

Proceedings of Symposia in APPLIED MATHEMATICS

Volume 60

Symbolic Dynamics and its Applications

American Mathematical Society
Short Course
January 4–5, 2002
San Diego, California

Susan G. Williams
Editor



American Mathematical Society

0189.3-53
5986
200 2

Proceedings of Symposia in **APPLIED MATHEMATICS**

Volume 60

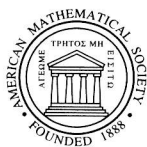
Symbolic Dynamics and its Applications

American Mathematical Society
Short Course
January 4-5, 2002
San Diego, California

Susan G. Williams
Editor



E200500027



American Mathematical Society
Providence, Rhode Island

Editorial Board

Peter S. Constantin (Chair) Eitan Tadmor Marsha J. Berger

LECTURE NOTES PREPARED FOR THE
AMERICAN MATHEMATICAL SOCIETY SHORT COURSE
SYMBOLIC DYNAMICS AND ITS APPLICATIONS
HELD IN SAN DIEGO, CALIFORNIA
JANUARY 4–5, 2002

The AMS Short Course Series is sponsored by the Society's Program Committee for National Meetings. The series is under the direction of the Short Course Subcommittee of the Program Committee for National Meetings.

2000 *Mathematics Subject Classification*. Primary 37B10, 37B50, 37A15, 37F45, 94B05, 19C99.

Library of Congress Cataloging-in-Publication Data

American Mathematical Society Short Course on Symbolic Dynamics and its Applications : (2002 : San Diego, Calif.)

Symbolic dynamics and its applications : American Mathematical Society, Short Course, January 4–5, 2002, San Diego, California / Susan G. Williams, editor.

p. cm. — (Proceedings of symposia in applied mathematics, ISSN 0160-7634 ; v. 60)

Includes bibliographical references and index.

ISBN 0-8218-3157-7 (alk. paper)

1. Symbolic dynamics—Congresses. I. Williams, Susan G., 1953– II. Title. III. Series.

QA614.85.A44 2002
514'.74—dc22

2003062891

Copying and reprinting. Material in this book may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to reprint-permission@ams.org.

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

© 2004 by the American Mathematical Society. All rights reserved.

The American Mathematical Society retains all rights
except those granted to the United States Government.

Printed in the United States of America.

⊗ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.

Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1 09 08 07 06 05 04

AMS SHORT COURSE LECTURE NOTES

Introductory Survey Lectures

published as a subseries of

Proceedings of Symposia in Applied Mathematics

Preface

The foundation of the field of symbolic dynamics is generally credited to Jacques Hadamard, who used infinite symbol sequences in his analysis of geodesic flow on negatively curved surfaces in 1898. Hadamard's symbolic techniques were soon adopted and extended by other authors. However, the field had to wait forty years for its christening by Marston Morse and Gustav Hedlund, who provided the first systematic study of symbolic dynamical systems as objects of interest in their own right. This rather prescient paper at the dawn of the computer age set the stage for the mathematical analysis of codes and finite-alphabet communication systems using the techniques of dynamics and ergodic theory, most notably in the pioneering work of C.E. Shannon on the mathematical theory of communication. Fifty years after Hadamard applied symbolic techniques to dynamics, Shannon and others were applying dynamical techniques to symbols.

In the fifty-odd years since then, symbolic dynamics has expanded its reach to apply, and be applied to, many areas. It has broken the confines of one dimension to encompass multi-dimensional arrays. The six chapters of this volume provide an introduction to the field as it is studied today and a sampler of its concerns and applications. They are expanded versions of the lectures given in the American Mathematical Society Short Course on Symbolic Dynamics and its Applications held in San Diego on January 4-5, 2002. I would like to take this opportunity to thank Jim Maxwell, Wayne Drady and the other AMS staff members who coordinated the short course and worked behind the scenes to make it run smoothly.

Susan G. Williams

Contents

Preface	vii
Introduction to symbolic dynamics SUSAN G. WILLIAMS	1
Combining modulation codes and error correcting codes BRIAN MARCUS	13
Complex dynamics and symbolic dynamics PAUL BLANCHARD, ROBERT L. DEVANEY, and LINDA KEEN	37
Multi-dimensional symbolic dynamics DOUGLAS LIND	61
Symbolic dynamics and tilings of \mathbb{R}^d E. ARTHUR ROBINSON, JR.	81
Strong shift equivalence theory J.B. WAGONER	121
Index	155

Introduction to Symbolic Dynamics

Susan G. Williams

ABSTRACT. We give an overview of the field of symbolic dynamics: its history, applications and basic definitions and examples.

