SYMBOLIC DYNAMICS AND MATRICES

MIKE BOYLE*

Contents

Introduction. The main purpose of this article is to give some overview of matrix problems and results in symbolic dynamics. The basic connection is that a nonnegative integral matrix A defines a topological dynamical system known as a shift of finite type. Questions about these systems are often equivalent to questions about "persistent" or "asymptotic" aspects of nonnegative matrices. Conversely, tools of symbolic dynamics can be used to address some of these questions. At the very least, the ideas of conjugacy, shift equivalence and strong shift equivalence give viewpoints on nonnegative matrices and directed graphs which are at some point inevitable and basic (although accessible, and even elementary).

My motivation for this article was to try to communicate some of this to matrix theorists. The earlier sections are more descriptive. The later sections move more to current frontiers and are oriented more to presenting open problems.

Trying to stay close to matrices, I've neglected huge parts of symbolic dynamics. Also even some matrix matters get short shrift. I've barely mentioned state splitting (3.4) and resolving maps (Sec. 10), which are important for constructions [AM,BMT,A2] and applications [ACH,AFKM,MSW]. Marker methods are a key to some inverse spectral results, but to avoid a nontrivial excursion into nonmatrix matters I just recommend some references (Sec.8). For more general background, one can dig into [DGS],[BMT],[PT1],[P2] and their references. Unfortunately, at present there is no single book which gives an appropriate introduction (although one by Adler, Lind and Marcus could appear any year now).

This article is a followup to the talk I was invited to give at the November 1991 I.M.A. Workshop on Combinatorial and Graph-theoretic problems in Linear Algebra. The talk was on joint work with David Handelman [BH1,2] on solving inverse spectral (and other) problems for nonnegative matrices, using tools from symbolic dynamics. My warm thanks go to the organizers Richard Brualdi, Shmuel

SECTION 1. SHIFTS OF FINITE TYPE

1.1 Subshifts. For the purposes of this paper, a topological dynamical system will be a continuous map T from a compact metric space X into itself. We can represent this as (X, T) or just T. Except in Section 9, T is a homeomorphism.

The system which is the full shift on n symbols (know more succinctly as the n-shift) is defined as follows. We endow a finite set of n elements–say, $\{0, 1, ..., n-1\}$– with the discrete topology. (This finite set is often called the alphabet.) We let X be the countable product of this set, indexed by $\mathbb{Z}$. We think of an element of X as a doubly infinite sequence $x = ...x_{-1}x_0x_1...$ where each x_i is one of the n elements. X is given the product topology and thus becomes a compact zero dimensional metrizable space. A metric compatible with the topology is given by (for x not equal to y)

$$\text{dist}(x, y) = 1/(k + 1), \quad \text{where } k = \min\{|i| : x_i \neq y_i\}.$$

That is, two sequences are close if they agree in a large stretch of coordinates around the zero coordinate.

A finite sequence of elements of the alphabet is called a word. If W is a word of length $j - i + 1$, then the set of sequences x such that $x_i...x_j = W$ is called a cylinder set. The cylinder sets are closed and open, and they give a basis for the product topology on X.

There is a natural shift map S sending X into X, defined by shifting the index set by one: $(Sx)_i = x_{i+1}$. It is easy to see that S is bijective, S sends cylinders to cylinders, and thus S is a homeomorphism. The full shift on n symbols is the system (X, S).

A subshift is a subsystem of some full shift on n symbols. This means that it is a homeomorphism obtained by restriction of the shift to some compact set Y invariant under the shift and its inverse. The complement of Y is open and is thus a union of cylinder sets. Because Y is shift invariant, it follows that there is a (countable) list of words such that Y is precisely the set of all sequences y such that for every word W on the list, for every $i \leq j$, W is not equal to $y_i...y_j$. If Y is a set which may be obtained by forbidding a finite list of words, then the subshift is called a subshift of finite type, or just a shift of finite type (SFT). For example, we get an SFT by restricting the two-shift to the set Y of sequences in which the word 00 never occurs.

1.2 Vertex shifts. We will define vertex shifts, which are examples of shifts of finite type. For some n, let A be an $n \times n$ zero-one matrix. We think of A as the adjacency matrix of a directed graph with n vertices; the vertices index the rows and the columns, and $A(i, j)$ is the number of edges from vertex i to vertex j. Let Y be the space of doubly infinite sequences y such that for every k in Z,

1.3 Edge shifts. Again let A be an adjacency matrix for a directed graph, but now allow multiple edges: so, the entries of A are nonnegative integers. Let the set of edges be the alphabet. Let Y be the set of sequences y such that for all k, the terminal vertex of y_k is the initial vertex of y_{k+1}. Again, we can think of Y as the space of doubly infinite walks through the graph, now presented by the edges traversed. The shift map restricted to Y is an edge shift and it is a shift of finite type: a sufficient list of forbidden words is the set of edge pairs ij which do not satisfy the head-to-tail rule.

In the sequel, unless otherwise indicated an SFT defined by a matrix A is intended to be the edge shift defined by A. We denote this SFT by S_A.

1.4 Codes. Suppose (X, S) and (Y, T) are subshifts. A map f from X to Y is called a code if it is continuous and intertwines the shifts, i.e. $fS = Tf$. We think of a code as a homomorphism of these dynamical systems.

Now suppose F is a function from words of length $2n + 1$ which occur in S-sequences into some finite set A. Then the rule

$$(fx)_i = F(x_{i-n}...x_{i+n}), \quad \text{for all } i \text{ in } \mathbb{Z},$$

defines a code f, called a block code, into the full shift on the alphabet A. This block code defines a code from S into any subshift T which contains the image of f. The "Curtis-Hedlund-Lyndon Theorem" asserts that every code is given by a block code. The argument is easy: given f, one obtains F and n above for $i = 0$ as a consequence of uniform continuity, and the formula for all i follows because f intertwines the shifts.

If a code f is surjective, then it is called a quotient or factor map. If it is injective, then it is called an embedding. If it is injective and surjective, then it is an isomorphism or conjugacy of subshifts. This notion of isomorphism is our fundamental equivalence relation.

To expand on this a little, think of a code f from S_A to S_B as a map of infinite paths in graphs. If we think of x_i as our location on this path at time i, then we think of $(fx)_i$ as our location on the image path at time i. The rule F above determines that location $(fx)_i$, knowing the location x_i with memory of the last n locations and anticipation of the next n locations. The same rule F works for any i. If f is an isomorphism, then in a strong sense the structure of these infinite path spaces is essentially the same.

1.5 Higher block presentations. Let S be a subshift. Suppose n is a positive integer and j, k are nonnegative integers such that $j + k + 1 = n$. Given this we will define a code f with domain the subshift S by the rule

clearly one-to-one.) The system T doesn't depend on the choice of j (although the map does). T is called the n-block presentation of S.

An easy exercise is to construct a one-block isomorphism between the n-block presentation of S and the subshift obtained by passing to the 2-block presentation $n-1$ times.

For an important example, let S be the edge shift defined by a matrix A. Let G be the graph with adjacency matrix A. A symbol in the alphabet of the two block presentation is a word uv, where u and v are edges and the terminal vertex of u equals the initial vertex of v (i.e. uv is a path of length 2 in G). We can define a new graph G' whose vertices are the edges of G, and where there is an arc from u to v if the terminal vertex of u equals the initial vertex of v. If we give such an arc the name uv, then we see that the two-block presentation of S_A is the SFT presented by the matrix which is the adjacency matrix of G'. That is, for SFT's defined by matrices, passing to the two-block presentation amounts to passing from the defining graph to its edge graph (remember, all our graphs are directed—the edge graph in this category is the directed graph we've just described, it rarely has a symmetric adjacency matrix).

Similarly, we can think of the n-block presentation of S_A as given by a graph $G(n)$ whose vertices are the paths of length $n-1$ in G. Here (for $n > 2$) there is an edge from vertex $a(1)...a(n-1)$ to a vertex $b(1)...b(n-1)$ iff $a(2)...a(n-1) = b(1)...b(n-2)$. (For $n = 2$, there is an edge from vertex $a(1)$ to vertex $b(1)$ iff the terminal vertex of $a(1)$ equals the initial vertex of $b(1)$.) Note, if the original subshift contains infinitely many points, then as n goes to infinity the size of the adjacency matrix for $G(n)$ must go to infinity. In particular matrices of very different size may define isomorphic SFT's.

1.6 One-block codes. Suppose f is a code from S to T. Then there is some n such that for all x, $(fx)_0$ is determined by the word $x_{-n}...x_n$ of length $2n + 1$. Define an isomorphism from S to its $(2n+1)$-block presentation S' as in §1.5, using $j = k = n$. Let h be the inverse of this isomorphism. Then fh is a code from S' to S_B, and fh is a one-block map. Often, given a code from S, by passing to a higher block presentation in this way we can assume that the code is just a one-block map.

For example, if there is a map from S_A to S_B, then there is a one-block map from S_C to S_B, where C is a matrix giving some higher-block presentation of A. That is, by passing to an iterated edge graph H of the graph with adjacency matrix A (H has adjacency matrix C), there is a graph homomorphism from H to the graph with adjacency matrix B which (applied at every edge along a path) gives a map from the set of infinite A-paths to the set of infinite B-paths.

1.7 Isomorphic SFT's. Any SFT (Y, S) is isomorphic to a vertex shift. To

$u = u_1...u_n$ to $v = v_1...v_n$ if $uv_n = u_1v$ and this word of length $n+1$ occurs in a sequence in Y. This graph defines a vertex SFT (X, T). An isomorphism f from Y to X is given by the rule $(fy)_i = y_i...y_{i+n-1}$.

Also, any SFT is isomorphic to an edge shift, because the two-block presentation of a vertex shift is an edge shift.

Even if one is only interested in SFT's defined from graphs, it is useful to consider general n-step SFT's, because working with these gives access to topological and combinatorial arguments which can in turn yield results about the graphically defined SFT's. The vertex shifts are sometimes more simple to work with than the edge shifts. The edge shifts are very useful. One reason is conciseness: an edge shift presented by a small matrix (perhaps though with large entries) may be presentable as a vertex shift only by a large matrix. Also, the set of zero-one matrices (the matrices which define vertex shifts) is not closed under various operations under which the set of nonnegative integer matrices is closed. Working only with zero-one matrices rules out some very useful matrix arguments (e.g. [F2]) and interpretations. For one of these, first a little preparation.

If S is a subshift, then we let S^n denote the homeomorphism obtained by iterating S n times. The homeomorphism S^n is isomorphic to a subshift T whose alphabet is the set of S-words of length n. An isomorphism from S^n to T is given by the map f which sends a point x to the point y such that for all k in $\mathbb{Z}$,

$$y_k = x_{kn}...x_{(k+1)n-1}.$$

Now, let an edge shift S be defined by a matrix A. Then the subshift S^n is conjugate to the edge shift defined by A^n. The number of edges from vertex i to vertex j in the directed graph with adjacency matrix A^n is just the (i, j) entry of the matrix A^n. This is also the number of paths of length n from i to j in the directed graph with adjacency matrix A. We can use bijections of these edges and paths to replace symbols y_k of the construction of the previous paragraph with edges in the directed graph defined by A^n. Then that construction provides the claimed isomorphism.

1.8 Topological Markov shifts. An SFT is also called a topological Markov shift, or topological Markov chain. This terminology is appropriate because an SFT can be viewed as the topological support of a finite-state stochastic Markov process, and also as the topological analogue of such a process. (This viewpoint was advanced in the seminal 1964 paper of Parry [P1].)

Roughly speaking, in a Markov process the past and future are independent if conditioned on the present (or more generally if conditioned on some finite time interval). We can say precisely why an SFT is a topological analogue of this. Suppose the SFT is n-step (given by forbidding a certain list of words of length at most $n+1$). Also suppose that x and y are points (doubly infinite sequences) in

must also be a point in the SFT. That is: the possibilities for the future (sequences in positive coordinates) and the past (sequences in negative coordinates) are independent conditioned on the near-present (i.e., the word in a certain finite set of coordinates).

