ALGEBRAIC ASPECTS OF SYMBOLIC DYNAMICS

Mike Boyle
Department of Mathematics
University of Maryland
College Park, MD 20742 U.S.A.
mmb@math.umd.edu

These four lectures are an introduction to some algebraic aspects of symbolic dynamics, with no assumption of previous background. They were delivered in Temuco, Chile in January 1997 as part of the School on symbolic dynamics organized by Francois Blanchard and Alejandro Maass. The intent is much more to describe ideas and invariants than to give proofs, and this tendency becomes more pronounced in the last two lectures.

Lecture I is elementary background on subshifts. The parts 1-3 are elementary (but cover a lot of ground) and the proofs left to the reader are feasible.

Lecture II describes and mostly explains the standard matrix invariants for shifts of finite type.

Lecture III explains the basics of dimension groups and shift equivalence, with some hints at further developments. There is a greater emphasis on the order structure of dimension groups than we need for the application in this lecture. This reflects the fundamental importance of the order structure of these groups in other situations.

In Lecture IV, we throw proof to the winds to describe one of the deepest results on SFT's, the Kim-Roush-Wagoner Factorization Theorem, and the Kim-Roush counterexamples to Williams' Conjecture.

We cover a lot, but still we provide only a narrow algebraic slice of symbolic dynamics. (For example we don't even mention systems of algebraic origin! See the lectures of Kitchens.) For a general introduction to the symbolic dynamics around SFT's, see the very fine introduction

> [LM] D. Lind and B. Marcus. An introduction to symbolic dynamics and coding (1995), Cambridge University Press.

In these few lectures, I give few references and don't mention many of the people who made contributions to what I describe. [LM] does cover the historical development, and contains a long appendix surveying advanced topics, which I strongly recommend as an antidote to my omissions. For something shorter, which in addition covers more on III.4 below, see my survey

> [B] M. Boyle. Symbolic dynamics and matrices, in Combinatorial and Graph-Theoretical Problems in Linear Algebra (1993), Vol. 50 of IMA Volumes in Math. and its Applications, eds. R. A. Brualdi, S. Friedland and V. Klee, pp. 1-38.

Both [LM] and [B] have extensive bibliographies.

Since the original delivery of these lectures, some errata were corrected and small changes made; and a final paragraph was added on the Kim-Roush counterexample to the irreducible Williams' Conjecture.

I thank Pierre-Paul Romagnoli for detailed and helpful comments on the original lectures.

I. THE WORLD OF SUBSHIFTS

1. General subshifts

1.1 Dynamical systems.

For the purposes of these lectures, 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. Apart from occasional remarks, T will be a homeomorphism.

1.2 Full shifts.

The system which is the full shift on n symbols (also called the n-shift) is defined as follows. We give a finite set of n elements — say, $\{0, 1, ..., n-1\}$ — the discrete topology. (This finite set is often called the *alphabet*.) We let X be the product of countably many copies of this set, with the copies indexed by $\mathbb{Z}$. We think of an element x 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 metrizable space. A metric compatible with the topology is given by defining, when x is not equal to y,

$$\mathrm{dist}(x,y) = 2^{-k}, \qquad \text{where } k = \min\{|i| : x_i \neq y_i\}.$$

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. Thus X is zero dimensional.

There is a natural shift map S sending X into X, defined by shifting the index set by one: $(Sx)_i = x_{i+1}$. (This is the “dynamics” in symbolic dynamics.) 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).

1.3 Subshifts.

A subshift is a subsystem of some full shift (X,T) on n symbols. This means that it is a homeomorphism obtained by restriction of T to some compact subset 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 $i \le j$, W is not equal to $y_i...y_j$. That is, Y is the subset of all sequences which avoid the forbidden words.

Concisely: any subshift may defined by excluding a countable collection of words.

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). Of course, if S is SFT, then S could also be defined by excluding an infinite set of words (for example, excluding 00 in the 2-shift is equivalent to excluding all words which contain 00). (But any list of forbidden words defining an SFT will by a compactness argument contain a finite list of forbidden words which defines the same SFT.)

1.4 Examples.

Let S be the subshift defined by restricting the two-shift to the set Y of sequences in which the word 00 never occurs. (That is: S is the subshift of the twoshift defined by excluding the word 00.) Then S is SFT.

Let T be the subshift defined by excluding all words 100...01 where the number of zeros is odd. Then T is a subshift, and we will see below it is not SFT. (T is the “even system” of B. Weiss.)

2. Block codes

2.1 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.

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

2.2 Block codes.

Now suppose F is a function from words of length $2n+1$ which occur in S-sequences into some finite set A. This function F gives a rule for taking an input sequence x of S and producing an output sequence fx. The sequence fx is determined by defining each of its coordinate symbols $(fx)_i$ by the rule

$$(fx)_i = F(x_{i-n} ... x_{i+n}).$$

It is easy to see that the map f produced in this way will be continuous and commute with the shift. That is, such a map f defines a code (called a block code) from S into the full shift on the alphabet A (or into any subshift which contains the image of f).

Example. The shift map itself is given by the code $(fx)_i = x_{i+1}$.

Example. Let S be the twoshift and $F(ijk) = j + k \pmod 2$, that is $(fx)_i = x_i + x_{i+1} \pmod 2$. Then at a sample point x we see

$$\begin{aligned} x &= x_0 x_1 \dots &&= 01101011 \dots \\ fx &= (fx)_0 (fx)_1 \dots &&= 1011110 \dots \end{aligned}$$

Notational caveat: sometimes "block code"is used to refer only to f such that $(fx)_i = F(x_i ... x_{i+n})$ for some function F.

2.3 Curtis-Hedlund-Lyndon.

The "Curtis-Hedlund-Lyndon Theorem" asserts that every code is a block code.

PROOF. Suppose f is a code from S to T. By continuity, for each symbol a in the alphabet of T, the inverse image of the closed open set $\{x : x_0 = a\}$ is closed open, i.e. compact open.

Any open set is a union of cylinders and any compact open set is a union of finitely many cylinders. If a cylinder C is defined on coordinates $[j, k]$ and $j' < j < k < k'$, then C is a union of cylinders defined on coordinates $[j', k']$. Because the alphabet of T is finite, it follows that we may choose N sufficiently large that for every a, the inverse image of $\{x : x_0 = a\}$ is a union of cylinders defined on coordinates $[-N, N]$.

Now define $F(i_{-N} \dots i_N) = a$ if $\{x : x_{-N} \dots x_N\}$ is contained in the inverse image of $\{x : x_0 = a\}$. Because f commutes with the shift, we have for any i that

$$\begin{aligned} (fx)_i = (T^{-i} f S^i x)_i &= (f S^i x)_0 \\ &= F(\ [S^i x]_{-N} \dots [S^i x]_N\ ) \\ &= F(\ x_{i-N} \dots x_N\ )\ . \qquad \text{QED} \end{aligned}$$

It is often convenient to abuse notation and use the same symbol for the maps we call f and F above.

2.4 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 $(fx)_i = [x_{i-j}...x_{i+k}]$. So, for the output sequence fx, the symbol $(fx)_i$ is the word of length n in x in the coordinates $i-j$ through $i+k$. For example, with $n = 2$ and $[j,k] = [0,1]$, we would see

$$\begin{aligned} x &= x_0x_1\dots &&= 012330\ ... \\ fx &= (fx)_0(fx)_1\dots &&= [01][12][23][33][30]\ ... \end{aligned}$$

The shift map on the set of all output sequences is a subshift T, and f is an isomorphism from S to T. (f is a block code and it is clearly one-to-one.) Given n, the map for the chosen $[j,k]$ is the map for $[0, n-1]$ followed by the jth power of the shift. So, given n, the system T doesn't depend on the choice of j (although the map does). T is called the n-block presentation of S.

