $\tilde{\chi}(\lambda_T) = 0$. Then

$$\chi_T(t)\phi_B(t) + c_{ij}(t) = \tilde{\chi}(t)t^{p(K+1)+l} + \tilde{\chi}(t)[(t^p - 1)\psi(t) - t^{p+l}] + c_{ij}(t).$$

If $\rho(t) = \tilde{\chi}(t)[(t^p - 1)\psi(t) - t^{p+l}] + c_{ij}(t)$, then $\rho(\lambda_T) = c_{ij}(\lambda_T) \neq 0$, so $\rho(t)$ is not identically zero. Also, deg $\rho(t) < r + 2p \leqq 3r$. Since

$$p(K+1) + l \geqq p\left(\frac{k-1}{p} - 1\right) = k - p - 1 \geqq 3r,$$

the highest power P of t dividing (4.1) coincides with that dividing $\rho(t)$, so $P \leqq 3r$ in this case. $\square$

5. Entropy. How does entropy change when one long block is removed? Since $h(\sigma_T) = \log \lambda_T$, this amounts to determining the change in spectral radius. Standard Perron–Frobenius theory [16, Thm. 1.1(e)] shows $\lambda_T > \lambda_{T\langle B\rangle}$. The following considerations allow us to guess that the difference is exponentially small in $|B|$. We can consider $T\langle B\rangle$ to be the end result of $U - tE$ as t varies from 0 to 1. To find the change in eigenvalue, it becomes important to know its rate of change, say at $t = 0$. Surprisingly, this rate is a simple function of the left and right eigenvectors.

LEMMA 6. *Let $A = [a_{ij}]$ be a real square matrix with simple eigenvalue λ and corresponding left eigenvector v and right eigenvector w. Then*

$$\left[\frac{\partial\lambda}{\partial a_{ij}}\right] = \frac{wv}{vw}.$$

Proof. Since λ is simple, it is known [12, Thm. II.1.8] that λ, v, and w are analytic functions of a_{ij}. Take ∂/∂_{ij} of $Aw = \lambda w$, use $\partial A/\partial a_{ij} = E_{ij}$, and multiply by v on the left to obtain

$$vA\frac{\partial w}{\partial a_{ij}} + vE_{ij}w = v\frac{\partial\lambda}{\partial a_{ij}}w + \lambda v\frac{\partial w}{\partial a_{ij}}.$$

Cancelling the terms involving $vA = \lambda v$ shows that $v_i w_j = (vw)(\partial\lambda/\partial a_{ij})$. Since λ is simple, $vw \neq 0$, concluding the proof. $\square$

Let the notation $f(k) \asymp g(k)$ as $k \to \infty$ mean that there are $a, b > 0$ so that for sufficiently large k we have $ag(k) < f(k) < bg(k)$.

Now fix T, and let $v > 0$ and $w > 0$ be the left and right eigenvectors for λ_T. Put $c = \min\{v_i, w_i\}$ and $d = \max\{v_i, w_i\}$. Recalling the map ψ defined in (3.1), we have that $\psi(v), \psi(w)$ are left and right eigenvectors for $U = T^{[k]}$, and have entries between c and d. Since these vectors have length about λ_T^k, it follows $\psi(v)\psi(w) \asymp \lambda_T^k$. Hence by Lemma 6, the derivative of λ_{U-tE} at $t = 0$ is within two positive constants of λ_T^{-k}. If we were able to extend this estimate of the derivative to all $0 \leqq t \leqq 1$, we would be done. However, the estimate runs into trouble as t becomes positive since the eigenvectors also vary with t.

Nevertheless, the linear algebra of § 3 enables us to prove a sharp exponential estimate.