1.9 Applications of SFT's. For completeness I'll mention in the most cursory fashion two ways in which SFT's appear in a natural and useful way.

First, imagining very long tapes of zeros and ones, consider infinite strings of zeros and ones (i.e., points in the full shift on two symbols). It is natural to think of a block code as a machine which takes an input tape and recodes it, and to suppose that somehow the study of block codes may be relevant to efficiently encoding and decoding data. This turns out to be the case [ACH,MSW], in fact I understand that some constructions arising from symbolic dynamics have actually been built into IBM hardware.

Second, imagine a homeomorphism (or diffeomorphism) h on some space X. One way to study h is by symbolic dynamics. Crudely, cut X into n pieces. Name the pieces $1, 2, ..., n$. To any given point x in X there is associated a sequence y in the full shift on n symbols, where y is defined by setting y_i to be the piece containing $h^i(x)$, for each integer i. This gives some set of sequences y. The sequence associated to $h(x)$ is the shift of y. This establishes some relation between the topological dynamics of the shift space and the dynamics of h. Sometimes a relation of this sort is very useful (for example for analyzing h-invariant measures or periodic points), when the shift space which arises is a shift of finite type [Bow1,2].

A variation on the last theme going back to Hadamard is the study of geodesic flows with symbolic dynamics [AF].

SECTION 2. MATRIX INVARIANTS

Throughout this section A will represent a matrix with integral entries. Unless otherwise indicated, we also assume that A is nondegenerate (every row has a nonzero entry and every column has a nonzero entry) and that the entries are nonnegative. (This is because if A were nonnegative with ith row or column zero, then A would define the same SFT as would the principal submatrix obtained by deleting row i and column i — it is only the "nondegenerate core" of A which carries information about the SFT defined by A.) We let S_A denote the shift of finite type defined by A.

By a matrix invariant of A we will mean something determined by A which is the same for matrices A and B which determine isomorphic shifts of finite type. Some of these matrix invariants correspond to dynamically important invariants of the associated shift. The matrix invariants usually have an algebraic flavor, often being defined for (not necessarily nonnegative) integral matrices. Then one has an

2.1 Mixing properties. The nonnegative matrix A is primitive if some power has all entries greater than zero. A is irreducible if for every position (i, j) there is $n > 0$ such that $A^n(i, j) > 0$. Otherwise A is reducible.

The associated SFT S is mixing if and only if A is primitive. It has a dense forward orbit if and only if A is irreducible. The most important class to understand is the class of mixing SFT's. Then one understands other SFT's by how they are built up from the mixing SFT's. This is analogous to the situation with nonnegative matrices, which one understands by first understanding the primitive case. (Caveat: often the general case of a problem for SFT's follows very easily from the mixing case, but sometimes the generalization is quite difficult.) In the sequel we will sometimes make the simplifying assumption that A is primitive. Sometimes this is only for simplicity and sometimes we are avoiding serious difficulties.

2.2 Entropy. The premier numerical invariant of a dynamical system S is its (topological) entropy, $h(S)$. For a subshift S,

$$h(S) = \limsup_n \frac{\log(\#W_n(S))}{n}$$

where $W_n(S)$ is the set of words of length n occurring in sequences of S. That is, the entropy is the exponential growth rate of the S-words. For a full shift on n symbols, the entropy is $\log(n)$. For an SFT defined by a matrix A, the entropy is the log of the spectral radius of A. This is easy to prove because the number of words of length n is the sum of the entries of A^n.

What numbers can be entropies of mixing SFT's? Equivalently, what numbers can be spectral radii of primitive integral matrices? This was settled by Lind [L]: a number is the spectral radius of a primitive integral matrix if and only if it is a Perron number. A Perron number is a positive real number which is an algebraic integer which is strictly greater than the modulus of any other root of its minimal polynomial.

2.3 Periodic points. The periodic points of a topological dynamical system are often involved in its dynamics in a crucial way. Let $\mathrm{Fix}(S)$ denote the set of fixed points of a map S, i.e. the set of points x such that $Sx = x$. Suppose that for every positive integer n, the set $\mathrm{Fix}(S^n)$ is finite. (This will be true for any subshift S, for which a fixed point of S^n is a periodic sequence of period dividing n.) Then the sequence $\#\mathrm{Fix}(S^n)$ contains all the information one has from restricting S to its periodic points and forgetting the topology. The favored choice in dynamics for compiling this information is the (Artin-Mazur) zeta function of S,

$$\zeta_S(z) = \exp \sum_{n=1}^{\infty} \frac{\#Fix(S^n)}{n} z^n.$$

then the zeta function is defined as an analytic function on the open disc of radius $1/a$ around the origin in the complex plane. For subshifts, such a number a always exists, not larger than the cardinality of the alphabet. For an SFT defined by a matrix A, a is the spectral radius of A.

If S is an SFT defined by a matrix A, then the number of fixed points of S is simply the trace of A: a fixed point is a sequence consisting of one edge repeated forever. Similarly, for all positive integers n

$$\#Fix(S^n) = tr(A^n).$$

From this one can compute

$$\zeta_S(z) = \exp \sum_{n=1}^{\infty} \frac{tr(A^n)}{n} z^n = \left[\prod (1 - az) \right]^{-1} = [\det(I - zA)]^{-1}$$

where the product is over the eigenvalues a of A, repeated according to multiplicity. (So the inverse zeta function of an SFT is a polynomial with integral coefficients and constant term 1.) The first equality follows from the definition of the zeta function and the previous equation. The last two equalities hold for any square matrix A with real entries, as we now argue. The last equality follows from dividing the equation

$$\det(zI - A) = \prod_a (z - a)$$

by z^k (where A is k by k) and then replacing $1/z$ with z. The second equality follows from three facts:

(1) $tr(A^n) = \sum_a a^n$,

(2) for any complex number a,

$$\exp \left(\sum_{n=1}^{\infty} \frac{(az)^n}{n} \right) = 1/(1 - az)$$

(to see this take the derivative of the log of each side),

(3) $\exp(\sum_{n=1}^{\infty} \sum_a \frac{(az)^n}{n}) = \prod_a \exp(\sum_{n=1}^{\infty} \frac{(az)^n}{n})$.

Problem. What are the nonzero spectra of primitive integral matrices?

2.4 Isomorphism. Matrices A and B are elementary strong shift equivalent over a semiring S if there are matrices U, V with entries from S such that $A = UV$ and $B = VU$. The matrices U, V need not be square. If the semiring is not specified, then it is understood to be the nonnegative integers. A and B are strong shift equivalent over S if they are linked by a finite chain of elementary strong shift equivalences— that is, strong shift equivalence is the transitive closure of elementary strong shift equivalence.

Two shifts of finite type S_A and S_B are isomorphic if and only if the matrices A and B are strong shift equivalent [W1]. It is not trivial to prove isomorphism gives strong shift equivalence, so we refer to [W1] or [PT1,Sec.V.3] for this direction. However the other direction is easy. Suppose $A = UV$, $B = VU$. Let G_A, G_B be the directed graphs with adjacency matrices A, B and let these graphs have no vertices in common. Let U be the adjacency matrix for a set of arcs with initial vertices in G_A and terminal vertices in G_B; similarly V describes arcs from G_B to G_A. Let lower case letters a, b, u, v represent arcs corresponding to A, B, U, V. From the matrix equations we may choose bijections of arcs and paths of length 2,

$$\{a\} \longleftrightarrow \{uv\}, \qquad \{b\} \longleftrightarrow \{vu\}$$

respecting initial and terminal vertex. We view a point of S_A as an infinite path $...a_{-1}a_0a_1...$ of edges a_i and apply the first bijection at each a_i to get the following picture.

(Here, for example, u_0v_0 is the path corresponding to a_0.) We apply the second bijection to give a correspondence $v_iu_{i+1} \longleftrightarrow b_i$ and get a larger picture:

This picture gives a rule which sends a point of S_A to a point of S_B. It is easy to check that the rule defines an isomorphism of the two SFT's.

Strong shift equivalence and shift equivalence (below) were introduced by Williams [W1]. These are crucial ideas in the subject.

2.5 Eventual isomorphism. Two matrices A and B are shift equivalent over a semiring S if there are matrices U, V over S and a positive integer ℓ (called the lag) such that the following equations hold.

$$A^\ell = UV \qquad B^\ell = VU$$
$$AU = UB \qquad BV = VA.$$

Again, S is the nonnegative integers if the semiring is not specified.

Two systems S and T are eventually isomorphic if S^n and T^n are isomorphic for all but finitely many n. If a pair U, V give a shift equivalence of lag ℓ from A to B, then A^ℓ and B^ℓ are strong shift equivalent. Also the pair $A^n U, V$ gives a shift equivalence of lag $l + n$ from A to B. Consequently, if A and B are shift equivalent, then the SFT's S_A and S_B are eventually isomorphic. The converse also holds because the shift equivalence of A^p and B^p implies the shift equivalence of A and B if p is a sufficiently large prime [KR1].

Regardless of the semiring S, strong shift equivalent matrices must be shift equivalent. This follows from manipulating a chain of ℓ elementary strong shift equivalences to produce a lag ℓ shift equivalence.

At first glance, shift equivalence may appear to be a more obscure and complicated equivalence relation than strong shift equivalence. In fact, it is just the opposite. (We will say more about this below.) Williams introduced the idea of shift equivalence with the intent of reducing strong shift equivalence to a manage-

Problem. Suppose A and B are irreducible matrices shift equivalent over $\mathbb{Z}_+$. Must they be strong shift equivalent over $\mathbb{Z}_+$?

It is not hard to show that the answer is yes to this question if and only if it is yes to the question for primitive matrices.

The classification of reducible SFT's will not follow easily from the classification of irreducible SFT's. However, work in progress by Kim and Roush indicates that one will be able to classify shift equivalent reducible SFT's in terms of the classification of irreducible SFT's and the range of the dimension representation [KRW] on their automorphism groups.

2.6 Flow equivalence. Two homeomorphisms are flow equivalent if they are cross sections of the same flow. If the matrices A and B are irreducible and neither is a permutation matrix, then S_A and S_B are flow equivalent if and only if

(i) $\det(I - A) = \det(I - B)$, and

(ii) the cokernels of $I - A$ and $I - B$ are isomorphic.

(Here, for example, if A is $n \times n$ then the cokernel of $I - A$ is the abelian group which is $\mathbb{Z}^n$ modulo the image of $I - A$.) This was ultimately proved by Franks [F2] following the earlier work of Bowen&Franks [BF] and Parry&Sullivan [PS].

The classification up to flow equivalence of SFT's defined by reducible matrices is difficult and interesting. Cuntz [C] has introduced algebraic invariants in a special case from the viewpoint of associated C^*-algebras. Danrun Huang, beginning from the work of Franks and Cuntz, has gone much further on this problem [Hu]. Huang's work is very much in the vein of discerning the right invariants and then showing they are complete by way of matrix constructions realizing prescribed algebraic actions subject to positivity constraints.

We won't be concerned with flow equivalence in this paper, except to flesh out certain algebraic patterns below.

2.7 Relations.

For matrices over $\mathbb{Z}_+$:

strong shift equivalence $\Rightarrow$ shift equivalence $\Rightarrow$

same zeta function $\Rightarrow$ same entropy.

Also, for irreducible matrices at least,

shift equivalence $\Rightarrow$ same flow equivalence class.

None of these implications can be reversed.

3.1 Shift equivalence and the dimension group. Suppose A is a $n \times n$ integral matrix. Then A acts on $\mathbb{Z}^n$ and from this we can form the direct limit group $G(A)$, on which A induces an automorphism A'. This gives a pair $(G(A), A')$. Two integral matrices A and B are shift equivalent over $\mathbb{Z}$ if and only if there is a group isomorphism f from $G(A)$ to $G(B)$ such that $f A' = B' f$.