> Exercise. 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.

2.5 One-block codes.

A one-block code $f : S \to T$ is a block code for which the symbol $(fx)_0$ is determined by the symbol x_0. So, f is given by a map from the alphabet of S to the alphabet of T.

Suppose $h : S \to T$ is a code, where $(hx)_0$ is determined by say $x_{-n}\dots x_n$. Let S' be the $(2n+1)$-block presentation of S. Let $g : S \to S'$ be the isomorphism given by $(gx)_i = [x_{i-n}\dots x_{i+n}]$. Let $h' : S' \to T$ be the one-block map given by

$$h' :\ [x_{-n}\dots x_n] \mapsto h(x_{-n}\dots x_n)\ .$$

Then $h = h'g$, and we have presented the map h as a one-block map on a higher block presentation. This is often convenient.

3. Shifts of finite type

3.1 Vertex shifts.

We now define vertex shifts, which are examples of shifts of finite type. Notation: throughout these lectures, "graph" means "directed graph".

For some n, let A be an $n\times n$ zero-one matrix. We think of A as the adjacency matrix of a 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, $A(y_k, y_{k+1}) = 1$. We think of Y as the space of doubly infinite walks through the graph, where the walks/itineraries are presented by recording the vertices traversed. The restriction of the shift to Y is a shift of finite type: a sufficient list of forbidden words is the set of words ij such that there is no arc from i to j.

CLAIM: Suppose (Y,S) is SFT, then it is isomorphic to a vertex shift.

PROOF: Let L be a finite list of words such that Y is the set of all sequences (from some full shift) in which no word of L occurs. Now, if a word W has length less than n, then excluding W is equivalent to excluding all words of length n which contain W. So without loss of generality, we can assume that all the words in the list L have the same length, say $n+1$. (Then we say S is an "n-step" SFT.)

It follows that the n-block presentation of S is a "1-step" SFT, definable by excluding some words of length two. (For example, a forbidden word 12345 in S corresponds to a forbidden word [1234][2345] in the 4-block presentation of S.) So after passing to the higher block presentation, WLOG we may suppose $n = 1$. We claim that S is then a vertex shift.

Let the alphabet of S be the vertex set of a graph. In this graph, there is an edge from i to j iff the word ij is not forbidden in S. Now the vertex shift for this graph has the same alphabet as S, and the same list of forbidden words of length 2, so it equals S. QED

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

The edge shift presentation is 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. For example, the n-shift as an edge shift is defined by the matrix $(\, n\,)$, as a vertex shift it is defined by the $n \times n$ matrix with every entry equal to 1.

Another reason is functoriality. Working only with zero-one matrices rules out some useful matrix operations (such as taking powers) 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 $S^{[n]}$ whose alphabet is the set of S-words of length n. An isomorphism from S^n to $S^{[n]}$ is given by the map f which sends a point x to the point y such that for all k in $\mathbb{Z}$,

$$y_k = [x_{kn} \dots x_{(k+1)n-1}].$$

CLAIM: Suppose an edge shift S is defined by a matrix A. Then the subshift $S^{[n]}$, after a renaming of symbols, is the edge shift defined by A^n.

PROOF: Let G be the graph with adjacency matrix A and let $G^{[n]}$ be the graph with adjacency matrix A^n. Let these graphs have the same vertex set, so that $A(i,j)$ is the number of edges from i to j in G and $A^n(i,j)$ is the number of edges from i to j in $G^{[n]}$.

An element $[e_1 e_2 \dots e_n]$ in the alphabet of S^n is a path in G of n edges from some vertex i to some vertex j. The number of such paths from i to j is $A^n(i,j)$. So there is a bijection from the alphabet of $S^{[n]}$ to the alphabet of S_{A^n} (i.e., the edge set of $G^{[n]}$) which respects initial and terminal vertex. This renaming of symbols defines a one-block isomorphism from $S^{[n]}$ to S_{A^n}. QED

3.3 Matrices.

Because any SFT is isomorphic to an edge SFT S_A, all information about the dynamics of S_A is determined by the matrix A. So we naturally will try to read dynamical invariants from the matrix A. This is the topic of the second lecture.

But it is worth noting right away that infinitely many matrices of arbitrarily large size can define isomorphic SFTs. For example, the n-block presentation of S_A is the edge shift defined by the graph $G(n)$ whose vertices are the words of length $n-1$ in S_A, and whose edges correspond to the obvious way to the words of length n. 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.

3.4 Markov characterization.

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.

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 the SFT such that $x_0 \ldots x_{(n-1)} = y_0 \ldots y_{n-1}$. Then it follows that the doubly infinite sequence $z = x_{(-\infty,n-1]}y_{[n,+\infty)}$ must also be a point in the SFT. (More precisely, this point z is defined by setting $z_i = x_i$ if $i < n$, and $z_i = y_i$ if $i \geq 0$.)

Put another way: 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). This by the way is a special case of the "local product structure" of hyperbolic dynamics.

Obviously an edge SFT has the Markov property above. One can check that a subshift isomorphic to a subshift with the Markov property also has the Markov property, and must be SFT.

3.5 Applications.

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.

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 $1, 2, \ldots, n$. 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.

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

4. SFT-like subshifts

4.1 Sofic shifts.

A subshift is *sofic* (this word is Hebrew for finite) if it is a quotient of an SFT.

Recall that if $f : S \to T$ is a block code, then there is a one-block code f' from a higher block presentation S' of S onto T. For n large enough, the n-block presentation of an SFT is an edge SFT, and f' is given as a function from the alphabet of S (the edge set) to the alphabet of T. We can think of this function as a labelling of the edges of a graph. So, T is a subshift whose domain is the set of sequences which can be read off labels along infinite walks through some labelled graph.

For an example, consider the graph on vertices u and v, with three edges: one edge from u to u labelled 1, one edge from u to v labelled 0, and one edge from v to u labelled 0. The sofic shift corresponding to this labelled graph is the *even system* of Weiss. It is easy to see that the even system does not have the Markov property and therefore gives an example of a sofic shift which is not SFT.

The sofic shifts can be characterized in several natural ways. They are perhaps the most natural "finitely defined" class of subshifts. A natural subclass, with several characterizations, are the AFT ("almost finite type") systems of Marcus. A sofic shift is AFT if it is the quotient of a transitive SFT by a finite-to-one open map. (We will say a subshift is transitive if it has a dense forward orbit.) The even system is AFT.

4.2 Specification.

Specification is a concept which comes from smooth hyperbolic dynamics, where it has a complicated formulation. A good introduction to this is

[DGS] M. Denker, C. Grillenberger and K. Sigmund. Ergodic Theory on Compact Spaces. Springer Lec. Notes in Math 527 (1970).

It was A. Bertand who showed that this property has a very simple and natural formulation in the special case of subshifts: a subshift S has the specification property if and only if it has a uniform transition length: i.e., there exists a positive integer M such that for any S-words U and W there exists and S-word V of length M such that UVW is an S-word.

4.3 Synchronized systems.

A subshift S is *synchronized* if it is transitive with a synchronizing block. A synchronizing block is a word W which gives a weakened version of the Markov property: whenever x and y are points in S such that $x_0 ... x_{(n-1)} = y_0 ... y_{n-1} = W$, then $z = x_{(-\infty,n-1]} y_{[n,+\infty)}$ is also be a point in S.