1. Origins

The field of symbolic dynamics evolved as a tool for analyzing general dynamical systems by discretizing space. Imagine a point following some trajectory in a space. Partition the space into finitely many pieces, each labeled by a different symbol. We obtain a *symbolic trajectory* by writing down the sequence of symbols corresponding to the successive partition elements visited by the point in its orbit. We may ask: Does the symbolic trajectory completely determine the orbit? Can we find a simple description of the set of all possible symbolic trajectories? And, most important, can we learn anything about the dynamics of the system by scrutinizing its symbolic trajectories? The answers to these questions will depend not only on the nature of our dynamical system, but on the judicious choice of a partition.

Hadamard is generally credited with the first successful use of symbolic dynamics techniques in his analysis of geodesic flows on surfaces of negative curvature in 1898 [Ha]. Forty years later the subject received its first systematic study, and its name, in the foundational paper of Marston Morse and Gustav Hedlund [MH]. Here for the first time symbolic systems are treated in the abstract, as objects in their own right. This abstract study was motivated both by the intrinsic mathematical interest of symbolic systems and the need to better understand them in order to apply symbolic techniques to continuous systems. However, a further impetus was given by the emergence of information theory and the mathematical theory of communication pioneered by C.E. Shannon [Sh].

Symbolic dynamics has continued to find application to an ever-widening array of continuous systems: hyperbolic diffeomorphisms, maps of the interval, billiards, complex dynamics and more. At the same time it contributes to, and finds inspiration in, problems arising in the storage and transmission of data, as we will see in Brian Marcus's chapter. Computer simulations of continuous systems necessarily involve a discretization of space, and results of symbolic dynamics help us

1991 *Mathematics Subject Classification*. Primary 37B10.

Key words and phrases. Symbolic Dynamics.

The author was supported in part by NSF Grant #0071004.

understand how well, or how badly, the simulation may mimic the original. And symbolic dynamics *per se* has proved a bottomless source of beautiful mathematics and intriguing questions.

There are two excellent texts on symbolic dynamics. *An Introduction to Symbolic Dynamics and Coding*, by Douglas Lind and Brian Marcus [LM], has the more modest prerequisites (for example, no prior knowledge of topology or measure theory is assumed), while B. Kitchens's more compact *Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Shifts* [Ki] assumes basic first-year graduate mathematics. For the most part, we have followed the notation and terminology of [LM] in this survey. Also highly recommended is the collection [BMN] of survey articles from the 1997 Summer School on Symbolic Dynamics and its Applications in Frontera, Chile. The selection of topics is largely complementary to that of this short course.

2. Two simple examples

Consider the unit interval $I = [0, 1]$ and the map f that sends $x \in I$ to $\{2x\}$, the fractional part of $2x$. We are interested in the orbit $x, f(x), f^2(x) = f(f(x)), \dots$. If we wanted to trace this orbit on a computer screen we might begin by resolving the interval into 2^{10} pixels. However, we will content ourselves with a much cruder discretization of space: we will break I into just two parts, $I_0 = [0, \frac{1}{2})$ and $I_1 = [\frac{1}{2}, 1]$. We assign to x a symbolic trajectory $x_0x_1x_2\dots$ where x_i is 0 or 1 according as $f^i(x)$ is in I_0 or I_1 . A little consideration will show that the expression $.x_0x_1x_2\dots$ is simply a binary expansion of the number x . Hence x is completely determined by its symbolic trajectory. We see here an exchange of spatial information for time series information mediated by dynamics: We can recover the complexity of the continuum I from our crude 2-element partition, provided that we observe the evolution of the system for all time.

What symbolic trajectories will appear in this scheme? All binary sequences except those that end in $111\dots$. This awkward exception can be removed by working instead with closed intervals $I = [0, 1]$, $I_0 = [0, \frac{1}{2}]$ and $I_1 = [\frac{1}{2}, 1]$, and mapping sequences to points instead of the other way around. Beginning with a binary sequence $x_0x_1x_2\dots$, we can assign to it the unique point

$$x = \bigcap_{i=1}^{\infty} f^{-i}(I_{x_i})$$

that has that symbolic itinerary. Then, for example, $\frac{1}{2}$ will arise from two symbolic trajectories, corresponding to the two binary expansions $\frac{1}{2} = .1000\dots = .0111\dots$. It is common practice in the application of symbolic techniques to sacrifice strict one-to-one correspondence for a simpler description of the set of symbolic trajectories.