This was pointed out by Krieger [Kr1,2] who also did something much deeper, giving a K-style construction of this group from certain compact subsets of the shift space. We will skip the (easy) proof of the previous paragraph and any explanation of the topological theory (see [BMT] for more).

We remark that $G(A)$ is called a dimension group (and therefore $(G(A), A')$ has been called the "dimension pair") for historical reasons. Krieger's topological construction was adapted from K-theoretic constructions in operator algebras–in fact, $G(A)$ is K_0 of an associated C^*-algebra [CuK2]. It is natural to think of K_0 of a ring R as a "dimension group", because K_0 is concerned with (stable) isomorphism classes of finitely generated projective R-modules and for R a field the isomorphism class of such a module is given by its (vector space) dimension.

We will give a concrete description of $(G(A), A')$. Of course A acts on $\mathbb{Q}^n$. Let V_A be the rational subspace of $\mathbb{Q}^n$ which is the image of A^n. (So, if A is not nilpotent, then V_A is the largest invariant subspace on which the action of A is nonsingular.) Now we can give the following presentation for $G(A)$:

$$G(A) = \{v \text{ in } V : \text{for some } k > 0, v A^k \text{ is in } \mathbb{Z}^n\}.$$

With this presentation, the automorphism A' is just multiplication by A. (We have arbitrarily chosen the action here to be on row vectors. The choice does matter, as a matrix need not be shift equivalent to its transpose [PT1]. However matrices A and B are (strong) shift equivalent if and only if their transposes are.)

For example, if $A = [2]$, then in this presentation $G(A)$ is the dyadic rationals–all rational numbers p/q where p and q are integers and q is a power of 2. If $|\det A| = 1$, then $G(A)$ is just $\mathbb{Z}^n$.

Note: if $(G(A), A')$ and $(G(B), B')$ are isomorphic, then the actions obtained by tensoring with $\mathbb{Q}$ are isomorphic. In other words, the restrictions of A and B to V_A and V_B are isomorphic as linear transformations of rational vector spaces. (In particular these restrictions have the same characteristic polynomial, which is just the characteristic polynomial of A divided by the appropriate power of the indeterminate. Equivalently, $\det(I - tA) = \det(I - tB)$.) Another way to say this is that A and B have the same Jordan form away from zero (i.e., the nonnilpotent parts of the Jordan forms of A and B are the same modulo conjugation by a permutation matrix).

To see this easily, suppose there is a shift equivalence. Note that because $|det A| = 1$, we must have $|det U| = 1$ (where U is the matrix in the defining equations for shift equivalence). If $AU = UB$, then

$$U^{-1}(A - I)U = (B - I).$$

Since every matrix on the left side is integral and $A - I$ is divisible by 2, the matrix on the right side must have every entry divisible by 2, a contradiction. (For an alternate proof, note that $\text{cok}(I - A) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$ and $\text{cok}(I - B) \cong \mathbb{Z}/4$, so that A and B do not even define flow equivalent SFT's.)

The general solution to the decision problem for shift equivalence is very difficult [KR1,3]. But there are many classes of tractable examples. For example, if A has a single nonzero eigenvalue k, then A is shift equivalent over $\mathbb{Z}$ to the one by one matrix $[k]$. If $p(t)$ is the minimal polynomial of an algebraic integer λ, then the shift equivalence classes over $\mathbb{Z}$ of integral matrices with characteristic polynomial $p(t)$ are in bijective correspondence with the ideal classes of the ring $\mathbb{Z}[1/\lambda]$ ([BMT]). Such algebraic aspects of shift equivalence turn out to correspond to coding relations among corresponding shifts of finite type ([BMT], [KMT], [As2]).

3.2 (Strong) Shift equivalence over $\mathbb{Z}_+$. If A and B are shift equivalent over $\mathbb{Z}$, then they are strong shift equivalent over $\mathbb{Z}$. The same is true for matrices over any principal ideal domain [E2],[W2] or Dedekind domain [BH2].

It is not known whether primitive matrices shift equivalent over $\mathbb{Z}_+$ must be strong shift equivalent over $\mathbb{Z}_+$. This is still unknown if $\mathbb{Z}$ is replaced by the rationals $\mathbb{Q}$, or even the reals $\mathbb{R}$! So the order requirement complicates the situation for strong shift equivalence drastically.

For shift equivalence, there are also complications but they are fewer. The best news is that two primitive matrices are shift equivalent over $\mathbb{Z}_+$ if they are shift equivalent over $\mathbb{Z}$. (This is still true by the same proof if $\mathbb{Z}$ is replaced by any unital subring of the reals.) This was first proved geometrically by Parry and Williams [PW], also see [KR1]. (Caveat: irreducible matrices may be shift equivalent over $\mathbb{Z}$ but not over $\mathbb{Z}_+$, as an example of Kaplansky and myself shows [B2].)

It is easy to sketch a proof of this. Suppose A and B are primitive and a pair U, V gives a shift equivalence of lag ℓ ($AU = UB$ etc.). The idea is, for large n the matrices $A^n U$, $V A^n$ are positive (possibly after replacing U, V with $-U, -V$) and they give a shift equivalence of lag $2n + \ell$. To understand positivity, remember that the Perron Theorem implies that for large n, A^n is "approximately" $(a^n)RL$ where R is a positive right eigenvector, L is a positive left eigenvector, $LR = 1$, and a is the spectral radius of A. Here "approximately" means that the error is growing at an exponentially smaller rate. Thus it suffices to show that LU and VR

One puts the requirement of nonnegativity into the dimension group context as follows [Kr1]. An ordered group is a group G with a set G_+ (called the positive set) such that G_+ is closed under addition and every element of G is a difference of elements of G_+. An isomorphism of ordered groups is a group isomorphism mapping the positive set of the domain group onto the positive set of the range group. Now one just adds to the "dimension group" structure an order structure on the group to reflect shift equivalence over $\mathbb{Z}_+$ rather than $\mathbb{Z}$. With this structure, the isomorphism A' above is an isomorphism of the ordered group $(G(A), G_+(A))$. The "dimension pair" now becomes a "dimension triple" (G, G_+, A').

We'll describe this in terms of the explicit presentation described in §3.1. Given a nonnegative matrix A, define the positive set

$$G_+(A) = \{v \text{ in } G(A) : \text{for some } k > 0, vA^k \text{ has all entries nonnegative}\}.$$

Multiplication by A induces an ordered-group automorphism A' on $(G(A), G_+(A))$. Now matrices A, B over $\mathbb{Z}_+$ are shift equivalent if and only if there is an isomorphism of their ordered groups intertwining A' and B'.

For more on dimension groups, see [E1].

3.3 The dimension module of an SFT. Here we make explicit a reformulation of the dimension data, which will seem trivial (but correct) from a homological viewpoint [Br]. This reformulation is by no means due to me (see [Wa1, pp.92,120]).

To an SFT defined by a matrix A, we associated a "dimension pair" $(G(A), A')$. The action of A' gives an action of the group $\mathbb{Z}$ on $G(A)$. Whenever a group H acts on an abelian group K by automorphisms of K, the group K acquires a $\mathbb{Z}H$-module structure, where $\mathbb{Z}H$ is the integral group ring of H [Br]. An isomorphism of such H-actions is equivalent to a $\mathbb{Z}H$-module isomorphism. So instead of referring to the "dimension pair" $(G(A), A')$ we can just refer to the dimension module. Here we mean the $\mathbb{Z}\mathbb{Z}$-module $G(A)$. The ring $\mathbb{Z}\mathbb{Z}$ is isomorphic to the ring $\mathbb{Z}[t, t^{-1}]$ of integral Laurent polynomials in one variable.

Since "dimension pair" and "dimension module" carry the same information, to some extent passing from the latter to the former is just a matter of cleaner terminology. But it is also a matter of a better functorial setup— "thinking right" as the group cohomologists say. We'll see more of this in Sections 5 and 7.

We encode the order information of the "dimension triple" by making the $\mathbb{Z}\mathbb{Z}$-module an ordered module in the natural way. First, $\mathbb{Z}\mathbb{Z}$ is an ordered ring in a natural way, with the semiring $\mathbb{Z}_+\mathbb{Z}$ (formal nonnegative integral combinations of the set $\mathbb{Z}$) the positive set. (If we think of $\mathbb{Z}\mathbb{Z}$ as the Laurent polynomials $\mathbb{Z}[t, t^{-1}]$, then the positive set is $\mathbb{Z}_+[t, t^{-1}]$, the Laurent polynomials whose coefficients are

3.4 Strong shift equivalence and state-splitting. Let A be an $n \times n$ matrix. Let A' be an $(n+1) \times n$ matrix related to A as follows: row i of A is the sum of row i and row $(n+1)$ of A', otherwise the rows of A and A' are equal. Now define an $(n+1) \times (n+1)$ matrix B related to A' as follows: column $(n+1)$ of B equals column i of A', and the first n columns of A' equal those of B. Then there is an elementary strong shift equivalence (U, V) between A and B with $A' = V$. For example,

$$A = \begin{pmatrix} 0 & 9 \\ 8 & 5 \end{pmatrix} \quad A' = \begin{pmatrix} 0 & 9 \\ 6 & 1 \\ 2 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 9 & 9 \\ 6 & 1 & 1 \\ 2 & 4 & 4 \end{pmatrix} \quad U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

In this case, or in the case where the roles of row and column are reversed, we say that B is obtained from A by an elementary state-splitting (the state i is split into two new states) and A is obtained from B by an elementary amalgamation. One of the fundamental tools introduced in Williams' paper [W1] is the following: if two SFT's S_A and S_B are topologically conjugate, then there is a finite sequence of state splittings and amalgamations which begins with A and ends with B. (In fact this can be chosen to be a finite sequence of row splittings followed by a finite sequence of column amalgamations [P2].)

3.5 The Masking Lemma. As one application of Williams' theorem that strong shift equivalence of nonnegative integral matrices is equivalent to isomorphism of the SFT's they represent, we state a case of Nasu's Masking Lemma. (This is more or less his original statement [N], but his argument gives a much more general result [BH1, App.1].)

THEOREM (NASU). *Suppose A and B are square nonnegative integral matrices, and there is a subsystem of S_A which is conjugate to S_B. Then there is a matrix A' which defines an SFT conjugate to S_A, such that A' contains B as a principal submatrix.*

I have no idea how one would prove this from scratch. With Williams' theorem, the basic idea is very simple. First one passes from A to a higher block presentation, which has as a principal submatrix a matrix B' defining an SFT conjugate to S_B. Williams' result gives a strong shift equivalence from B to B'. Now one simply experiments with 2×2 block forms and sees that each elementary strong shift equivalence along this chain can be extended. For details, see [N] or [BH1].

3.6 Algebraic topology. Wagoner [Wa1-5] has introduced ideas of algebraic topology into the study of shift equivalence, strong shift equivalence and other matters beyond the scope of this survey. Essentially, these let one make arguments and constructions by way of topological objects constructed as quotient spaces of certain infinite simplices from the relations of (strong) shift equivalence. (Incidentally in

SECTION 4: ZETA FUNCTIONS AND SPECTRA

Recall the zeta function of §2.2,

$$\zeta_S(z) = \exp\left(\sum_{n=1}^{\infty} \frac{\#\mathrm{Fix}(S^n)}{n} z^n\right).$$

At first glance this may seem an unnecessarily tortured way to encode the information of the periodic points. We'll consider some justification for this, particularly from the viewpoint of matrices and shifts of finite type.

4.1 "Higher mathematics". The zeta function comes to us from more exalted zeta functions in algebraic geometry and number theory. The zeta function was introduced by Artin and Mazur [AM]. They used its natural relationship to certain algebraic geometric systems to obtain constraints on the periodic points of large sets of diffeomorphisms by proving rationality of the zeta function. The potential connections with such systems, and other algebraically defined systems, are one reason to use the zeta function to count periodic points. Algebraic topology also appears as a powerful tool for analyzing zeta functions even of systems which do not arise from algebra [F2,Fri1].