THEOREM 3. *There are constants $c_T, d_T > 0$ so that if k is sufficiently large, for every $B \in \mathscr{B}_k(X_T)$ we have*

$$c_T\lambda_T^{-k} < h(\sigma_T) - h(\sigma_{T\langle B\rangle}) < d_T\lambda_T^{-k}.$$

Proof. In the following, the a_j are suitable positive constants. Let us first consider the case $T = [n]$ of the full n-shift. Then $\lambda_T = n$ and $\lambda_{T\langle B\rangle}$ satisfies $(t - n)\phi_B(t) + 1 = 0$. By [8, Lemma 3], $|\phi_B(z)| > a_1(1.7)^k$ for $|z| \geqq 1.7$ and suitable $a_1 > 0$. Thus if $|z| = 1.7$ we have

$$(5.1) \qquad\qquad |(z-n)\phi_B(z)| \geqq (0.3)a_1(1.7)^k > 1$$

for k large enough. We use this to count roots by invoking the following result from complex analysis.

ROUCHÉ'S THEOREM. *Let $f(z)$ and $g(z)$ be analytic on $\{|z| \leq R\}$ and satisfy $|f(z) - g(z)| < |f(z)|$ for $|z| = R$. Then $f(z)$ and $g(z)$ have the same number of zeros in $\{|z| < R\}$.*

Using (5.1) and Rouché's Theorem, it follows that $(z - n)\phi_B(z)$ and $(z - n)\phi_B(z) + 1$ have the same number of roots in $|z| \geq 1.7$, namely 1. This shows that $\lambda = \lambda_{T\langle B\rangle} \geq 1.7$. Hence

$$\frac{1}{\lambda^{k-1} + \lambda^{k-2} + \cdots + 1} \leq |n - \lambda| = \frac{1}{\phi_B(\lambda)} = \frac{1}{\lambda^{k-1} + c_{k-2}\lambda^{k-2} + \cdots + c_0} \leq \frac{1}{\lambda^k}.$$

Thus

$$|n - \lambda_{T\langle B\rangle}| \asymp \lambda_{T\langle B\rangle}^{-k} \quad \text{for } B \in \mathscr{B}_k(X_T) \text{ as } k \to \infty.$$

We claim that actually $|n - \lambda_{T\langle B\rangle}| \asymp n^{-k}$, for which it suffices to show that $|n - \lambda_{T\langle B\rangle}| \leq a_2 n^{-k}$ for suitable $a_2 > 0$ since $\lambda = \lambda_{T\langle B\rangle} < n$. Now

$$|n - \lambda| \leq a_3 \lambda^{-k} = a_3 n^{-k \log \lambda / \log n},$$

and since $|n - \lambda| < a_4 \lambda^{-k}$, it follows, using differentiability of $\log x$ at $x = n$, that $-\log \lambda / \log n < -1 + a_5 \lambda^{-k} < -1 + a_5(1.7)^{-k}$. Hence

$$|n - \lambda| \leq a_3 n^{-k} n^{a_5 k(1.7)^{-k}}.$$

Putting

$$a_2 = a_3 \min_{1 \leq k < \infty} \{n^{a_5 k(1.7)^{-k}}\} < \infty$$

completes the proof for the full shift.

To extend these ideas to general T, we need a version of (5.1) that works for smaller $|z|$.

LEMMA 7. *Fix $\rho > 1$. Then*

$$\inf_{B \in \mathscr{B}_k(X_T)} \inf_{|z| \geq \rho} |\phi_B(z)| \to \infty \quad \text{as } k \to \infty.$$