4.4 Coded systems.

A *coded system* is a subshift S for which there is a countable collection of words such that the sequences which are concatenations of these words are a dense subset in S. Every synchronized system is coded. The coded systems, introduced with the synchronized systems by Blanchard and Hansel, have several natural characterizations. For example, a subshift is coded iff it contains a dense increasing union of transitive SFT's.

4.5 The progression

For mixing subshifts we have a list of implications (which could be lengthened)

$$SFT \Longrightarrow AFT \Longrightarrow \text{sofic} \Longrightarrow \text{specification} \Longrightarrow \text{synchronized} \Longrightarrow \text{coded} .$$

(A system with specification must be mixing; apart from specification, the implications hold for merely transitive subshifts.) In an SFT, a point can be built up from words satisfying local constraints. This property grows weaker as we move along the list above.

There are only countably many sofic systems. There are uncountably many shifts with specification.

Specification guarantees a unique measure of maximal entropy with full support, which is a natural limit of certain measures on periodic points. Also specification guarantees the subshift is *almost sofic*: its entropy is the supremum of the entropies of the SFT's it contains as subsystems.

All this fails for synchronized shifts, and even more for general coded shifts. Synchronized systems must still have a large group of symmetries (D. and U. Fiebig), but this breaks down for coded systems.

Coded systems still have dense periodic points, and a transitive system with dense periodic points need not be coded, even if it has positive entropy given by the growth rate of the periodic points (K. Petersen). A transitive subshift with dense periodic points can have little in the way of SFT-like properties.

For the "transsofic" subshifts above, a good start are the papers of Blanchard and Hansel and the Fiebigs, referenced in [LM].

5. Minimal subshifts

At the other end of the world are the minimal subshifts. A system is minimal if it has no proper subsystem; equivalently, every orbit is dense. An easy argument with Zorn's Lemma shows that every topological dynamical system contains a subsystem which is minimal.

The class of minimal subshifts is absolutely vast. [One indication: Krieger proved that every ergodic finite entropy measure preserving isomorphism of Lebesgue space is isomorphic as a measurable system to some uniquely ergodic minimal system.]

In one approach, these systems are studied with abstract, transcendental structures. This sort of approach is useful in applying topological dynamical ideas to combinatorial number theory.

There are also classes and approaches of a more finite nature. Among the most beautiful and historically rich of these classes are the substitution systems.

II. MATRIX INVARIANTS FOR SFTS.

We recall that a shift of finite type (SFT) can be presented as an edge shift S_A. In this lecture we build up a dictionary between various dynamical properties of the SFT and invariants (mostly algebraic) of the matrix A. By a matrix invariant of A we will mean a property of A which is the same for matrices A and B which determine isomorphic shifts of finite type. After an initial section on nonnegative matrices, we'll go through the dictionary of corresponding properties by sections as follows.

	Dynamics	**Matrices**
2.	Transitivity	[irreducible]
3.	Mixing	[primitive]
4.	Entropy	[spectral radius]
5.	Periodic points	[traces of powers]
6.	Zeta function	[nonzero eigenvalues]
7.	Isomorphism	[SSE]
8.	Eventual isomorphism	[SE]
9.	Flow equivalence	[cok$(I-A)$ + sign]
10.	Relations	

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 a matter of avoiding trivialities. 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. For example, the following two matrices define the same SFT:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad , \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Restricting to nondegenerate matrices allows statements about the matrix itself rather than its nondegenerate core.

1. Nonnegative Matrices

The matrix A is *irreducible* if for every (i,j) there exists $n>0$ such that $A(i,j)>0$. (This n can depend on (i,j) — consider a cyclic permutation matrix.) The matrix A is *primitive* if there exists $n>0$ such that A^n has all entries strictly positive.

By the Perron-Frobenius Theorem,
— the spectral radius λ_A of A is an eigenvalue of A
— if A is irreducible, then λ_A has a strictly positive eigenvector
— if A is primitive, then λ_A is strictly larger than the modulus of any other eigenvalue.

2. Transitivity

The basic fact is that an SFT S_A is transitive if and only if A is irreducible. (This relies on our standing assumption that A is nondegenerate.) Now we explain.

The word "transitivity" has been used to describe two properties of a dynamical system (X,T):

(a) For all nonempty open sets U,V, there exists n such that $(T^nU)\cap V\neq\emptyset$

(b) For all nonempty open sets U,V, there exists $n\geq 0$ such that $(T^nU)\cap V\neq\emptyset$

We remark that a standard Baire category argument shows that these properties are equivalent respectively to the following:

(α) T has a dense orbit (i.e. some set $\{T^nx : n\in\mathbb{Z}\}$ is dense)

(β) T has a dense forward orbit (i.e. some set $\{T^nx : n\in\mathbb{Z}, n\geq 0\}$ is dense)

The properties are in general slightly different. In these lectures, we use transitivity to refer to the property (b).

For a subshift S, cylinders give a basis for the topology. It follows that the condition (b) is equivalent to the condition that for any S-words W,W' there is an S-word of the form WUW'. (This tells us that some point in a cylinder set defined by W will visit a cylinder set defined by W'.)

For a SFT S_A, words correspond to paths in the graph with adjacency matrix A. It is routine to check that A is irreducible if and only if there is a path in the graph from any vertex to any other, i.e., if and only if for any two S-words W,W' there is an S-word of the form WUW'.

3. Mixing

The basic fact: S_A is mixing if and only if A is primitive ("only if" relies on our assumption of nondegeneracy of A). We explain.

A system (X,T) is *mixing* if for any two nonempty open sets U,V the set $\{n\in\mathbb{N} : (T^{-n}U)\cap V=\emptyset\}$ is finite. For a subshift T, this condition will hold if for any two words W,W', for all but finitely many positive integers n, there is a T-word WUW' such that U is a word of length n. Now it is not hard to verify from the graph interpretation that this is equivalent to A being primitive.

The most important class to understand is the class of mixing SFT's. 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.)

4. Entropy

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

$$h(S) = \limsup_n \ (1/n)\ \log\ [\#W_n(S)]$$

where $W_n(S)$ is the set of words of length n occuring 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 S_A,

$$h(S) = \ \log\ (\lambda_A)$$

where λ_A is the spectral radius of A. This follows from the spectral radius theorem because the number of words of length n is the sum of the entries of A^n, which is the L^1 norm of A^n.

5. Periodic Points

The basic fact here is that the periodic point counts for S_A are determined by the sequence $\mathrm{tr}(A^n)$, $n \in \mathbb{N}$.

Let $\pi_n(S)$ be the number of points of least period n in S. The sequence $\pi_n(S)$ contains all the information one has from restricting S to its periodic points and forgetting the topology.

Let $t_n(S)$ be the number of fixed points of S^n. One can recover the sequence $\pi_n(S)$ from the sequence $t_n(S)$. The sequence $t_n(S)$ is less directly dynamical but it is much better suited to finite summary and algebra.

In a subshift S, a fixed point is a sequence with the same symbol in every coordinate. A fixed point of S^n is a sequence which is the infinite concatenation of some word of length n. In particular, if the alphabet of S consists of M symbols, then S has at most M^k points of period n.

A fixed point x of S^n is determined by the word $x_0 \dots x_{n-1}$ (which is repeated forever to form x). For the edge shift S_A, such a word $x_0 \dots x_{n-1}$ is a path of length n (i.e., a path of n edges) which begins and ends at the same vertex. Since $A^n(i,i)$ is the number of paths of length n which begin and end at vertex i, we have

$$t_n(S_A) = \sum_i A^n(i,i) = \ \mathrm{tr}\ (A^n)\ .$$

6. Zeta function

We want a good way to present the periodic point data given by the sequence $t_n(S)$. The favored choice in dynamics for this is the (Artin-Mazur) zeta function of S,

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

For $S = S_A$,

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

where the product is over the eigenvalues λ of A, repeated according to multiplicity. (Of course, in this product only the nonzero eigenvalues matter.) So, the inverse zeta function of an SFT is a polynomial with integral coefficients and constant term 1.