4.2 Rationality constraints. Given a system S, consider the sequence $f_n = \#\mathrm{Fix}(S^n)$, $n = 1, 2, \dots$. Assume each f_n is finite. We can capture the information in this sequence in a generating function or a zeta function,

$$g(z) = \sum_{n=1}^{\infty} f_n z^n$$

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{f_n}{n} z^n\right).$$

These functions do carry the same information. Also, if the zeta function is a ratio of polynomials, then so is the generating function (it is the derivative of the log of the zeta function, multiplied by z). But the converse is false.

The generating function is rational if and only if the sequence f_n eventually satisfies some recursion relation. It turns out [BowL] that the zeta function is rational if and only if there are integral matrices C, D such that for all n,

$$f_n = tr C^n - tr D^n.$$

It is sometimes the case in dynamics that one can prove the rationality of the zeta function for interesting systems, precisely by finding such matrices [F1,Fri1,Fri2]. The rationality of the zeta function then sharply and transparently captures this constraint.

4.3 Product formula. The zeta function can be written as a (usually infinite) product

4.4 Why not the characteristic polynomial? If A is a matrix then its characteristic polynomial $\chi_A(z) = \det(zI - A)$ is almost the zeta function of $S = S_A$; if A is $n \times n$, then

$$\zeta_S^{-1}(z) = \det(I - zA) = z^n \det(z^{-1}I - A) = z^n \chi_A(z^{-1}).$$

However, the characteristic polynomial contains extraneous, noninvariant information; only the nonzero spectrum matters for the traces of powers of A. Also, sometimes working with $\det(I - zA)$ rather than $\det(zI - A)$, one avoids extraneous but nontrivial complications of sign (e.g., §4.5). Still, sometimes for practicality one works not with the zeta function but with the characteristic polynomial of A away from zero—this is the unique polynomial with nonzero constant term which can be obtained from the characteristic polynomial by dividing by a power of the indeterminate.

4.5 Cycles. Let A (for simplicity) be a nonnegative integer matrix. One can check that $\det(I - zA)$ is a sum of terms $(-z)^k$, where there is one term for each set of pairwise disjoint simple cycles in the directed graph with adjacency matrix A, and k is the sum of the lengths of the cycles in the set. Here a cycle is simple if it visits no vertex twice, and two cycles are disjoint if they have no vertex in common. (I learned this viewpoint from [W3] and [Ara]; these authors consider matrices with more general entries, corresponding to graphs labelled by polynomials—sometimes commuting, sometimes not—where the viewpoint is especially useful and the terms acquire coefficients which are products of the labels along the cycles.) So there is some nice connection between the structure of simple cycles and the zeta function.

4.6 Matrices with polynomial entries. We discuss this in the next section.

SECTION 5. GRAPHS VIA POLYNOMIAL MATRICES

5.1 Introduction. Recall, a $n \times n$ matrix over $\mathbb{Z}_+$ can be considered as the adjacency matrix of a directed graph with n vertices. Using such matrices to represent SFT's (as edge shifts) allows a more concise presentation of SFT's than one has using only zero-one matrices (for vertex shifts), and gives access to additional arguments.

There is a still more general way to present a directed graph (hence an SFT), by using matrices with entries in $t\mathbb{Z}_+[t]$ (polynomials in one variable t, with nonnegative integer coefficients, with every term divisible by t—i.e., the only constant term allowed is zero). This allows still more concise presentations, additional access to matrix arguments and algebra, and a concordance of formal patterns which make it a convincing candidate for the "right" general way to present a directed graph (or SFT).

preparation by Kim, Roush and Wagoner, which we will not preempt with further discussion here, develops very interesting and useful new constructions of conjugacies by methods which appeal in a fundamental way to this polynomial matrix setting.

I thank Hang Kim and Fred Roush for suggesting to me that these polynomial matrices may also be important for studying the inverse spectral problem and related problems for nonnegative matrices.

5.2 The construction. The basic idea is extremely simple. Let A be an $n \times n$ matrix over $t\mathbb{Z}_+[t]$. From A we will construct a directed graph. Its vertex set will include a set of n vertices (say $1, 2, ..., n$) which index the rows and the columns of $A(t)$. If for example, $A(1,2) = 3t^2 + t^4$, then there will be three paths of length 2 and one path of length 4 from vertex 1 to vertex 2. At each interior vertex on one of these paths (a path of length k has $k-1$ interior vertices), there is just one incoming edge and one outgoing edge. These interior vertices are disjoint from $1, 2, ..., n$. This recipe produces a graph which can have many more than n vertices (hence the conciseness of the presentation). For example, the matrix

$$A(t) = \begin{bmatrix} 0, & t^2 \\ 2t^3, & t + t^3 \end{bmatrix}$$

produces the directed graph

Note as in [BGMY] (fondly referred to as the "bigamy" paper) that the distinguished set of n vertices (corresponding to the indices of the rows and the columns of A) is a "rome": any sufficiently long path in the graph hits the rome. Given a rome in a directed graph, one can reverse the procedure and produce a presenting matrix A over $t\mathbb{Z}_+[t]$, where A is $n \times n$ if the rome has cardinality n. If B is a matrix over $\mathbb{Z}_+$ which is the adjacency matrix of a directed graph, then the matrix tB is one matrix over $t\mathbb{Z}_+[t]$ which presents the graph in the new formalism.

Another viewpoint is to think of A as giving a directed graph G' with labelled edges. The number of edges from i to j is the (i,j) entry of A evaluated at $t = 1$. An edge is labelled by a power of t. The power corresponds to the length of a path.

(Of course by inverting we get the zeta function of the associated SFT.) We can picture the argument in terms of the graph G' described in §5.2. We choose, if possible, some arc from i to j labelled by t^{k+1}, with $k > 0$; then we delete this arc, add a vertex i', add an arc labelled t from i to i', and add an arc labelled t^k from i' to j. It is clear that a finite sequence of such moves produces the graph G, with every edge labelled by t. So we are done if we show the invariance of $\det(I - A)$ under one such elementary move.

This is a simple computation. With i, i', j and A as above, we may assume $i' = 1, i = 2$. Let A' be the matrix derived from A by the elementary move above. Adding t times row 1 of $(I - A')$ to row 2 does not change the determinant. The resulting matrix M has determinant equal to $\det(I - A)$, because the upper left entry is 1, every other entry in the first column is zero, and the lower right diagonal block is $I - A$.

Below, by the zeta function of a matrix (over $\mathbb{Z}$ or $t\mathbb{Z}_+[t]$) we will mean the inverse of the quantity (*) above. By its spectral radius we will mean the spectral radius of B in (*)—i.e., $1/a$, where a is the smallest positive root of (*).

5.4 Shift equivalence and flow equivalence.

In this part let B be the adjacency matrix for a graph which is also presented by a matrix A over $t\mathbb{Z}_+[t]$. Let B have size N. Let $L = \mathbb{Z}[t, t^{-1}]$ be the ring of Laurent polynomials over $\mathbb{Z}$. Let L^N represent the N-fold direct sum.

The matrix $I - tB$ maps L^N into itself. It is well known [Wa1, pp.92,120] that the group cokernel$(I - tB)$ is isomorphic to the dimension group $G(B)$ of §3.1. It is not hard to check that an isomorphism is determined by the map which sends an element $[t^n e_i]$ of cok$(I - tB)$ (where e_i represents the usual canonical basis vector) to the vector $e_i(B')^{-n}$ (where B' is the isomorphism of $G(B)$ in §3.1). Moreover, if we let cok$(I - tB)$ be an L-module in the obvious way, then under this correspondence the action of t^{-1} on cok$(I - tB)$ corresponds to the action of B' on $G(B)$. So, the L-module cok$(I - tB)$ is a version of the dimension module of §3.3. (Similarly, we could use the module cok$(I - t^{-1}B)$; then the action of t would correspond to the action of B.) For integral matrices C and D, there is a module isomorphism between cok$(I - tC)$ and cok$(I - tD)$ if and only if C and D are shift equivalent over $\mathbb{Z}$.

To summarize: the L-module cok$(I - tB)$ is the shift-equivalence-over-$\mathbb{Z}$ class. Also $\det(I - tB)$ is the (inverse of the) zeta function.

Obviously we would like these facts to hold for A in place of tB. They do! (This was observed independently by, at least, Wagoner and myself.) One proof that cok$(I - tB)$ and cok$(I - A)$ give isomorphic modules follows the pattern of the proof of §5.3 for $\det(I - A)$. Again check each elementary step. Note that the

will sketch how shift equivalence and flow equivalence are nicely unified in this polynomial setting. Recall, $\det(I - B)$ and $\operatorname{cok}(I - B)$ are complete invariants of flow equivalence for S_B within the class of irreducible SFT's. Obviously we can get these from $\det(I - tB)$ and $\operatorname{cok}(I - tB)$ by setting $t = 1$. Can we recover these invariants from $\det(I - A)$ and $\operatorname{cok}(I - A)$ by setting $t = 1$? Yes.

A satisfying way to express this is the following. Given A, form the L-module $\operatorname{cok}(I - A)$. This is the shift equivalence class and one invariant of it is $\det(I - A)$. Now, in the language of group cohomology [Br], apply the coinvariants functor to this $\mathbb{Z}[t, t^{-1}]$ module to get a $\mathbb{Z}$-module. This $\mathbb{Z}$-module (abelian group) $\operatorname{cok}(I - A(1))$ is the Bowen–Franks group. This group and $\det(I - A(1))$ are complete invariants of flow equivalence. The latter invariant is obtained by applying the coinvariants functor to $\det(I - A)$.

In other words, at least in this primitive case we get the flow equivalence class by applying the coinvariants functor to the shift equivalence class. It will be interesting to see how well this viewpoint carries over to more general situations.

5.5 Powers. Let B be a nondegenerate adjacency matrix for a graph which is also presented by a matrix A over $t\mathbb{Z}_+[t]$. Also, let n be a positive integer greater than 1. Recall (§1.7) that the matrix B^n over $\mathbb{Z}_+$ presents an SFT isomorphic to $(S_B)^n$.

In contrast, the SFT T_n presented by the polynomial matrix A^n will never be isomorphic to $(S_B)^n$. It turns out that the SFT T_1 is a quotient of T_n by a map which is everywhere n-to-1! (In particular, A and A^n present SFT's of equal entropy.) To see this, let A have size k, and let $1, 2, ..., k$ represent the vertices comprising the natural rome in the graph G_n presented by A^n. Similarly, let $1, 2, ..., k$ represent the vertices comprising the natural rome in the graph G_1 presented by A. An entry $A^n(i, j)$ represents the paths in G_n from vertex i which end at vertex j and whose interior vertices do not intersect the rome in G_n. This entry also represents the paths from i to j in G_1 whose interior vertices hit the rome in G_1 exactly $n - 1$ times. This correspondence gives us a map from paths in G_n to paths in G_1. It is easy to check this map determines a code from T_n to T_1 which is everywhere n-to-1.

The inverse zeta function of T_n is $\det(I - A^n)$; this polynomial is divisible by the polynomial $\det(I - A)$, which is the inverse zeta function of S_B. The ratio is $\det(I + A + ... + A^{n-1})$. It is possible for the latter polynomial to be trivial (so $\det(I - A) = \det(I - A^n)$). An example for $n = 2$ (i.e., $\det(I + A) = 1$) is given by

$$A = \begin{bmatrix} t^2, & t \end{bmatrix}$$

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

Finally, we indicate briefly how the shift equivalence data for $(S_B)^n$ may be recovered from the polynomial matrix A. In this polynomial setting, as in §5.4 we think of the shift equivalence class of $(S_B)^n$ as the isomorphism class of a certain ordered L-module. We derive a representative M of this class from the ordered L-module $\operatorname{cok}(I - A)$ as follows. As an ordered abelian group, we let M equal $\operatorname{cok}(I - A)$. We define the action of L on M by defining the action of t on M to be the action of t^n on the original module $\operatorname{cok}(I - A)$.