Proof. Recall from the proof of Theorem 2 that if p is the fundamental period for B, then

$$\phi_B(t) = \sum_{m=0}^{\lfloor (k-p-1)/p\rfloor} t^{k-1-mp} + \psi(t),$$

where $\deg \psi(t) < 2p$, and $\psi(t)$ has coefficients that are 0 or 1. Fix $M > 0$. If $p > k/10$, then for all $|z| \geq \rho$ we have

$$|\phi_B(z)| \geq |z|^{k-1} - 2\sum_{m=0}^{\lfloor 9k/10\rfloor} |z|^m \geq M$$

provided k is large enough. If $p \leq k/10$, then for $K = \lfloor (k-1)/p\rfloor$ we have

$$\phi_B(t) = t^l\left(\frac{t^{Kp} - 1}{t^p - 1}\right) + \psi(t),$$

where $l < p$ and deg $\psi(t) < 2p$. Hence for all $|z| > \rho$ we have that

$$|\phi_B(z)| \geqq \left|z^l\left(\frac{z^{Kp}-1}{z^p-1}\right)\right| - \sum_{m=0}^{2p}|z|^m$$

$$\geqq \frac{|z|^{9k/10} - |z|^{k/10}}{|z|^{k/10}+1} - \frac{k}{5}|z|^{k/5} \geqq M$$

if k is sufficiently large. $\square$

We now prove Theorem 3 for general T. We know that $\lambda = \lambda_{T\langle B\rangle}$ satisfies

(5.2) $\chi_T(t)\phi_B(t) + c_{ij}(t) = 0$,

where $c_{ij}(t) = \text{cof}_{ij}(tI - T)$ is one of a finite set of polynomials. Since $\chi_T(t) = (t - \lambda_T)q(t)$, where $q(\lambda_T) \neq 0$, we have

$$|\lambda_T - t| = \frac{|c_{ij}(t)|}{|\phi_B(t)q(t)|}.$$

Let the roots of $\chi_T(t)$ be $\lambda_1 = \lambda_T, \lambda_2, \cdots, \lambda_r$. Fix ρ with $\lambda_T > \rho > \max_{2 \leq j \leq r}|\lambda_j|$. Apply Lemma 7 to conclude by Rouché's Theorem that for sufficiently large k (5.2) has exactly one solution for $|t| > \rho$, namely $t = \lambda_{T\langle B\rangle}$. Hence

$$|\lambda_T - \lambda_{T\langle B\rangle}| = \frac{|c_{ij}(\lambda_{T\langle B\rangle})|}{|\phi_B(\lambda_{T\langle B\rangle})q(\lambda_{T\langle B\rangle})|} \to 0 \quad \text{as } k \to \infty.$$

Since $q(z) \neq 0$ if $|z| > \rho$, we have $|q(\lambda_{T\langle B\rangle})| \asymp 1$ as $k \to \infty$. Furthermore, as remarked above, the Perron–Frobenius theory shows that $c_{ij}(\lambda_T) \neq 0$ for all i, j, so $|c_{ij}(\lambda_{T\langle B\rangle})| \asymp 1$ as $k \to \infty$. Since $|\phi_B(\lambda_{T\langle B\rangle})| \asymp \lambda_{T\langle B\rangle}^k$, we obtain

$$|\lambda_T - \lambda_{T\langle B\rangle}| \asymp \lambda_{T\langle B\rangle}^{-k}.$$

The conclusion that this forces

$$|\lambda_T - \lambda_{T\langle B\rangle}| \asymp \lambda_T^{-k} \quad \text{as } k \to \infty$$

follows exactly as in the full shift case. $\square$

We remark that these estimates may also be proved using generating functions, in the spirit of the calculations in [8, § 2] for the full shift. Let $f_n^{(ij)}$ be the number of blocks in $\mathscr{B}_n(X_T)$ from i to j not containing B, and put $F_{ij}(z) = \sum_{n=0}^{\infty} f_n^{(ij)} z^{-n}$, and set $F(z) = [F_{ij}(z)]$. Then developing matrix analogues of the polynomial relations in [8, § 1], we can show that

$$F(z) = (I - z^{-(k-1)}E)[zI - T + \phi_B(z)^{-1}TE]^{-1},$$

where $E = E_{ij}$. The largest root of the denominator is $\lambda_{T\langle B\rangle}$, and we can derive the exponential estimate of Theorem 3 from this. The details are more involved than for the proof given here. This generating function technique could also be used when omitting a finite number of blocks, as in [10, § 2]. This yields a nonsingular system of matrix equations whose solution would allow an estimate of the resulting spectral radius. However, the results are not nearly as explicit as for one block.