For examples, we give some matrices and the corresponding zeta functions:

$$A = (\,3\,) \qquad B = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}$$

$$\zeta_{S_A}(z) = \frac{1}{1-3z} \qquad \zeta_{S_B}(z) = \frac{1}{(1-3z)^2} \qquad \zeta_{S_C}(z) = \frac{1}{1-2z-15} = \frac{1}{(1-5z)(1+3z)}$$

PROOF of (*)
The first equality follows from $t_n(S_A) =$ tr (A^n) .
The last equality follows from dividing the factorization of the characteristic polynomial

$$\det(zI - A) = \prod_\lambda (z - \lambda)$$

by z^k (where A is k by k) and then replacing $1/z$ with z. This equality is saying that the inverse zeta function is the characteristic polynomial "written backwards":

$$\begin{array}{llllll} \text{if} & \det(zI - A) = z^k & + a_{k-1}z^{k-1} & + a_{k-2}z^{k-2} & + \ldots + a_j z^j , \\ \text{then} & \det(I - zA) = 1 & + a_{k-1}z & + a_{k-2}z^2 & + \ldots + a_j z^{k-j} . \end{array}$$

The second equality of $(*)$ follows from three facts:
(1) $tr(A^n) = \sum_\lambda \lambda^n$,
(2) for any complex number λ,

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

(to see this take the derivative of the log of each side)
(3) $\exp(\sum_{n=1}^{\infty} \sum_\lambda (1/n)(\lambda z)^n) = \prod_\lambda \exp(\sum_{n=1}^{\infty} (1/n)(\lambda z)^n)$. QED

7. Isomorphism

The basic, fundamental result, due to Williams, is that S_A and S_B are isomorphic if and only if the matrices A and B are strong shift equivalent. This is completely general, nondegeneracy of A and B need not be assumed.

Matrices A and B are elementary strong shift equivalent if there are matrices U, V with entries from $\mathbb{Z}_+$ such that $A = UV$ and $B = VU$. The matrices U, V need not be square. A and B are strong shift equivalent (SSE) 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.

It takes some time to prove that isomorphism of S_A and S_B implies A and B are strong shift equivalent. However the other direction is easy. It suffices to show S_A and S_B are isomorphic under the assumption $A = UV$, $B = VU$.

Let G_A, G_B be the directed graphs with adjacency matrices A, B. We can name the vertices so that these graphs have no vertices in common. We can also view U as the adjacency matrix for a set of edges with initial vertices in G_A and terminal vertices in G_B: now $U(i,k)$ denotes the number of edges drawn from the A-vertex i to the B-vertex k. Similarly let $V(k,j)$ is the number of edges from the B-vertex k to the A-vertex j.

Let lower case letters a, b, u, v represent arcs corresponding to A, B, U, V. We have for any A-vertices i and j,

$$A(i,j) = (UV)(i,j) = \sum_k U(i,k)V(k,j) .$$

This equation has a graphical interpretation. The left hand side $A(i,j)$ is the number of A-edges from i to j. The right hand side is the number of UV paths from i to j.

If follows from this equation (and the other equation $B = VU$) that we may choose bijections of arcs and paths of length 2,

$$\alpha : \{a\} \longleftrightarrow \{uv\}, \qquad \beta : \{b\} \longleftrightarrow \{vu\}$$

which respect initial and terminal vertex.

Now we view a point of S_A as an infinite path $...a_{-1}a_0a_1...$ of edges a_i, and we view a point of S_B as an infinite path $...b_{-1}b_0b_1...$ of edges b_i. Using the chosen bijections above, we can induce bijections of points as follows.

$$\begin{aligned} \dots a_0a_1a_2a_3\dots \longleftrightarrow & \dots [u_0v_0][u_1v_1][u_2v_2][u_3v_3]\dots \\ \longleftrightarrow & \quad \dots [v_0u_1][v_1u_2][v_2u_3][v_3u_4]\dots \longleftrightarrow \dots b_0b_1b_2b_3\dots \end{aligned}$$

Here the left bijection comes from applying α ($a_i \longleftrightarrow u_iv_i$), the middle bijection is the given regrouping, and the right bijection comes from applying β^{-1} ($v_iu_{i+1} \longleftrightarrow b_i$). This bijection of points is given by a block code, so it is an isomorphism of S_A and S_B.

It is a remarkable fact that ANY isomorphism of edge SFTs can be expressed as a composition of maps of exactly the form above, followed by a power of the shift.

Strong shift equivalence is easy to define but, it turns out, very difficult to understand completely. This led Williams to introduce another matrix relation, shift equivalence, which we consider next.

8. Eventual isomorphism

Two systems S and T are *eventually isomorphic* if S^n and T^n are isomorphic for all but finitely many n.

Nonnegative integral matrices A and B are *shift equivalent* if there are nonnegative integral matrices U, V and a positive integer ℓ (called the lag) such that the following equations hold.

$$\begin{aligned} A^\ell &= UV & \qquad B^\ell &= VU \\ AU &= UB & BV &= VA\ . \end{aligned}$$

Given the equations above, we see that A^ℓ and B^ℓ are strong shift equivalent. Moreover, so are higher powers, since for any positive integer n,

$$U(VA^n) = A^{\ell+n} \qquad \text{and} \qquad (VA^n)U = (B^nV)U = B^{\ell+n}\ .$$

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. We will skip that proof.

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. Williams introduced the idea of shift equivalence with the intent of reducing strong shift equivalence to a manageable equivalence relation on matrices. He conjectured that shift equivalence implies strong shift equivalence for matrices over $\mathbb{Z}_+$. After many years, this conjecture was finally refuted by Kim and Roush in the reducible case, and very recently in the irreducible case (again by Kim and Roush). We will say a little about how this was done in our fourth lecture.

The crucial irreducible case still remains open for many classes. For example, it is open in the special case where the matrices under consideration have just one nonzero eigenvalue, and it is open in the case that both matrices are two by two, with determinant less than -1.

9. Flow equivalence

Two homeomorphisms are flow equivalent if they are cross sections of a common flow. If C is a square (say, $n \times n$) integral matrix, then by its cokernel $\mathrm{cok}(C)$ we mean the finitely generated abelian group $\mathbb{Z}^n/C\mathbb{Z}^n$.

If A and B are irreducible and neither is a permutation matrix, then S_A and S_B are flow equivalent if and only if

$$\begin{aligned}&\text{(i) } \det(I-A) = \det(I-B) \quad \text{and}\\ &\text{(ii) } \mathrm{cok}(I-A) \cong \mathrm{cok}(I-B).\end{aligned}$$

Here, the only information in (i) which is not determined by (ii) is the sign of the determinant. This is because for a square integral matrix C, the cardinality of $\mathrm{cok}(C)$ is $|\det C|$ if C is nonsingular, and it is infinite if C is singular.