5.6 Small presentations. If a matrix is $n \times n$, then we say it has size n. What is the smallest size of a matrix with a given nonzero spectrum? With a given shift equivalence class? These are difficult questions with unhappy answers.

For example, consider the 4-tuple $(\sqrt{2}, i, -i, \epsilon)$, where ϵ is small and positive. This will be the nonzero spectrum of a primitive real matrix, but as ϵ goes to zero the minimum size of such a matrix goes to infinity [BH1, Ap.3]. The problem of determining the minimum size at which a primitive matrix can realize a given nonzero spectrum is notoriously difficult [BH1, Ap.3].

For an example over the integers, consider the 3-tuple $(5,1,1)$. There is an infinite collection of primitive integral matrices which have nonzero spectrum $(5,1,1)$ but which are pairwise not shift equivalent. (This follows from the general results of [BH2]. It is also given by a barehanded construction in [B2,Example 3.4], which works for tuples (M,a,a) when M and a are positive integers with M > a+3 .) In particular, as one runs through the possible shift equivalence classes, the sizes of the realizing primitive matrices must go to infinity (since the entries of primitive integral matrices with bounded spectral radius and size are uniformly bounded).

It is a plausible and exciting prospect that one can obtain much more control on the size of a presenting matrix over $t\mathbb{Z}_+[t]$. As a quick example, we remark that it is easy to convert the above-mentioned primitive matrices (those constructed in [B2,Example 3.4]) into polynomial matrix presentations of size 4. That is, we get infinitely many shift equivalence classes with the same zeta function, all presented by polynomial matrices of size 4. (Of course, the degrees of the polynomials in these matrices must become unbounded.)

The striking theorems below of Handelman and Perrin (§5.8 and §5.9), and the

Problem 1. Suppose C is a primitive integral matrix and A is a matrix over $t\mathbb{Z}_+[t]$, where C is strong shift equivalent over $\mathbb{Z}_+$ to the adjacency matrix for the graph presented by A. Given C, what is the minimum size possible for A?

Problems 2 & 3. Suppose an integral matrix C of size n is shift equivalent over $\mathbb{Z}$ to a primitive matrix. Find good bounds for the minimum size of a matrix A over $t\mathbb{Z}_+[t]$ such that A presents a graph with primitive adjacency matrix B, where

(Problem 2) B is shift equivalent over $\mathbb{Z}$ to C

(Problem 3) B and C have the same nonzero spectrum.

We remark that for problems 1 and 2, the minimum number of generators for the group $\mathrm{cok}(I - C)$ is a lower bound for the size of A. There are additional and independent constraints involving the sequence trC^n and the spectral radius of C. For example, if C has spectral radius less than 2 and $trC > k$, then the size of A is greater than k.

5.7 Sparse matrices and labelled graphs.

Suppose for the moment, for concreteness and simplicity, that U is a unital subsemiring of the reals (e.g., $\mathbb{Q}$ or $\mathbb{R}_+$). Suppose C is a square matrix whose entries lie in $tU[t]$–i.e., they are polynomials over U with zero constant term. Then we can think of C as presenting a labelled graph much as above. If C has size k, then there are k distinguished vertices in the graph. A monomial term ct^n in $C(i,j)$ contributes a path of n arcs from vertex i to vertex j, whose interior vertices connect to no additional vertices. The first arc on the path is labelled c and the others are labelled 1.

Let B be the matrix which is the adjacency matrix of this graph. So, B is $N \times N$, where N is the total number of vertices in the graph. $B(i,j)$ is zero if there is no arc from i to j, otherwise it is the label on the arc from i to j. Now B might be quite sparse, and so we can think of C as a concise presentation of B. Also, it is an easy exercise following §5.3 to check that $\det(I - C) = \det(I - tB)$. (Similarly, one can prove an analogue of §5.4.)

The point we wish to make is that matrices over polynomial rings may be a good way to present sparse matrices. In particular, one can hope for more satisfying results on the intractable problem of determining the smallest size nonnegative matrix with a given nonzero spectrum, by allowing polynomial matrix presentations.

Problems 2$\mathbb{R}$ & 3$\mathbb{R}$. Suppose a real matrix C of size n is shift equivalent over $\mathbb{R}$ to a primitive matrix. Find good bounds for the minimum size of a matrix A over $t\mathbb{R}_+[t]$ such that A presents a graph with primitive adjacency matrix B, where

(Problem 2$\mathbb{R}$) B is shift equivalent over $\mathbb{R}$ to C

(Problem 3$\mathbb{R}$) B and C have the same nonzero spectrum.

Of course, in problems 2$\mathbb{R}$ and 3$\mathbb{R}$ above, in place of $\mathbb{R}$ we could consider any

real number. The matrix A he constructs presents a graph with primitive adjacency matrix.

5.9 Perrin's theorem. One can ask, given a Perron number λ, what is the smallest matrix A over $t\mathbb{Z}_+[t]$ with spectral radius λ? Dominique Perrin has explained to me that for any Perron number λ there is a matrix A over $t\mathbb{Z}_+[t]$ which has size 2 and has spectral radius λ! This follows from his construction in [Pe]. Given λ, the nonnegative integral matrix B on page 364 of [Pe] has spectral radius λ^n, for some positive n. The desired matrix A is obtained by noticing that the labelled graph for $t^n B$ has a rome consisting of the two vertices 1 and k.

To be honest, there is a nontrivial imperfection to Perrin's striking result: the adjacency matrix for the corresponding graph will be irreducible but not necessarily primitive, and the period of the matrix (which will be the integer n above) may be large.

Problem. Can one prove Perrin's result, but with the realizing matrix A corresponding to a graph of period 1?

5.10 A theorem on extensions.

Lind [L] proved that every Perron number (positive algebraic integer with modulus strictly greater than that of any conjugate) is the spectral radius of a primitive integer matrix (the converse is obvious from the Perron theorem). He did not bound the size of such a matrix. Using the polynomial matrix presentation, we'll produce a realizing matrix whose size is the degree of the Perron number. Basically the proof just lifts a corollary [BMT 5.14] of work of Lind and Handelman into the polynomial matrix setting.

As in [BMT], we say a matrix is IEP if it is square, every entry is an integer, and it is eventually positive (i.e., all sufficiently large positive powers of the matrix have every entry strictly positive).

LEMMA. *Let B be an IEP matrix of size m. Then there is a primitive matrix C of size m over $t\mathbb{Z}_+[t]$ with the same spectral radius as B such that the dimension module for B is a quotient of that for C.*

Proof. Following the lines of Lind's proof [L] (but avoiding most of the difficulties via the IEP hypothesis), one can construct a primitive matrix A with spectral radius equal to that of B such that the dimension module for B is a quotient of that for A. This was done in [BMT, 5.14]. The matrix A constructed there is the adjacency matrix of a $(\mathbb{Z}_+)$-labelled graph in which m vertices (denoted there as $v(i,n), 1 \leq i \leq \beta$; β there corresponds to m here) comprise an obvious rome. Using this rome we pass to the size m matrix presentation over $t\mathbb{Z}_+[t]$.

$\square$

(2) *Suppose C is an $n \times n$ integral matrix whose spectral radius λ is a simple root of the characteristic polynomial which is strictly greater than the modulus of any other root. Then there is a directed graph with primitive adjacency matrix B (over $\mathbb{Z}_+$) which can be presented by a matrix A over $t\mathbb{Z}_+[t]$, such that the following hold*

> *- B has spectral radius λ*
>
> *- the dimension module for C is a quotient of the module for B*
>
> *- if λ is irrational, then A has size n*
>
> *- if λ is rational, then A has size at most $n + k$,*
>
>   *where k is the smallest integer such that $\lambda^k \geq n + k$.*

Proof. First we prove (2) using results of Handelman. If λ is irrational, then C is similar over the integers to an IEP matrix [H1]. If λ is rational, then C is shift equivalent over the integers to an IEP matrix of size at most $n + k$ (k as defined in (2)) [H2]. In either case, the lemma provides the desired matrix A.

Now we prove (1). If λ is rational, then λ is a positive integer, and we let $A = [t\lambda]$. If λ is irrational, then we let C be the companion matrix of the minimal polynomial of λ and appeal to (2). This finishes the proof.

□

SECTION 6. MORE WILLIAMS' PROBLEMS

6.1 Introduction. The equations defining (strong) shift equivalence can be used to define (strong) shift equivalence for morphisms in any category. Always, strong shift equivalence implies shift equivalence. By a Williams' problem we mean the problem of whether the converse is true.

It turns out that shift equivalence and strong shift equivalence arise in several natural ways from problems in symbolic dynamics. We'll consider them in this section. In each case we get a Williams' problem. Often there is a dynamical interpretation for strong shift equivalence and shift equivalence (isomorphism and eventual isomorphism) which gives the problem direct dynamical meaning. Usually shift equivalence turns out not to imply strong shift equivalence (but understanding the difference is fundamental).

First we recall a definition. If G is a semigroup, then the integral semigroup ring $\mathbb{Z}G$ of G is the free abelian group with generator set G, with the multiplication defined on G by the semigroup operation and then extended to $\mathbb{Z}G$ by the distributive law. We picture an element of $\mathbb{Z}G$ as a formal integral combination of elements of G. We let $\mathbb{Z}_+G$ denote nonnegative integral sums of elements of G. We make $\mathbb{Z}G$

6.2 Markov shifts and matrices of Laurent polynomials. By a Markov shift we will mean an irreducible shift of finite type (SFT) together with a shift-invariant Markov measure with full support. It turns out that the analysis of Markov shifts is intimately related to understanding matrices of integral Laurent polynomials, and in particular their shift equivalence classes.

A Markov shift can be defined by an irreducible stochastic matrix P. Let A be the zero-one matrix such that $A(i, j)$ is zero if and only if $P(i, j)$ is. Then S_A is the underlying SFT. We view P as giving labels to the edges in the graph with adjacency matrix A—that is, P is a function from arcs into the reals. P determines the Markov measure on the SFT S_A as follows. Let ℓ be the positive left eigenvector for P whose entries sum to 1. Then for any i, the measure of the set of points which see a given word $a_0...a_k$ in coordinates $i, ..., i + k$ is $\ell(v_0)P(a_0)P(a_1)...P(a_k)$, where v_0 is the initial vertex of the arc a_0.

A code between Markov shifts $S(P)$ and $S(Q)$ is a code between their underlying SFT's which sends the P-measure to the Q-measure. (For more on these codes and their relatives, we recommend [P2] and [MT1].)

It is natural to try to generalize the ideas of (strong) shift equivalence to this category by using the (strong) shift equivalence equations on stochastic matrices. To see why this fails, suppose we have $P = UV$ and $Q = VU$ with P and Q stochastic. We would like to build up some elementary isomorphism between $S(P)$ and $S(Q)$ with these equations. We could try to follow the construction of §2.3. But now we don't know how to break up UV-paths into arcs. An entry of UV can be interpreted as a sum of terms (weights) on paths. Such a weight has the form $U(i, j)V(j, k)$. But $UV(i, k)$ could be the sum of, say, several such small terms or just a few larger terms. It turns out that what one really needs to know are the path weights with multiplicities—information which is lost on multiplying the real matrices.

The solution [PT2] is to regard the entries of the stochastic matrix P as lying not in the reals $\mathbb{R}$ but in a larger ring, $\mathbb{Z}\mathbb{R}_+^*$, the integral group ring of the group $\mathbb{R}_+^*$ of positive reals under multiplication. (Parry and Tuncel [PT2] actually used an isomorphic ring with a more analytic flavor.) It turns out [PT2] that in a natural way, P and Q are strong shift equivalent over $\mathbb{Z}_+\mathbb{R}_+^*$ if and only they define isomorphic Markov shifts. Also, P and Q are shift equivalent as matrices over $\mathbb{Z}_+\mathbb{R}_+^*$ if and only if all but finitely many powers of the Markov shifts are isomorphic (i.e., they are eventually isomorphic) [MT1].