We also remark that much finer information about $|\lambda_T - \lambda_{T\langle B\rangle}|$ can be obtained from the proof of Theorem 3. For example, if $T = [n]$ and $B \in \mathscr{B}_k(X_T)$ has only the trivial overlap with itself, then

$$\lambda_{T\langle B\rangle} = n - \frac{1}{n^{k-1}} - \frac{k-1}{n^{2k-1}} + O\left(\frac{k^2}{n^{3k}}\right).$$

Such estimates can be proved directly [8, Lemma 4], or with the Lagrange inversion formula [17, Thm. 2].

6. Jordan forms. Let $J^\times(T)$ be the invertible part of the Jordan form $J(T)$ for T, and $J^0(T)$ be its nilpotent part. Williams [18] proved that $J^\times(T)$ is a topological invariant of σ_T. Although $J^0(T)$ is not an invariant, our analysis allows us to compute $J^0(T^{[k]})$ and, in many cases, $J^0(T\langle B\rangle)$.

Denote by $J_0(m)$ the elementary $m \times m$ Jordan block for eigenvalue 0 as in (3.3), and by $J_0(m)^{\oplus n}$ the direct sum of n copies of $J_0(m)$. We first show that in passing from T to $T^{[2]}$, each Jordan nilpotent block of $J^0(T)$ increases by one dimension, and enough one-dimensional nilpotent blocks are added to account for the rest of dim $\mathbb{R}^{[2]}$.

LEMMA 8. *Suppose T has Jordan form $J^\times(T) \oplus J^0(T)$, where*

$$J^0(T) = J_0(1)^{\oplus n_0} \oplus J_0(2)^{\oplus n_1} \oplus \cdots \oplus J_0(p)^{\oplus n_{p-1}}.$$

Then $T^{[2]}$ has Jordan form $J^\times(T) \oplus J^0(T^{[2]})$, where

$$J^0(T^{[2]}) = J_0(1)^{\oplus n_{-1}} \oplus J_0(2)^{\oplus n_0} \oplus \cdots \oplus J_0(p+1)^{\oplus n_{p-1}},$$

and

$$(6.1) \qquad n_{-1} = \sum_{i,j=1}^{r} T_{ij} - \dim J^\times(T) - \sum_{l=1}^{p} (l+1)n_{l-1}.$$

Proof. Since T is assumed irreducible, for each $i \in \mathcal{S}$ we can choose $\xi(i) \in \mathcal{E}$ with $\eta(\xi(i)) = i$. Define $g: \mathbb{R}^{[1]} \to \mathbb{R}^{[2]}$ by $g(\sum_{i=1}^{r} a_i e_i^{[1]}) = \sum_{i=1}^{r} a_i e_{\xi(i)}^{[2]}$. Recalling the map ψ defined in (3.1), note that $T^{[2]} \circ g = \psi$. If $\{v_1, \cdots, v_p\}$ is a Jordan basis for $J^\times(T)$, then $\{\psi(v_1), \cdots, \psi(v_p)\}$ is one for $J^\times(T^{[2]})$ by Lemma 1. Let V denote the span of the $\psi(v_m)$. Suppose $w \in \mathbb{R}^{[1]}$ generates one of the Jordan nilpotent blocks of size s in $J^0(T)$. Then $g(w)$ generates a Jordan nilpotent block of size $s+1$ in $J^0(T^{[2]})$ since $g(w)T^{[2]} = \psi(w)$. Thus by mapping each Jordan nilpotent generator in $J^0(T)$ to its image under g, we obtain vectors in $\mathbb{R}^{[2]}$ with nilpotency ≥ 2 under $T^{[2]}$. On the subspace W generated by powers of $T^{[2]}$ on these vectors the matrix of $T^{[2]}$ is

$$J_0(2)^{\oplus n_0} \oplus J_0(3)^{\oplus n_1} \oplus \cdots \oplus J_0(p+1)^{\oplus n_{p-1}}.$$