The characterization above is due to Franks, following earlier work of Bowen-Franks and Parry-Sullivan. We do not have time for a proof. For examples, consider the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \qquad C = (\,5\,) \qquad D = \begin{pmatrix} 5 & 2 \\ 2 & 3 \end{pmatrix} \qquad E = (\,25\,)\,.$$

It is easy to compute

$$\begin{aligned}
\det(I-A) &= -4 \quad \text{and} \quad \mathrm{cok}(I-A) \cong \mathbb{Z}/4\mathbb{Z}\\
\det(I-B) &= -4 \quad \text{and} \quad \mathrm{cok}(I-B) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}\\
\det(I-C) &= -4 \quad \text{and} \quad \mathrm{cok}(I-C) \cong \mathbb{Z}/4\mathbb{Z}\\
\det(I-D) &= 4 \quad \text{and} \quad \mathrm{cok}(I-D) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}\\
\det(I-E) &= -24 \quad \text{and} \quad \mathrm{cok}(I-E) \cong \mathbb{Z}/24\mathbb{Z}\quad .
\end{aligned}$$

Thus S_A and S_C are flow equivalent and no other pair here is. Note, the SFTs S_A and S_B have the same zeta function, and the SFTs S_A and S_C have different entropies. Also, S_E is isomorphic to the square of the S_C.

The classification of general (possibly reducible) SFTs up to flow equivalence was completed by D. Huang, but the algebraic invariants are too complicated for us to describe here.

10. Relations

For matrices over $\mathbb{Z}_+$: SSE $\Rightarrow$ SE $\Rightarrow$ same zeta function $\Rightarrow$ same entropy.

None of these implications can be reversed.

It also turns out that SE $\Rightarrow$ FE; in the reducible case, this is a difficult result. As we can see from the examples above, the zeta function does not determine the flow equivalence class, and (of course) the flow equivalence class does not determine even the entropy.

III. DIMENSION GROUPS AND SHIFT EQUIVALENCE

1. Dimension groups

1.1 Ordered groups

An *ordered group* is a pair (G, G_+) such that G is a countable abelian group and G_+ is a subset of G (the "positive set") containing 0 such that

- $G_+ + G_+ = G_+$ [G_+ is a monoid]
- $G_+ - G_+ = G$ [G_+ generates]
- $G_+ \cap -G_+ = \{0\}$ [((G, G_+) is directed]

Example. Let $G = \mathbb{Z}^n$ with the usual positive set $G_+ = \mathbb{Z}^n_+ = \{v \in \mathbb{Z}^n : v \geq 0\}$. Such a group is called *simplicial.*

1.2 Direct limit groups

Suppose $G_1, G_2, \dots$ is a sequence of groups and for $n \geq 1$ we are given homomorphisms $\phi_n : G_n \to G_{n+1}$. We call the homomorphisms ϕ_n *bonding maps* and picture the information as the diagram

$$G_1 \longrightarrow G_2 \longrightarrow G_3 \longrightarrow G_4 \longrightarrow G_5 \cdots$$

where an arrow $G_n \longrightarrow G_{n+1}$ represents a bonding map ϕ_n. We will describe the construction of the direct limit group G from the given data.

We let ϕ_{ij} denote the composition $\phi_{j-1} \cdots \phi_{i+1}\phi_i$, which takes G_i to G_j.

Let X be the disjoint union of the G_n; concretely,

$$X = \bigcup_n \{(h, n) : h \in G_n,\ n \in \mathbb{N}\} .$$

Define a relation $\sim$ by declaring $[h, i] \sim [h', i']$ if for some j, $\phi_{ij}(h) = \phi_{i'j}(h')$. (In the picture, this means that h and h' are pushed to a common image at some point to the right.) It is easy to check that $\sim$ is an equivalence relation. As a set, G is the set of equivalence classes for $\sim$.

Note, if an equivalence class contains an element $[h,n]$ of X, then it also contains $[\phi_{nj}(h), j]$ for all $j > n$. We can use this to define addition in G.

Suppose g and g' are two equivalence classes in G. Choose some positive integer n such that g contains some element $[h,n]$ and g' contains some element $[h',n]$. Define $g+g'$ to be the equivalence class containing $[h+h',n]$. It is straightforward to check that $g+g'$ is independent of the particular choice of representatives, and with this operation G is a group.

If each of the groups G_n is actually an ordered group, and each of the bonding maps is order-preserving (maps the positive set of G_n into the positive set of G_{n+1}), then we can define the the direct limit order on G. Here an equivalence class is in the positive set G_+ if it the equivalence class contains an element $[h,n]$ such that $h \in (G_n)_+$. It is easy to verify that (G, G_+) is now an ordered group.

1.3 Dimension groups

A dimension group is an ordered group which is isomorphic to a direct limit of simplicial groups. So, up to isomorphism, a dimension group is presented by a sequence of homomorphisms of simplicial groups. Such a homomorphism can be presented as matrix multiplication, for example

$$\mathbb{Z}^3 \to \mathbb{Z}^2$$
$$\begin{pmatrix} x & y & z \end{pmatrix} \mapsto \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{pmatrix} .$$

(We will have our matrices act from the right, on rows, to read composition from left to right.) The condition that the homomorphism of groups be order preserving is exactly the condition that the presenting integral matrix have nonnegative entries.

So: up to isomorphism, a dimension group is presented by a sequence $A_1, A_2, A_3, \dots$ of nonnegative integral matrices.

The dimension groups also have an abstract, useful characterization due to Effros, Handelman and Shen: an ordered group is a dimension group if and only if it satisfies the following two properties:

Riesz interpolation:	for any u_1, u_2, v_1, v_2 from G with all $u_i \le v_j$, there exists w in G with $u_i \le w \le v_j$ for all i, j.
Unperforation:	if $n \in \mathbb{N}$ and $g \in G$ with $ng \in G_+$, then $g \in G_+$.

These two properties are what must survive from the lattice ordering of a simplicial group on passage to a direct limit. Notice an ordered group with unperforation must be torsion free. Sometimes unperforation is built into the definition of an ordered group.

The EHS characterization provides a way of recognizing dimension groups independent of matrix constructions. For example, it is trivial to check from the EHS conditions that the group $\mathbb{Q} \oplus \mathbb{Q}$ with positive set $\{(x,y) : x > 0, y > 0\}$ is a dimension group; but constructing a matrix sequence here takes a little thought.

1.4. Uses for dimension groups

The dimension groups come to us from C^*-algebras, where (among other important uses) they are used to classify the important class of AF (approximately finite) C^*-algebras. The dimension groups appear as the K_0 groups of these algebras. (The term "dimension group" can be rationalized somewhat by the view that K_0 is a generalization of vector space dimension to more complicated modules.) The old exposition of Effros is still a nice introduction (E. Effros. Dimensions and C^*-algebras. AMS-CBMS Reg. Conf. 46 (1981)).

Dimension groups have played a significant role in the study of SFT's; we will see one example of this in the fourth lecture.

The dimension groups have also played a dramatic role in recent study of topological orbit equivalence of minimal homeomorphisms of the Cantor set, and in the study of substitution systems. Here, if X is the Cantor set and T the minimal homeomorphism, then the dimension group (G, G_+) is given by

$$\begin{aligned} G &= C(X, \mathbb{Z})/B(X, \mathbb{Z}) \\ G_+ &= C(X, \mathbb{Z}_+)/B(X, \mathbb{Z}) \end{aligned}$$

The dramatic results here are beyond the scope of this lecture and I mention their existence for context. (See the lectures of Host for references.) Two remarks are worth making.

First: the order structure of the groups is crucial to their use in the minimal setting. (This is much less true in the SFT setting we are considering now.) One should view the order not as an unwelcome guest to the algebra but as a powerful part of it.

Second: it is not only the definition of the groups which the C^* theory contributes, the groups come to dynamics along with with potent suggestions for their use.

2. Dimension modules

2.1 Stationary dimension groups

A dimension group is *stationary* if it is isomorphic to one which can be defined by a sequence of matrices $A_1, A_2, \ldots$ which are all equal. If this common matrix is A, then we can call the dimension group the dimension group of A.