Finally there is a crucial simplification. Given P, Parry and Schmidt [PSc] showed it is possible to pass in a canonical way from $\mathbb{R}_+^*$ to a finitely generated subgroup of $\mathbb{R}_+^*$—that is, to consider only matrices over a certain finitely generated subgroup of $\mathbb{R}_+^*$. This group must be isomorphic to $\mathbb{Z}^n$ for some n. After choos-

of these things, we recommend [MT1].)

Much of the structure for SFT's generalizes here. For example, the information carried by the periodic points is perfectly encoded in the stochastic zeta function of Parry and Tuncel [PT2] , given by the formula

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{tr(P^n)}{n} z^n = [\det(I - zP)]^{-1}.$$

Here the entries of P lie not in $\mathbb{R}$ but in its integral group ring (or, if one prefers, in the isomorphic ring $\mathbb{Z}[\exp]$ of Parry and Tuncel, see [PT2] or [MT1, Defn. 4.2]). The equation above makes sense at the level of formal power series. (In earlier work of Parry and Williams [PW], a "stochastic zeta function" was offered which was given by the same formula but with the entries of P still regarded as lying in $\mathbb{R}$. This gives an invariant, but one which does not capture all the desired information. We have appropriated the term "stochastic zeta function" for the Parry-Tuncel function because we regard it as the correct end product of this line of development.)

For SFT's, the dimension module (§5.3) was a certain $\mathbb{Z}[t, t^{-1}]$-module. The dimension module of a Markov shift is a $R[t, t^{-1}]$-module, where R may be taken to be the ring of Laurent polynomials in n variables (by identifying the variables with generators of the canonical subgroup mentioned earlier)—it is a version of Tuncel's dimension module [Tu]. The variable-length graphs still work—but now they are labelled by elements of $tR_+[t]$ rather than $t\mathbb{Z}_+[t]$. Recall if A is a matrix over $t\mathbb{Z}_+[t]$ presenting an SFT, then the shift-equivalence-over-$\mathbb{Z}$ data for the SFT is encoded as the isomorphism class of the $\mathbb{Z}[t, t^{-1}]$-module $\mathrm{cok}(I - A)$. In the stochastic case, the entries of A lie in $tR_+[t]$, and the shift-equivalence-over-R data is encoded as the isomorphism class of the $R[t, t^{-1}]$-module $\mathrm{cok}(I - A)$. (There is even a notion of stochastic flow equivalence, for which one invariant [P2,Ara] can be interpreted in the following way: apply the coinvariants functor to the $R[t, t^{-1}]$-module $\mathrm{cok}(I - A)$ to obtain the R-module $\mathrm{cok}(I - A(1))$, which is an invariant of stochastic flow equivalence. Here $A(1)$ denotes the matrix obtained by substituting 1 for t in A. Similarly the element $\det(I - A(1))$ is an invariant of stochastic flow equivalence [P2,Ara].)

Even in the primitive case, shift equivalence does not imply strong shift equivalence for matrices over R_+ [B3]. (A matrix over R_+ is primitive if some power has every entry nonzero.) Thus even in the primitive case we must be concerned with order in the classification of matrices up to shift equivalence over R_+. We can express the order information in the module framework by considering the $R_+[t, t^{-1}]$-module $\mathrm{cok}(I - A)$ as an ordered module. Here the positive set of $\mathrm{cok}(I - A)$ is the set of vectors which have all entries in $R_+[t]$ after multiplication by a sufficiently large power of A. Now shift equivalence of matrices over $tR_+[t]$ is equivalent to

To think of this in a finite way, consider an irreducible matrix A over R_+ as a matrix with entries in $\mathbb{Z}_+[x_1, x_1^{-1}, ..., x_n, x_n^{-1}]$ and let $p(t)$ be its characteristic polynomial. Factor $p(t)$ in $\mathbb{Z}[x_1, x_1^{-1}, ..., x_n, x_n^{-1}][t]$. There will be one factor such that for any substitution of positive reals for the variables $x_1, ..., x_n$, this factor will have the largest root. You may regard that factor as the beta function. What are the beta functions of irreducible matrices over R_+? There is not even a good conjecture at present. Progress to date has rested on a blend of techniques from algebra, geometry and analysis [H3-6,deA] .

Problem. What are the beta functions of primitive matrices of Laurent polynomials?

Problem. What are the zeta functions arising from primitive matrices of Laurent polynomials? (That is, which polynomials can arise as $\det(I - tA)$ for some primitive matrix A whose entries are Laurent polynomials with nonnegative integral coefficients?)

6.3 Boolean matrices. The Boolean semiring $\mathbb{B}$ is the set $\{0, 1\}$, with addition and multiplication defined as the quotient of these operations on the nonnegative reals by the map sending 0 to 0 and sending all positive numbers to 1. So, 1 is a multiplicative identity, 0 is an additive identity, $1 + 1 = 1$, $0 \times 0 = 0$. Relations among nonnegative matrices project to relations among Boolean matrices (matrices over $\mathbb{B}$). For example, if A and B are shift equivalent nonnegative matrices, then their Boolean images are shift equivalent over $\mathbb{B}$; and if A and B are to be strong shift equivalent, then their Boolean images must be. So it makes sense to look at Williams' problem for matrices over $\mathbb{B}$, if only to check necessary conditions on Williams' problem for matrices over $\mathbb{Z}_+$ (or $\mathbb{R}_+$, or $\mathbb{Q}_+$).

Kim and Roush have completely classified all Boolean matrices up to shift equivalence and also up to strong shift equivalence [KR5]. Their general result is lovely, but for simplicity we will just state the answer for primitive Boolean matrices (those for which some power is the matrix with every entry 1). They are all shift equivalent! And two such matrices are strong shift equivalent if and only if they have the same powers with zero trace (there are only finitely many such powers). In particular, shift equivalence does not imply strong shift equivalence in this setting.

One useful spinoff of their work is an intriguing tool for controlling the sign patterns of nonnegative matrices. As remarked in [KR5, p.154], if A and B are two nonnegative matrices whose Boolean projections are strong shift equivalent, and their entries lie in a nondiscrete unital subring S of the reals, then B is strong shift equivalent over S_+ to a matrix with the same block sign pattern as A. For example, this with the classification result [KR5] shows that a matrix over the nonnegative rationals $\mathbb{Q}_+$ with positive trace is strong shift equivalent over $\mathbb{Q}_+$ to a matrix with

shift equivalence and strong shift equivalence for SFT's over infinite alphabets. Here a crucial feature is that the morphisms are required to be uniformly continuous.

6.5 Sofic shifts. There are also notions of shift equivalence and strong shift equivalence available for sofic shifts (these are the subshifts which are quotients of SFT's) [BK]. Again strong shift equivalence corresponds to isomorphism and shift equivalence corresponds to isomorphism of all large powers. The equations of (strong) shift equivalence are now applied to elements of an integral semigroup ring, where the semigroup is the semigroup under multiplication of infinite zero-one matrices with all row sums at most one and with all but finitely many entries zero. Now, instead of considering matrices over a commutative integral group ring, we are looking at the (noncommutative) integral semigroup ring of a nonabelian semigroup. Because of this noncommutativity, it is a serious problem even to define an appropriate zeta function [B4]. Nevertheless, Kim and Roush showed that shift equivalence in this setting is decidable [KR4], which matches their result in the SFT setting [KR3]. A key reduction in their proof is an analogue of the Parry-Schmidt result [PSc] for Markov chains, which allowed one to restrict to modules over smaller rings: in the sofic case, for considering shift equivalence (but not strong shift equivalence), Kim and Roush observed by appeal to work of Nasu [N] that for two given systems it suffices to consider modules over the integral semigroup ring of a certain *finite* semigroup (the zero-one matrices of a certain bounded size).

For an explanation of these ideas, we refer to [BK,Sec.1] and [KR4]. Our aim here is primarily to indicate by yet another example that the basic ideas of shift and strong shift equivalence return in various guises to describe symbolic dynamical structures.

SECTION 7. GENERAL MATRICES

In this section we'll consider shift equivalence and strong shift equivalence a bit more generally. This has algebraic and order aspects. We justify the investigation by the symbolic dynamical relevance of (strong) shift equivalence in various settings, and because these relations are algebraically natural in general (e.g., $sr \sim rs$ generates a definition of $K_0(R)$ via projections in $GL(R)$).

First we consider the algebraic aspect. Let R be a unital ring (unital means that R has a multiplicative identity). We define shift equivalence and strong shift equivalence of matrices over R by the same equations we use for $R = \mathbb{Z}$. Does shift equivalence imply strong shift equivalence?

If $R = \mathbb{Z}$, the answer is yes. This was proved in old unpublished preprints by Williams and by Effros [E2],[W2]. They offer related but different arguments, both of which go through for any principal ideal domain. The answer is still yes for a Dedekind domain [BH2]. On the other hand the following is open.

two elements in R_+. We have already looked at shift and strong shift equivalence for $R = \mathbb{Z}$ and $R_+ = \mathbb{Z}_+$. There Williams' problem is open in the irreducible case. Even for matrices of size 2, after all this time, it is settled (in the affirmative) only if the determinant is not less than -1 [Ba1,CuKr,W1].

Problem [Ba1]. Suppose A and B are primitive size 2 matrices over $\mathbb{Z}_+$ which are shift equivalent over $\mathbb{Z}_+$, with $\det A < -1$. Must A and B be strong shift equivalent over $\mathbb{Z}+$?

There are additional ideas [Ba2] but the problem is tough (as can be attested by the calibre of some of the mathematicians who have spent months or years on Williams' problem for matrices over $\mathbb{Z}_+$—Williams, Parry, Franks, Krieger, Marcus, Baker, Handelman, Kim&Roush....).

A sane response is to back off and consider Williams' problem for matrices over the nonnegative rationals or reals. Here Kim and Roush proved that shift equivalent matrices over $\mathbb{Q}_+$ are strong shift equivalent—if each has exactly one nonzero eigenvalue, with multiplicity one (equivalently, the inverse zeta function is $1 - at$, where a is the nonzero eigenvalue) [KR6]. (It is some indication of the difficulty of the problem that this seemingly simple case was open for so long.) They have also added to Baker's viewpoint [Ba2] the development of approximation and homotopic techniques [KR8,9].

But still: strong shift equivalence remains very poorly understood.

Problem. Does shift equivalence over $\mathbb{R}_+$ imply strong shift equivalence over $\mathbb{R}_+$ for strictly positive matrices over the reals?

Section 8. Inverse Problems for Nonnegative Matrices

In [BH1,2], Handelman and I studied certain inverse problems for nonnegative matrices, using tools from symbolic dynamics. As we took pains to explain and motivate the problems there, in this section the discussion will be brief, and we refer to [BH1,2] for more.

8.1 The inverse spectral problem. Let $\Delta = (d_1, ..., d_n)$ be an n-tuple of complex numbers. An old problem asks, when is Δ the spectrum of a nonnegative real matrix of size n? (Δ is the spectrum of a matrix A if the characteristic polynomial is $\chi_A(t) = \prod_i (t - d_i)$. So Δ includes the information about multiplicities and we don't care about the order in which the d_i are listed.)

Necessary conditions on Δ are discussed in [BH1], especially in Appendix 3. The best reference to the literature on this problem is still [BePl], for a more recent discussion see [Mi]. To my knowledge the problem first appears in print in Suleimanova's 1949 paper [Su] (if we neglect the glorious work of Perron and Frobe-

§2.3), and it is appropriate to ask when an n-tuple Δ of nonzero complex numbers can be the nonzero part of the spectrum of some nonnegative matrix. This is also natural from just the matrix viewpoint and certainly occurred to matrix theorists working on the problem (e.g. Charles Johnson). I would guess this viewpoint was never seriously pursued because constructions which could make much use of the relaxed condition were unavailable.