For each $i \in \mathcal{S}$ let

$$\mathcal{S}(i) = \{\xi \in \mathcal{E} : \eta(\xi) = i, \xi \neq \xi(i)\}.$$

The vectors $\{e_\xi^{[2]} - e_{\xi(i)}^{[2]} : \xi \in \mathcal{S}(i), i \in \mathcal{S}\}$ are each annihilated by $T^{[2]}$ and span the remaining part of $\mathbb{R}^{[2]}$. It follows as in the proof of the Jordan form [10, § 7.3] that $V \oplus W$ has a $T^{[2]}$-invariant complement on which $T^{[2]}$ is 0. Since $\dim \mathbb{R}^{[2]} = \sum_{ij} T_{ij}$, the dimension n_{-1} of this complement is given in (6.1), concluding the proof. $\quad\square$

Since $T^{[k]} \cong (T^{[k-1]})^{[2]}$, Lemma 8 allows the inductive determination of $J^0(T^{[k]})$. It is convenient to introduce the sequence $\{n_q(T) : q \in \mathbb{Z}\}$ of integers defined as follows. For $q \geq 0$ the $n_q(T)$ are determined by

$$J^0(T) = \bigoplus_{q=0}^{\infty} J_0(q+1)^{\oplus n_q(T)},$$

so $n_q(T) = 0$ for sufficiently large q. For $q < 0$ we use the backward recursion

$$(6.2) \qquad n_q(T) = \sum_{i,j=1}^{r} (T^{|q|})_{ij} - \dim J^\times(T) - \sum_{l=1}^{\infty} (l+1)n_{l+q},$$

where the infinite series is really finite since $n_{l+q} = 0$ for large l.

THEOREM 4. *If $n_q(T)$ is defined as above, then*

$$J^0(T^{[k]}) = \bigoplus_{q=0}^{\infty} J_0(q+1)^{\oplus n_{q+1-k}(T)}.$$

Proof. If $k = 1$ this follows by definition, while Lemma 8 shows this to be true for $k = 2$. The general result follows by induction on k. □

For a concrete example, consider the full 2-shift $T = [2]$. The theorem implies that for all $k \geq 3$ we have

$$J^0(T^{[k]}) = J_0(1)^{\oplus 2^{k-3}} \oplus J_0(2)^{\oplus 2^{k-4}} \oplus \cdots \oplus J_0(k-2)^{\oplus 2^0} \oplus J_0(k-1).$$

Mike Boyle (private communication) has pointed out to us that

(6.3) $\text{rk}\,(T^{[k]})^p = \text{rk}\,(T^{[k+l]})^{p+l}$

for $l, p \geq 0$ and $k \geq 1$. The proof is analogous to that of Lemma 8. If S is any matrix, then the number of Jordan nilpotent blocks of size $m \geq 1$ is $\text{rk}\,(S^{m+1}) - 2\,\text{rk}\,(S^m) + \text{rk}\,(S^{m-1})$. Hence (6.3) yields an alternative proof of Theorem 4.

We now show that under some circumstances we can determine the Jordan form of $T\langle B \rangle$.

THEOREM 5. *Suppose T is invertible, that $B \in \mathscr{B}_k(X_T)$ is reduced with initial state i and terminal state j, and that*

(6.4) $(\det T)\phi_B(0) + \text{cof}_{ij}(-T) \neq 0.$

This condition is met by all blocks in a full shift. Then $T\langle B \rangle$ is invertible on W_k, $J^\times(T\langle B \rangle)$ is just the Jordan form of the restriction of $T\langle B \rangle$ to W_k, and $J^0(T\langle B \rangle)$ is obtained from $J^0(T^{[k]})$ by deleting one copy of $J^0(k - 1)$.