There is a very concrete way to visualize this group in terms of linear transformations, which we describe now.

Let A be an $n \times n$ integral matrix. Then A defines a linear transformation on $\mathbb{Q}^n$ by matrix multiplication, $v \mapsto vA^n$. The *eventual image* of A is the vector subspace of elements which are in all forward images of A,

$$V_A = \bigcap_{k>0} \{vA^k : v \in \mathbb{Q}^n\} .$$

Because $\mathbb{Q}^n \supset \mathbb{Q}^n A \supset \mathbb{Q}^n A^2 \supset \cdots \supset \mathbb{Q}^n A^n$, we have $\dim(\mathbb{Q}^n) \geq \dim(\mathbb{Q}^n A) \geq \cdots$. As soon as $\dim(\mathbb{Q}^n A^k) = \dim(\mathbb{Q}^n A^{k+1})$, the restriction of A to $\mathbb{Q}^n A^k$ must be an isomorphism of vector spaces, and $V_A = \mathbb{Q}^n A^k$. In particular, $V_A = \mathbb{Q}^n A^n$.

Now define $G_A = \{v \in V_A : \exists k > 0 \text{ such that } vA^k \in \mathbb{Z}^n\}$, so G_A is the subgroup of V_A which is eventually mapped by A into the integral lattice. Also define

$(G_A)_+ = \{v \in G_A : \exists k > 0 \text{ such that } vA^k \in \mathbb{Z}^n_+\}$.

It is not hard to check that $(G_A, (G_A)_+)$ is isomorphic to the dimension group of A. (Here is an explicit isomorphism from $G \to G_A$: if $g \in G$ and g contains an element $[v, k]$, then send g to $vA^n(A|V_A)^{-(n+k)}$.) It's worth noticing that the pair $(G_{A^n}, (G_{A^n})_+)$ is the same for any positive integer n.

Consider some examples.

Example. Suppose $A = (\,2\,)$. Then $V_A = \mathbb{Q}$ and G_A is the group of dyadic rationals, $\{m2^{-k} : m, k \in \mathbb{Z}\}$.

Example. Suppose $B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Then $V_B = \{q(\,1 \quad 1\,) : q \in \mathbb{Q}\}$. The action of B on V_B is just multiplication by 2, so an element $q(\,1 \quad 1\,)$ of V_B is eventually mapped into $\mathbb{Z}^2$ only if q is an integer divided by a power of 2, i.e., q is a dyadic rational. Thus G_B is isomorphic to G_A of the previous example.

Example. Suppose A is $n \times n$ and det(A)= +/-1. Then $G_A \cong \mathbb{Z}^n$.

2.2 The dimension module of A

Again suppose A is square nonnegative integral. It is an easy exercise to check that multiplication by A defines an automorphism ($\hat{A}$, say) of the ordered group $(G_A, (G_A)_+)$. For an example, think of multiplication by 2 on the dyadic rationals.

There is a more algebraic way of formulating the information given by the "dimension pair" $\{(G_A, (G_A)_+)\ ,\ \hat{A}\}$.

First we use $\hat{A}$ to make G_A an L-module, where L denotes the Laurent polynomials in one variable, $L = \mathbb{Z}[t, t^{-1}]$. Here the action of L on G_A is induced by requiring t^{-1} to act as $\hat{A}$.

Second we make L an ordered ring by declaring $\mathbb{Z}_+[t, t^{-1}]$ to be its positive set L_+. Then L_+ sends $(G_A)_+$ to itself and our L-module is an ordered module.

We call this ordered L-module the *dimension module* of A. Fortunately, in the most important SFT case we will be able to forget the order structure here.

We remark that a matrix A and its transpose can define nonisomorphic dimension modules (even disregarding the order). In other words, the modules defined by way of the action of A on row vectors vs. column vectors can be nonisomorphic.

3. Shift equivalence

3.1 The relationship

Recall that square nonnegative integral matrices A, B are shift equivalent if there exist nonnegative integral matrices U, V and a positive integer ℓ such that

$$\begin{aligned} A^\ell &= UV & \qquad B^\ell &= VU \\ AU &= UB & \qquad BV &= VA\ . \end{aligned}$$

The basic fact is that A and B are shift equivalent if and only if their dimension modules are isomorphic as ordered modules. That is, there is a group isomorphism from G_A to G_B which sends the positive set $(G_A)_+$ onto $(G_B)_+$ and which intertwines the isomorphisms $\hat{A}$ and $\hat{B}$.

The proof of this fact is routine. Given a shift equivalence as above, the required isomorphism from G_A to G_B is given by matrix multiplication by the matrix U. In particular: if $A = B$, then multiplication by U defines an automorphism of the dimension module of A. This fact will be important in the next lecture.

For example, with $A = (\,2\,), \ell = 5, U = (\,8\,)$ and $V = (\,32\,)$, the matrix U induces an automorphism on the dyadic rationals G_A which is simply multiplication by 8.

The connection between shift equivalence and dimension groups was made by Krieger. There are deeper aspects to this connection which we will not go into.

3.2. The primitive case

We say that A and B are algebraically shift equivalent over $\mathbb{Z}$ if there are (not necessarily nonnegative) matrices U, V which satisfy the shift equivalence equations for A and B. Similarly we can define algebraic strong shift equivalence over $\mathbb{Z}$.

FACT (Parry-Williams): if A and B are primitive, then they are shift equivalent if and only if they are algebraically shift equivalent.

PROOF

By the Perron-Frobenius Theorem, the spectral radius λ of A is an eigenvalue of A, for which there are positive left and right eigenvectors, say x and y. We choose these so that $xy = 1$. By the Perron-Frobenius Theorem, λ is a simple root of the the characteristic polynomial of A, and λ is strictly greater than the modulus of any other root. It follows that the sequence $[(1/\lambda)A]^n$ converges to the matrix which fixes the dominant eigendirection and annihilates the complementary generalized eigenspaces. This matrix is the positive matrix yx.

Now we prove the nontrivial direction. Suppose a pair of integral matrices U, V satisfy the shift equivalence equations for A and B. Shift equivalent matrices have the same nonzero eigenvalues, so A and B have the same spectral radius λ. The vectors xU and Vy are eigenvectors of B with eigenvalue λ, so each is strictly positive or strictly negative. If one is strictly negative, then so is the other, since $xUVy = xA^\ell y > 0$; so, if necessary after replacing U and V with $-U$ and $-V$, we can assume $xU > 0$ and $Vy > 0$. Since

$$\lim_n [(1/\lambda)A]^n U = yxU > 0 \ ,$$

it follows that $A^n U$ is strictly positive for large n. Similarly VA^n is strictly positive for large n. The pair $A^n U, VA^n$ give a lag $\ell + 2n$ shift equivalence of A and B. QED

The dimension formulation of the Parry-Williams result is that the dimension modules of A and B are isomorphic as ordered modules if and only if they are isomorphic as unordered modules. So, in the most important case (the case of mixing SFT's), we have the great luxury of ignoring the order.

Although the relationship of SE and SSE over $\mathbb{Z}_+$ is mysterious, the relationship of SE and SSE over $\mathbb{Z}$ is not: over $\mathbb{Z}$, shift equivalence and strong shift equivalence define the same equivalence relation. From a lag ℓ SE over $\mathbb{Z}$, it is possible to extract a lag ℓ SSE over $\mathbb{Z}$ (Williams).

3.3 Ideal classes

We give a useful example of algebraic structure for SE.

Let p be an irreducible monic polynomial with integral coefficients and let λ be a root of p. Consider the class $\mathcal{M}_\lambda$ of square integral matrices with characteristic polynomial p.