Aside from motivating the pursuit of the nonzero spectrum, symbolic dynamics enters the picture by providing some tools which let one exploit the extra room provided by passing to arbitrarily large matrices subject to a given nonzero spectrum. In some cases there are matrix constructions for which codes need not be mentioned, but whose inspiration comes from coding constructions. There are also the ideas around strong shift equivalence, which provide some direct constructions, and which let one translate coding constructions into matrix results. In this vein there is especially one result, the Submatrix Theorem, which plays an essential role in [BH1,2] and for which the only proof known (so far) relies in a fundamental way on coding ideas independent of matrices. These ideas are the marker methods of the proof of Krieger's Embedding Theorem [Kr3]. Because the Embedding Theorem has been a basic and useful tool in the study of shifts of finite type (read: asymptotic theory of nonnegative integral matrices), it may be that the Submatrix Theorem will have other applications in the study of "asymptotic" aspects of more general nonnegative matrices.

8.2 Submatrix theorem. Let S be a unital subring of the reals. Given square nonnegative matrices A and B over S, with A primitive, the Submatrix Theorem produces (subject to "obvious" necessary conditions on the nonzero spectra) a matrix A' with B as a principal submatrix, where A' is strong shift equivalent over S to A. If we are interested only in invariants of strong shift equivalence (such as the nonzero spectrum), then this provides tremendous control over submatrices for constructions.

If the ring S is discrete, then it must be the integers. The necessary trace condition for the Submatrix Theorem is different in this case, and to avoid discussing it we just refer to [BH1, Thm.1.10].

Recall (§2.3) the polynomial $\det(I - tC)$ determines the nonzero spectrum of a matrix C and vice versa.

SUBMATRIX THEOREM (NON-DISCRETE CASE) [BH1]. *Let S be a dense unital subring of the reals. Suppose that A and B are nonnegative matrices with entries from S, such that A is primitive. Then there exists a primitive matrix C with entries from S such that B is a proper principal submatrix of C and $\det(I-tC) = \det(I-tA)$ if and only the following three conditions hold:*

(1) The spectral radius of B is strictly smaller than that of A.

8.3 Spectral conjecture. Handelman and I [BH1] conjectured that certain "obvious" necessary conditions are sufficient for an n-tuple $\Delta = (d_1, ..., d_n)$ of complex numbers to be the nonzero spectrum of a primitive matrix whose entries lie in a given unital subring S of the reals. (The general case follows easily from the primitive case [BH1].) It seems to us that the supporting evidence [BH1] is fairly overwhelming. (Not that we can prove it.) For example, the conjecture is true if S is the reals, or if S is nondiscrete and $\sum d_i$ is nonzero [BH1].

Those necessary conditions on Δ are a Perron condition, a Galois condition, and a trace condition. The Perron condition is that there be a positive real number which is listed just once in Δ and which is strictly larger than the modulus of any other entry. The Galois condition is simply that the degree n polynomial $\prod(t - d_i)$ must have its coefficients in S. The trace condition is different if $S = \mathbb{Z}$, so here we will just state the trace condition when S is nondiscrete. For $k > 0$, let $t(k)$ denote the sum of the kth powers of the entries of Δ. Then the trace condition is that for all positive integers m and k, two things hold:

(1) $t(k) \geqslant 0$, and

(2) if $t(m) > 0$, then $t(mk) > 0$.

8.4 Generalizing the spectral conjecture. Again let S be a unital subring of the reals. Suppose A is a square matrix over S. The Spectral Conjecture asserts that there exists a primitive matrix B over S with the same nonzero spectrum as A if two necessary conditions, a Perron condition and a trace condition, are satisfied. (Here the Galois condition is automatically satisfied.) The Generalized Spectral Conjecture of [BH2] asserts that under these same necessary conditions, one can require the primitive matrix B to be shift equivalent over S to A. Handelman and I take this opportunity to make a further generalization: we conjecture that under these same necessary conditions, there exists a primitive matrix B strong shift equivalent over S to A.

But, remember, at this moment we do not know if two matrices shift equivalent over S must be strong shift equivalent over S (Sec.7)—i.e., it may be that this generalization is equivalent to the Generalized Spectral Conjecture of [BH2].

Section 9. One-sided Shifts

Let A be a square nonnegative integral matrix. Then A defines a one-sided SFT T_A. This is defined just as the two-sided SFT S_A was, with one difference: now the sequences are $x = x_0 x_1...$ (that is, the coordinates are indexed only by nonnegative integers, not by all the integers). Here the shift map is still continuous and (if A is nondegenerate) surjective, but usually it is not invertible. (In fact, it is invertible only when the space of sequences is finite. When A is nondegenerate, this means

words of length N to the alphabet of T_B such that for all x and for all $i \geq 0$,

$$(fx)_i = F(x_i...x_{i+N-1}).$$

Williams classified these systems up to isomorphism [W1]. (There is an exposition of this in [BFK], which also has a good deal more about these systems.) The classification is beautifully simple, so we will describe it.

Given the matrix A, if possible choose two equal columns i and j; then add row j to row i, and then erase row j and column j. This produces a smaller matrix A_1. The matrix A_1 may also have a pair of equal columns, which we can "amalgamate" as before to get a yet smaller matrix. If B is a matrix obtained by a (finite) sequence of column amalgamations in this way from A, then we call B an amalgamation of A. If B has no pair of equal columns (i.e. cannot be further column-amalgamated), then B is called a total amalgamation of A.

Williams [W1] proved that the total amalgamation of A is independent (up to conjugation by a permutation matrix) of the choices of columns at each step, and that T_A is isomorphic to T_B if and only if the total amalgamations of A and B are the same (up to conjugation by a permutation matrix). So for one-sided SFT's the classification has a clean and simple solution.

However there is another fundamental problem which seems much harder in the one-sided case. Recall a matrix is nondegenerate if it has no zero rows or columns, and a primitive matrix is a square nonnegative matrix some power of which has every entry strictly positive.

Problem. Suppose B is a square nondegenerate nonnegative matrix and A is a primitive matrix. Give necessary and sufficient conditions under which the one-sided SFT T_B is isomorphic to a proper subsystem of the one-sided SFT T_A.

Let us immediately express this problem in purely matrix terms. With no loss of generality, we may (and should) assume that B is a total amalgamation. Then T_B is isomorphic to a proper subsystem of T_A if and only if for some matrix C presenting a higher block presentation of T_A, the matrix C has a proper principal submatrix D such that B is a total amalgamation of D.

The solution of the corresponding two-sided case is Krieger's embedding theorem [Kr3]. There, simple necessary conditions on entropy and periodic points are sufficient for the proper embedding: the entropy of S_A must strictly exceed that of S_B, and for every positive integer n the number of orbits of cardinality n for S_B cannot exceed the corresponding number for S_A.

These necessary conditions do not suffice in the one-sided case. For example, a point with k preimages must be sent to a point with at least k preimages. Similarly there are constraints counting preimages of preimages, etc.; and these are mixed

systems [W1,BFK]. So it is plausible that in the one-sided case a proof of the embedding theorem may be much more closely tied to graphs and matrices (and therefore could lead to better algorithms for constructing the embedding codes which exist by Krieger's theorem in the two-sided case). There have only been a few papers on one-sided SFT's ([W1],[BFK],[As1]), so we have a situation common in symbolic dynamics: there's not too much to learn, but there is something to invent.

The problem above is posed for embedding SFT's only (rather than general subshifts), in contrast to the statement of Krieger's theorem. This is to emphasize the matrix aspect. There is almost surely no loss of generality. It is an exercise to check that an embedding of a subshift (one-sided or two) into a shift of finite type always extends to an embedding of some SFT containing the subshift. Every subshift is a nested intersection of SFT's in a well-understood way. So from the solution of the problem above, one should be able without great difficulty to give the solution to the general problem. (This is certainly the case for two-sided shifts—but in that case, at present there is no real simplification of Krieger's proof obtained by embedding only SFT's.)

Section 10. Quotients

Let S_A and S_B be (for simplicity) mixing SFT's. When is S_B a quotient of S_A? That is, when is there a code from S_A onto S_B? It is easy to check that a necessary condition is the entropy inequality

$$h(S_A) \geq h(S_B).$$

There is a fundamental dichotomy for such maps. If $h(S_A) > h(S_B)$ and f is a quotient map from S_A onto S_B, then the points with uncountably many preimages under f comprise a residual subset of S_B. If $h(S_A) = h(S_B)$ and f is a quotient map from S_A onto S_B, then there is a positive integer N such that no point in S_B has more than N preimages under f [CP].

As it turns out, the case of strict inequality is relatively easy [B1]. If $h(S_A) > h(S_B)$, then S_B is a quotient of S_A if and only if the trivially necessary periodic point condition holds: for all $n > 0$,

(*) $$tr(A^n) > 0 \Longrightarrow tr(B^n) > 0.$$

The equal-entropy case is much more rigid, subtle and algebraic. The periodic point condition (*) is still necessary for S_B to be a quotient of S_A. But also, the dimension module of S_B must be a quotient of a closed submodule of the dimension module of S_A [KMT]. (In the terminology of §3.1, a closed submodule is given by restricting A' to a subgroup H of $G(A)$ which is the intersection of $G(A)$ with an A'-invariant subspace of V_A. A submodule given by restricting A' to some A'-invariant

Then S_B is a quotient of S_A.

This conjecture is strongly supported by the work of Jonathan Ashley [As2]. For primitive integral matrices A and B of equal spectral radius satisfying (*), he showed that S_B is a quotient of S_A by a closing map if and only if the dimension module of S_B is a quotient or closed subsystem of the dimension module of S_A. (Closing maps are topologically conjugate to resolving maps, which can be constructed from certain matrix equations and are the most useful codes for industrial applications [ACH]. See [AM] and [BMT] for background.)

Ashley's proof begins with a construction (the Eventual Factors Theorem of [BMT]) which produces interrelated quotient maps of higher powers of the shifts. This construction is derived from matrix equations which capture the quotient relation for the dimension modules. One scheme for approaching the conjecture above is to mimic this pattern: first find matrix equations for the dimension condition, then find an "eventual" construction from these analogous to the starting result from [BMT], and then adapt Ashley's arguments for coding into and out of long periodic blocks. (Caveat: this may be a red herring.)

There are, at least, some nice matrix equations for the dimension condition. Let A and B be primitive integral matrices of equal spectral radius. Then the dimension module for B is a quotient of a closed submodule of the dimension module for A if and only if there are positive integral matrices R, S and a positive integer n such that for all positive integers k,

$$RA^kS = B^{n+k}.$$

We leave a proof to the interested reader (it may help to consult [BMT,Sec.2]).

REFERENCES

[ACH]   R.ADLER, D.COPPERSMITH & M.HASSNER, *Algorithms for sliding block codes*, IEEE Trans. Info.Th., Vol. IT-29, pp. 5-22, 1983.

[AF]   R.ADLER & L.FLATTO, *Geodesic flows, interval maps, and symbolic dynamics*, Bulletin (New Series) A.M.S. Vol. 25, No.2 (1991), pp.229-334.

[AFKM] R.ADLER, J.FRIEDMAN, B.KITCHENS & B.MARCUS, *State splitting for variable-length graphs*, IEEE Trans. Info. Th. Vol. IT-32, No. 1, pp. 108-113, 1986.

[Ara]   P. ARAUJO, *A stochastic analogue of a theorem of Boyle's on almost flow equivalence*, Preprint, 1991.

[ArMa] M.ARTIN & B.MAZUR, *On periodic points*, Annals of Math. 81 (1969), 82-99.

[As1]   J.ASHLEY, *Marker automorphisms of the one-sided d-shift*, Erg.Th.& Dyn.Syst. 10 (1990), 247-262.

[As2]   J.ASHLEY, *Resolving factor maps for shifts of finite type with equal entropy*, Ergod. Th. & Dyn. Syst. 11 (1991), 219-240.