Proof. Use the notations of § 3, with $U = T^{[k]}$, $e = e_{\eta B}^{[k]}$, $E = E_{\beta B, \eta B}$, and V_k and W_k the subspaces described there. Since T is invertible, there is a $v \in V_k$ such that $vU^{k-1} = eU^{k-1}$. Since $eU^{k-2} \notin V_k$, it follows that $w = e - v$ generates a nilpotent Jordan block under U of maximal size, and that W_k is the direct sum of the corresponding subspace and V_k. The proof of the Jordan Theorem [10, § 7.3] shows that there is a Jordan basis for U with w as one of its generating basis vectors. Since E vanishes on the direct complement of W_k, it follows that passing from U to $U - E = T\langle B \rangle$ preserves the direct complement Jordan structure. By (6.4), it follows as in Theorem 2 that $U - E$ is invertible on W_k. Thus in passing from U to $U - E$, one Jordan nilpotent block of size $k - 1$ combines with V_k to form the subspace W_k on which the invertible part of $U - E$ acts, while the remaining Jordan nilpotent blocks remain undisturbed. □

REFERENCES

[1] R. ADLER, D. COPPERSMITH, AND M. HASSNER, *Algorithms for sliding block codes*, IEEE Trans. Inform. Theory, 29 (1983), pp. 5–22.

[2] M. ARTIN AND B. MAZUR, *On periodic points*, Ann. of Math., 81 (1965), pp. 82–89.

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

[4] R. BOWEN AND O. LANFORD III, *Zeta functions of restrictions of the shift transformation*, AMS Proc. Symp. Pure. Math., 14 (1970), pp. 43–49.

[5] M. BOYLE, D. LIND, AND D. RUDOLPH, *The automorphism group of a shift of finite type*, Trans. Amer. Math. Soc., 306 (1988), pp. 71–114.

[6] J. E. DAYHOFF AND G. L. GERSTEIN, *Favored patterns in spike trains*, J. Neurophysiol., 49 (1983), pp. 1334–1363.

[7] E. N. GILBERT, *Synchronization of binary messages*, IRE Trans. Inform. Theory, IT-6 (1960), pp. 470–477.

[8] L. J. GUIBAS AND A. M. ODLYZKO, *Maximal prefix-synchronized codes*, SIAM J. Appl. Math., 35 (1978), pp. 401–418.

[9] L. J. GUIBAS AND A. M. ODLYZKO, *Periods in strings*, J. Combin. Theory Ser. A, 30 (1981), pp. 19–42.

[10] L. J. GUIBAS AND A. M. ODLYZKO, *String overlaps, pattern matching, and non-transitive games*, J. Combin. Theory Ser. A, 30 (1981), pp. 183–208.

[11] K. HOFFMAN AND R. KUNZE, *Linear Algebra*, 2nd ed., Prentice-Hall, Englewood Cliffs, 1971.

[12] T. KATO, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1982.

[13] D. E. KNUTH, J. H. MORRIS, AND V. R. PRATT, *Fast pattern matching in strings*, SIAM J. Comput., 6 (1977), pp. 323–350.

[14] B. MARCUS, *Sofic systems and encoding data*, IEEE Trans. Inform. Theory, 31 (1985), pp. 366–377.

[15] W. PARRY AND S. TUNCEL, *Classification Problems in Ergodic Theory*, LMS Lecture Notes 67, Cambridge University Press, Cambridge, 1982.

[16] E. SENETA, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[17] H. S. WILF, *Strings, substrings, and the "nearest integer" function*, Amer. Math. Monthly, 94 (1987), pp. 855–860.

[18] R. WILLIAMS, *Classification of subshifts of finite type*, Ann. of Math., 98 (1973), pp. 120–153; *errata*, 99 (1974) pp. 380–381.