Then, there is a bijection between algebraic shift equivalence classes of matrices in $\mathcal{M}_\lambda$, and ideal classes in the ring $\mathbb{Z}[1/\lambda]$. (This is a part of the classification of mixing SFT's up to regular isomorphism.)

This is an analogue of a classical result of Taussky-Todd: there is a bijection between similarity-over-$\mathbb{Z}$ classes of matrices in $\mathcal{M}_\lambda$, and ideal classes of the ring $\mathbb{Z}[\lambda]$. In both cases one checks rather easily that a bijection is given by the map which sends a matrix to the ideal class of the ideal generated by entries of an eigenvector for λ which has entries in $\mathbb{Z}[\lambda]$.

The class numbers of the rings $\mathbb{Z}[\lambda]$ and $\mathbb{Z}[1/\lambda]$ are finite and the relations under consideration are computable.

4. Further developments

4.1 Polynomial matrices

The presentation of an SFT as a vertex shift allows one to extract algebraic invariants from a defining zero-one matrix. Defining edge shifts with nonnegative integral matrices, one makes a significant advance in conciseness and functoriality of presentation.

There is another advance in conciseness and functoriality gained by presenting SFT's with polynomial matrices. Here is an example illustrating the general situation. Let A be the polynomial matrix

$$A = \begin{pmatrix} 2t & t^3 + t \\ 3t^2 & 0 \end{pmatrix} .$$

To this 2×2 matrix, we associate a graph G with two distinguished vertices, say 1 and 2. A term t^k in A corresponds to a path of length k in G. From the term $2t$, G acquires two edges from 1 to 1. From the term t^3, G acquires a path of three edges from 1 to 2. On this path are two new, intermediate vertices which have no further adjoining edges. Similarly G acquires an edge from 1 to 2 and three paths of length two from 2 to 1.

Clearly we can describe some very complicated graphs with polynomial matrices of small size. Also, the algebra behaves beautifully with this presentation. Let S be the edge shift for the graph constructed as above from A. Then the inverse zeta function is $\det(I - A)$.

Moreover, if A is $n \times n$, then by matrix multiplication A maps L^n into itself (where $L = Z[t, t^{-1}]$). Let

$$\operatorname{cok}(I - A) = L^n/(L^n(I - A)) .$$

This is an L-module and (suitably ordered) it is isomorphic to the dimension module for the SFT S. If we set t equal to 1, then L becomes $\mathbb{Z}$ and our dimension module becomes the Bowen-Franks invariant of flow equivalence. For a more thorough discussion of all this, see the survey [B].

4.2 SE and SSE for sofic systems

The utility of shift equivalence and strong shift equivalence does not end with SFT's. For at least two other classes of symbolic systems, one can find suitable ordered rings in which the equations of SSE and SE correspond to isomorphism and eventual isomorphism in other cases.

These classes are the irreducible Markov shifts (meaning: irreducible SFT's with Markov measure) and the sofic systems. In the Markov case, the role played (in the SFT case) by integral matrices is played by matrices whose entries are Laurent polynomials in several variables. In the sofic case, one works in a certain complicated integral semigroup ring and the objects which play the role of matrices are derived from canonical covers of the sofic systems by SFTs.

IV. AUTOMORPHISMS AND CLASSIFICATION OF SFTS

1. Automorphisms

1.1 Automorphisms

An *automorphism* of a dynamical system (X, S) is a homeomorphism $U : X \to X$ such that $US = SU$. So, an automorphism is a self-conjugacy or symmetry of the dynamical system. With the operation of composition, the automorphisms of S are a group, $\mathrm{Aut}(S)$. In this lecture we will always restrict to systems which are nontrivial mixing SFTs (nontrivial means not just a fixed point).

It is difficult to get hold of structure for SFTs, and automorphisms provide one opportunity. How big is the group? How does it act? When are $\mathrm{Aut}(S)$ and $\mathrm{Aut}(T)$ isomorphic groups?

1.2 Aut(S) as an abstract group

It is countable (there are only countably many block codes), but very complicated (containing for example copies of all finite groups and many others).

It is residually finite (this is because the periodic points are dense and for each n the points of period n are a finite invariant set).

The center of $\mathrm{Aut}(S)$ equals the powers of S (Ryan's theorem). Thus, for example, the 2-shift $S_{[2]}$ and the 4-shift $S_{[4]}$ have nonisomorphic automorphism groups. The center of $\mathrm{Aut}(S_{[4]})$ has a generator which is a square in the group $\mathrm{Aut}(S_{[4]})$ (because $S_{[4]} = (S_{[2]})^2$). But the center of $\mathrm{Aut}(S_{[2]})$ does not have a generator which is a square in $\mathrm{Aut}(S_{[2]})$ (it is not hard to show $S_{[2]}$ has no square root).

One measure of the mystery of $\mathrm{Aut}(S)$ is the following: we know NO way to show $\mathrm{Aut}(S)$ and $\mathrm{Aut}(T)$ are nonisomorphic except by some straightforward application of

Ryan's Theorem. For example, we do not know whether $\mathrm{Aut}(S_{[2]})$ and $\mathrm{Aut}(S_{[3]})$ are isomorphic as abstract groups.

1.3 Action on finite subsystems

Here is a seminal problem posed by Williams in the 1970's (in a 1977 lecture in Boulder and perhaps earlier). Suppose S has fixed points x and y. Does there exist an automorphism of S which exchanges them? Williams proposed this as one approach to the classification problem: perhaps different answers to this question would distinguish SFTs defined by shift equivalent matrices.

Well, there's no reason to stop at fixed points. Let $F_n[S]$ denote the (finite) set of points of period at most n. Let $S_{<n>}$ be the restriction of S to $F_n[S]$. Now $S_{<n>}$ has its own little group of automorphisms, and we can ask: when does an automorphism of $S_{<n>}$ extend to an automorphism of S? Always?

For possible obstructions, we look for useful representations of the automorphism group. By a representation, here we just mean a homomorphism into a group. The only useful representations found so far involve the dimension module or periodic points.

2. Representations

2.1 The dimension representation ρ

Suppose ϕ is an automorphism of S_A. Recall that ϕ is given by a power of the shift composed with a map constructed from a strong shift equivalence from A to A in the explicit fashion we earlier described. (That is, recall we have already claimed this without proof.) Now, if we have a strong shift equivalence from A to itself, say

$$A = U_1V_1,\ V_1U_1 = A_1 \qquad A_1 = U_2V_2,\ V_2U_2 = A_2 \qquad \dots \qquad A_\ell = U_\ell V_\ell, V_\ell U_\ell = A$$

then we get a shift equivalence

$$\begin{array}{ll} A^\ell = UV & B^\ell = VU \\ AU = UB & BV = VA \end{array}$$

by setting $U = U_1U_2\dots U_\ell$ and $V = V_\ell \dots V_2V_1$.

Now, this shift equivalence defines an isomorphism of the dimension module of S_A (from multiplication by U) and it turns out that this isomorphism, composed with the isomorphism induced by that power of the shift, is canonical: it depends only on the automorphism ϕ.

Thus we have a homomorphism ρ from $\mathrm{Aut}(S_A)$ to $\mathrm{Aut}(M_A)$, where $\mathrm{Aut}(M_A)$ is the group of automorphisms of the dimension module. This homomorphism ρ is called the *dimension representation.*

2.2 The kernel of ρ

The group of automorphisms of the dimension module is algebraic, "small" (typically a finitely generated abelian group), and varies from one SFT to another.

Thus most of the vast complexity of $\mathrm{Aut}(S)$ lies in the kernel of ρ. Think of this group as combinatorial, big and somehow homogeneous in its properties across nontrivial mixing SFTs. (This "homogeneity" is not a precise property, but a feeling one gets from experience with the actions of $\mathrm{Ker}(\rho)$.)