[Ba1]   K.A.BAKER, *Strong shift equivalence of 2x2 matrices of non-negative integers*, Ergod.Th.& Dyn.Syst. 3 (1983), 541-558.

[Bow1]   R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Springer Lec. Notes in Math. 470, Springer-Verlag (1975).

[Bow2]   R. Bowen, *On Axiom A diffeomorphisms*, CBMS Regional Conf. Ser. in Math., No. 35, Amer.Math.Soc., Providence, RI (1977).

[BowF]   R. Bowen & J. Franks, *Homology for zero-dimensional basic sets*, Annals of Math. 106 (1977), 73-92.

[BowL]   R. Bowen & O.E.Lanford, *Zeta functions of restrictions of the shift transformation*, Proc. Symp. Pure Math. A.M.S. 14 (1970), 43-50.

[B1]   M. Boyle, *Lower entropy factors of sofic systems.*, Erg.Th.&Dyn.Syst. 4, (1984), 541-557.

[B2]   M. Boyle, *Shift equivalence and the Jordan form away from zero*, Erg.Th.& Dyn. Syst. 4 (1984), 367-379.

[B3]   M. Boyle, *A zeta function for homomorphisms of dynamical systems*, J. London Math. Soc. (2) 40 (1989) 355-368.

[B4]   M. Boyle, *The stochastic shift equivalence conjecture is false*, Proceedings of the 1991 Yale Adlerfest, to appear.

[BFK]   M.Boyle, B.Kitchens & J.Franks, *Automorphisms of one-sided subshifts of finite type*, Ergod.Th.&Dyn.Syst. 10 (1990), 421-449.

[BH1]   M. Boyle & D. Handelman, *The spectra of nonnegative matrices via symbolic dynamics*, Annals of Math. 133 (1991), 249-316.

[BH2]   M. Boyle & D. Handelman, *Algebraic shift equivalence and primitive matrices*, Trans. AMS, to appear.

[BMT]   M.Boyle, B. Marcus & P. Trow, *Resolving Maps and the Dimension Group for Shifts of Finite Type*, Memoirs AMS 377 (1987).

[BK]   M.Boyle & W. Krieger, *Almost Markov and shift equivalent sofic systems*, in *Dynamical Systems, Proc. Special Year in Dynamics*, 1986-87 at the University of Maryland, Springer Lec. Notes in Math 1342, Springer-Verlag.

[Br]   K. Brown, *Cohomology of Groups*, Springer GTM 87. Springer-Verlag 1982.

[CP]   E. Coven & M. Paul, *Endomorphisms of irreducible subshifts of finite type*, Math Systems Th. 8(1974), 167-175.

[Cu]   J.Cuntz, *A class of C*-algebras and topological Markov chains II: reducible chains and the Ext-functor for C*-algebras*, Inventiones Math. 63, 25-40 (1981).

[CuKr1]   J.Cuntz & W.Krieger, *Topological Markov chains with dicyclic dimension groups*, J.fur Reine und Angew. Math. 320 (1980), 44-51.

[CuKr2]   J.Cuntz & W.Krieger, *A class of C*-Algebras and topological Markov chains*, Inventiones Math. 56, 251-268 (1980).

[deA]   V. De Angelis, *Polynomial beta functions and positivity of polynomials*, PhD. dissertation, University of Washington, 1992.

[DGS]   M. Denker, C. Grillenberger & K.Sigmund, *Ergodic Theory on Compact Spaces*, Springer Lec. Notes in Math. 527, Springer-Verlag (1976).

[E1]   E.G.Effros, *Dimensions and C*-Algebras*, CBMS 46 (1981), A.M.S., Providence, Rhode Island.

[E2]   E.G.Effros, *On Williams' problem for positive matrices*, Unpublished manuscript, ca. 1981.

[F1]   J. Franks, *Homology and Dynamical Systems*, CBMS 49 (1982), A.M.S., Providence, Rhode Island.

[F2]   J. Franks, *Flow equivalence of subshifts of finite type.*, Erg.Th.Dyn. Syst. 4(1984), 53-66.

[Fri1]   D. Fried, *Rationality for isolated expansive sets*, Advances in Math. 65, 35-38 (1987).

[Fri2]   D. Fried, *Finitely presented dynamical systems* Erg.Th.Dyn.Syst.7(1987), 489-507.

[H1]   D. Handelman, *Positive matrices and dimension groups affiliated to C*-Algebras and topological Markov chains*, J. Operator Th. 6 (1981), 55-74.

[H2]   D. Handelman, *Eventually positive matrices with rational eigenvectors*, Erg.Th.&Dyn.Syst. 7 (1987), 193-196.

[H3]   D. Handelman, *Positive polynomials, convex integral polytopes and a random walk problem*, Springer Lec. Notes 1282, Springer Verlag (1987).

[HMS]   C. HEEGARD, B. MARCUS & P. SIEGEL, *Variable length state splitting with applications to average run-length constrained (ARC) codes*, IEEE-Information Theory, v. 37 (1991), pp. 759-777.

[Hu]   D. HUANG, *Flow equivalence of reducible shifts of finite type*, In preparation.

[Ke]   R.B.KELLOGG, *Matrices similar to a positive or essentially positive matrix*, Lin.Alg.Applic. 4 (1971), 191-204.

[KR1]   K.H.KIM & F.ROUSH, *Some results on decidability of shift equivalence*, J. Combinatorics, Info.Sys.Sci. 4 (1979), 123-146.

[KR2]   K.H.KIM & F.W.ROUSH, *On strong shift equivalence over a Boolean semiring*, Erg.Th.& Dyn.Syst. 6 (1986), 81-97.

[KR3]   K.H.KIM & F.W.ROUSH, *Decidability of shift equivalence. Dynamical Systems (Proceedings, Univ. of Maryland 1986-87)*, Springer Lec. Notes 1342, Springer(1988), 374-424.

[KR4]   K.H.KIM & F.W.ROUSH, *An algorithm for sofic shift equivalence*, Erg. Th. & Dyn. Syst. 10 (1990), 381-393.

[KR5]   K.H.KIM & F.W.ROUSH, *Strong shift equivalence of boolean and positive rational matrices*, Linear Algebra Appl. 161: 153-164 (1992).

[KR6]   K.H.KIM & F.W.ROUSH, *Full shifts over R_+ and invariant tetrahedra*, PU.M.A. Ser. B, Vol.1 (1990), No. 4, pp. 251-256.

[KR7]   K.H.KIM & F.W.ROUSH, *Williams' conjecture is false for reducible subshifts*, Journal AMS 5 (1992), 213-215.

[KR8]   K.H.KIM & F.W.ROUSH, *Path components of matrices and strong shift equivalence over Q_+*, Linear Algebra Appl. 145: 177-186 (1991).

[KR9]   K.H.KIM & F.W.ROUSH, *Strong shift equivalence over subsemirings of Q_+*, PU.M.A. Ser.B, Vol.2 (1991), 33-42.

[KRW]   K.H.KIM, F.W.ROUSH & J.B.WAGONER, *Automorphisms of the dimension group and gyration numbers of automorphisms of a shift*, Journal AMS 5 (1992), 191-211.

[KMT]   B. KITCHENS, B. MARCUS & P. TROW, *Eventual factor maps and compositions of closing maps*, Erg.Th.Dyn.Syst. 11 (1991), 85-113.

[Kr1]   W. KRIEGER, *On a dimension for a class of homeomorphism groups*, Math. Ann. 252 (1980), 87-95.

[Kr2]   W. KRIEGER, *On dimension functions and topological Markov chains*, Invent. Math. 56 (1980), 239-250.

[Kr3]   W. KRIEGER, *On the subsystems of topological Markov chains*, Erg.Th.Dyn. Syst. 2 (1982), 195-202.

[L]   D. LIND, *The entropies of topological Markov shifts and a related class of algebraic integers*, Erg. Th. Dyn. Syst. 6(1986),571-582.

[LL]   R. LOEWY & D. LONDON, *A note on an inverse problem for nonnegative matrices*, Linear and Multilinear Algebra 6 (1978),83-90.

[M]   B. MARCUS, *Factors and extensions of full shifts*, Monats. fur Mathematik 88 (1979), 239-247.

[MSW]   B.H.MARCUS, P.H.SIEGEL & J.K.WOLF, *Finite-state modulation codes for data storage*, IEEE Journal of Selected Areas in Communications, Vol. 10 (1992), 5-37.

[MT1]   B.MARCUS & S.TUNCEL, *The weight-per-symbol polytope and embeddings of Markov chains*, Ergod.Th. & Dyn.Syst. 11 (1991), 129-180.

[MT2]   B.MARCUS & S.TUNCEL, *Entropy at a weight-per-symbol and embeddings of Markov chains*, Invent. Math. 102 (1990), 235-266.

[Mi]   H. MINC, *Nonnegative Matrices*, Wiley Inter-Science (1988).

[N]   M. NASU, *Topological conjugacy for sofic systems and extensions of automorphisms of finite subsystems of topological Markov shifts*, Proceedings of Maryland Special Year in Dynamics 1986-87. Springer Lecture Notes 1342 (1988), Springer-Verlag.

[Ne]   M. NEWMAN, *Integral Matrices*, Academic Press, New York, 1972.

[P1]   W. PARRY, *Intrinsic Markov chains*, Trans. AMS 112 (1964), 55-66.

[P2]   W. PARRY, *Notes on coding problems for finite state processes*, Bull. London Math. Soc. 23 (1991) 1-33.

[PT2]   W. PARRY & S. TUNCEL, *On the stochastic and topological structure of Markov chains*, Bull. London Math. Soc. 14 (1982) 16-27.

[PW]   W.PARRY & R.F.WILLIAMS, *Block coding and a zeta function for Markov chains*, Proc. LMS 35 (1977), 483-495.

[Pe]   D. PERRIN, *On positive matrices*, Theoretical Computer Science 94 (1992) 357-366.

[R]   H.J.RYSER, *Combinatorial Mathematics*, Carus Monographs No. 14, Math. Assoc. of America, 1963.

[Sh]   C. SHANNON, *A Mathematical Theory of Communication*, Bell Sys. Tech. J. 27 (1948) 379-423; 623-656.

[ShWe]   C. SHANNON & W. WEAVER, *The Mathematical Theory of Communication*, University of Illinois Press (1949) (many reprintings).

[Su]   H.R. SULEIMANOVA, *Stochastic matrices with real eigenvalues*, Dokl. Akad. Nauk. SSSR 66 (1949), 343-345 (Russian).

[Tu]   S. TUNCEL, *A dimension, dimension modules and Markov chains*, Proc. LMS 3, Vol. 46 (1983), 100-116.

[Wa1]   J.WAGONER, *Markov partitions and K2*, Pub. Math. IHES No. 65 (1987), 91-129.

[Wa2]   J.B.WAGONER, *Topological Markov chains, C*-algebras and K_2.*, Advances in Math. Vol. 71, No. 2 (1988), 133-185.

[Wa3]   J.B.WAGONER, *Eventual finite generation for the kernel of the dimension group representation*, Trans. AMS 317 (1990), 331-350.

[Wa4]   J.B.WAGONER, *Triangle identities and symmetries of a subshift of finite type*, Pacific J. Math. Vol.44, No.1, 1990, 181-205.

[Wa5]   J.B.WAGONER, *Higher-dimensional shift equivalence and strong shift equivalence are the same over the integers*, Proc. AMS Vol 109, No.2, (1990), 527-536.

[W1]   R.F.WILLIAMS, *Classification of subshifts of finite type*, Annals of Math. 98 (1973), 120-153; Errata, Annals of Math. 99 (1974), 380-381.

[W2]   R.F.WILLIAMS, *Strong shift equivalence of matrices in $GL(2,\mathbb{Z})$*, Unpublished manuscript, ca. 1981.

[W3]   R.F.WILLIAMS, *A new zeta function, natural for links*, To appear in the Proceedings of the Conference in honor of Smale's 60th birthday.