In particular the richness of action of $\mathrm{Aut}(S)$ on finite subsystems comes almost entirely from the action of $\mathrm{Ker}(\rho)$. Constraining the action of $\mathrm{Aut}(S)$ is mostly a matter of finding general constraints on the action of $\mathrm{Ker}(\rho)$; then it is a matter of fine structure to work out the remaining possibilities for a particular SFT.

2.3 Sign

Now we need some algebraic invariants for the action of $\mathrm{Aut}(S)$ on periodic points. Suppose U is an automorphism of S. Let $P_n^o(S)$ be the set of points of period exactly n for S. This finite set is invariant under U.

U permutes the S-orbits of length n (here length means cardinality). If this permutation is even, then define $\mathrm{sign}_n(U) = 0$. If the permutation is odd, then set $\mathrm{sign}_n(U) = 1$. This gives us a homomorphism

$$\begin{aligned} \mathrm{sign}_n : \ \mathrm{Aut}(S) &\to \mathbb{Z}/2\mathbb{Z} \\ U &\mapsto \ \mathrm{sign}_n(U). \end{aligned}$$

(If there are no S-orbits of size n, then we adopt the convention $\mathrm{sign}_n = 0$.) Since the shift S maps each S-orbit to itself, $\mathrm{sign}_n(S) = 0$ for every n.

2.4 Gyration

We will now use the action on $P_n^o(S)$ to define a homomorphism from $\mathrm{Aut}(S)$ into $\mathbb{Z}/n\mathbb{Z}$. From each S-orbit of length n pick a single point, giving us say k points $x_1, x_2, \dots x_k$. An automorphism U will send x_i into the orbit of some x_j, say $j = \pi(i)$, so for some integer $m(i)$,

$$U : x_i \mapsto S^{m(i)} x_{\pi(i)} \ .$$

We define

$$\begin{aligned} \mathrm{gy}_U : \mathrm{Aut}(S) &\to \mathbb{Z}/n\mathbb{Z} \\ U &\mapsto \sum_i m(i). \end{aligned}$$

It is an easy and pleasant exercise to check that this map is well defined (does not depend on the choice of representatives x_i), and is a group homomorphism.

2.5 SGCC

Now we can define the "sign-gyration-compatibility-condition" homomorphism SGCC. This will be a homomorphism

$$\begin{aligned} \mathrm{SGCC} : \mathrm{Aut}(S) &\to \oplus_n \mathbb{Z}/n\mathbb{Z} \\ U &\mapsto \oplus_n \mathrm{SGCC}_n \end{aligned}$$

where

$$\begin{aligned} \mathrm{SGCC}_n : \mathrm{Aut}(S) &\to \mathbb{Z}/n\mathbb{Z} \\ U &\mapsto \mathrm{gy}_n(U) + \frac{n}{2} \sum_{(n/k)=2^j} \mathrm{sign}_k(U) \ . \end{aligned}$$

The last sum is over the positive integers k such that $k < n$ and n/k is a positive integer which is a power of 2. For example, if $n = 12$, then the sum is over $k = 6, 3$. An empty sum is zero.

We will see in a moment that under some conditions SGCC must vanish. This is an astonishing constraint. We do not have time to motivate where in the world it could come from. But, do notice, for n even the condition $\mathrm{SGCC}_n(U)=0$ implies that the action of U on orbits of length n is constrained by its action on orbits of smaller length.

3. The KRW Factorization Theorem

3.1 Statement of the theorem

Kim, Roush and Wagoner ("KRW") proved that there is a homomorphism

$$\mathrm{sgcc}_n : \ \mathrm{Aut}(M_A) \to \oplus_n \mathbb{Z}/n\mathbb{Z},$$

such that $\mathrm{SGCC}_n = \mathrm{sgcc}_n \circ \rho$. Taking the direct sum over n, we get $\mathrm{SGCC} = \mathrm{sgcc} \circ \rho$. In other words, SGCC factors through the dimension representation. Especially, SGCC vanishes on the kernel of the dimension representation.

3.2 Nonsurjectivity of ρ

The KRW maps sgcc_n can be computed (with difficulty: for $n = 2$, KRW computations of examples were computer assisted). With the computability, KRW could produce examples of mixing SFTs S for which ρ is not surjective.

The rough idea for nonsurjective examples is simple: produce (for example) an SFT S in which a certain automorphism ψ of the dimension module has nonzero image under sggc_2, but S has just one fixed point, and no points of period two. An automorphism ϕ of S must then satisfy $\mathrm{SGCC}_2(\phi) = 0$, so $\rho(\phi) = \psi$ would give

$$0 = \mathrm{SGCC}_2(\phi) = \mathrm{sgcc}_2(\rho(\phi)) = \mathrm{sgcc}_2(\psi) \neq 0$$

which is a contradiction.

3.3 A long story

The Factorization Theorem did not appear overnight, it was a culmination of a long development also involving Hedlund, Williams, Boyle, Krieger, U. Fiebig, Nasu, Lind and Rudolph. (See the references in [LM].) There is also another dramatic recent development on the constructive side in which KRW use innovative constructions with polynomial matrices to show that $\mathrm{SGCC} = 0$ is the *only* constraint on the action of $\mathrm{Ker}(\rho)$ on finite subsystems. (A result of Boyle and Krieger then implies that $\mathrm{SGCC} = 0$ remains the only constraint on the action of $\mathrm{Ker}(\rho)$ on infinite subsystems.)

4. Classification

4.1 SE does not imply SSE: the reducible case

Kim and Roush were able to use one of the nonsurjectivity examples above to get a counterexample to the shift equivalence conjecture in the reducible case. Here is the rough idea.

Let A be a primitive matrix for which ρ_A is not surjective and consider a pair of reducible matrices B, C with the block forms

$$B = \begin{pmatrix} A & X \\ 0 & A \end{pmatrix} \quad , \quad C = \begin{pmatrix} A & Y \\ 0 & A \end{pmatrix} .$$

As we have seen, a strong shift equivalence from B to C induces a shift equivalence from B to C. Also, the SSE from B to C restricts to give SSE's of the upper left blocks A and of the lower right blocks A. These two SSE's produce isomorphisms from S_A to S_A, which we can interpret as automorphisms.

The action of these automorphisms comes from SE's obtained by restriction to diagonal blocks of the SE from B to C (which was induced by the SSE from B to C). The matrices involved in the diagonal block SE's are not independent, they satisfy some additional matrix equations arising from the blocks X and Y.

Kim and Roush got their counterexample by making very judicious choices of A, X and Y to arrange the following:

(i) For S_A, the dimension representation is not surjective;

(ii) B and C are shift equivalent; and

(iii) for any shift equivalence of B and C, one of the induced diagonal-block shift equivalences from A to A must give an action on the dimension module of A which cannot arise from an isomorphism of A.

4.2 SE does not imply SSE: the irreducible case

After the original delivery of these lectures, Kim and Roush obtained a counterexample to the shift equivalence conjecture for irreducible matrices over $\mathbb{Z}_+$ [K.H.Kim and F.W.Roush, Williams conjecture is false for irreducible subshifts, Annals of Math. to appear]. The study of automorphisms of the shift, the dimension representation, and the Factorization Theorem again underly this work; but this work is even more deeply embedded in the developed theory and especially the CW complexes of strong shift equivalences developed over several years by Wagoner. These complexes provide the setting for the work of the Factorization Theorem and the irreducible counterexample of Kim and Roush, and appear to be fundamental to the study and classification of SFTs.