Our map f now has a very pleasant symbolic representation. If $x = .x_0x_1x_2\dots$ then $f(x) = \{2x\} = .x_1x_2x_3\dots$. We shift the symbolic sequence to the left and lop off the initial symbol. The key to the utility of symbolic dynamics is that the dynamics is given by a simple coordinate shift. Dynamic properties that might have seemed elusive in the original setting now become transparent. For instance, we can immediately identify the points of period 3 (that is, with $f^3(x) = x$). They are the eight points with repeating symbolic trajectories $x_0x_1x_2x_0x_1x_2\dots$. Note that the set of points with symbolic representation beginning with some fixed

initial string $x_0x_1\dots x_n$ is the dyadic interval $[k/2^{n+1}, (k+1)/2^{n+1}]$, where $k = x_0 \cdot 2^n + x_1 \cdot 2^{n-1} + \dots + x_n$. The orbit of a point is dense in I , visiting every interval no matter how small, if and only if its symbol sequence contains all possible finite strings of 0's and 1's.

As a variation on the first example, consider the map $g(x) = \{\gamma x\}$ on I , where $\gamma = (1 + \sqrt{5})/2$ is the golden mean. We let $I_0 = [0, \frac{1}{\gamma}]$ and $I_1 = [\frac{1}{\gamma}, 1]$. Since $\gamma = 1 + \frac{1}{\gamma}$, we have $g(I_0) = I$ and $g(I_1) = I_0$. A point that lands in I_1 under some iterate of g must move to I_0 at the next iteration. In fact, it is not hard to see that the set of symbolic trajectories is exactly the set of binary sequences that do not contain the string 11.

The symbolic trajectory $x_0x_1x_2\dots$ corresponds to a series expansion $x = x_0\gamma^{-1} + x_1\gamma^{-2} + x_2\gamma^{-3} + \dots$. Expansions of numbers with respect to a non-integer base β are called *beta expansions*. There is a very interesting literature relating dynamic properties of symbolic systems obtained by beta expansions to the number-theoretic properties of beta. The chapter by C. Frougny in [BMN] provides an up-to-date survey.

3. Full shifts and subshifts

We let \mathcal{A} denote a symbol set or *alphabet*, which for now we assume to be finite. The (two-sided) *full \mathcal{A} -shift* is the dynamical system consisting of the set of biinfinite symbol sequences, together with the *shift map* σ that shifts all coordinates to the left. More formally, our space is

$$\mathcal{A}^{\mathbb{Z}} = \{x = (x_i)_{i \in \mathbb{Z}} : x_i \in \mathcal{A} \text{ for all } i \in \mathbb{Z}\}$$

and the map $\sigma : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ satisfies $(\sigma x)_i = x_{i+1}$. If $\mathcal{A} = \{0, 1, \dots, n-1\}$ we call $\mathcal{A}^{\mathbb{Z}}$ the full n -shift.

The advantage of working with biinfinite sequences is that the shift map is invertible. However, we may also consider the one-sided \mathcal{A} -shift $\mathcal{A}^{\mathbb{N}}$, with the truncating shift map described in the previous section. These arise naturally as symbolic representations of noninvertible maps like the map $x \rightarrow \{2x\}$ on I . For simplicity we state most of our definitions for two-sided shifts; the one-sided analogue is generally clear.

We often think of an element of $\mathcal{A}^{\mathbb{Z}}$ as a time series, with x_0 representing the present location or state of our trajectory, $(x_i)_{i < 0}$ its past history and $(x_i)_{i > 0}$ its future. The action of the shift map is like a tick of the clock, moving us one step into the future.

We consider two points of $\mathcal{A}^{\mathbb{Z}}$ to be close to one another if they agree on a large central block $x_{-n} \dots x_n$ of coordinates. To be more concrete, we can define the distance between distinct points x and y to be $d(x, y) = 2^{-n}$ where n is the smallest integer with $x_{-n} \neq y_{-n}$ or $x_n \neq y_n$. This is a metric, and induces the product topology on $\mathcal{A}^{\mathbb{Z}}$. The map σ and its inverse are continuous: if x and y agree on their central $2n+1$ coordinates, then σx and σy agree at least on their central $2n-1$ coordinates.

A *subshift* or *shift space* is a closed subset of some full shift $\mathcal{A}^{\mathbb{Z}}$ that is invariant under the action of σ . For example, the set of binary sequences that do not contain the string 11 is a subshift of the 2-shift. It is closed because its complement is open: if a sequence contains 11 then every sequence sufficiently close to it does as well. This subshift is often called the *golden mean shift*, in part because of its

connection to the golden mean beta expansion. More generally, let \mathcal{F} be any set of finite strings (also called *words* or *blocks*) of symbols of \mathcal{A} . The set of sequences that do not contain any word of \mathcal{F} is a subshift $X_{\mathcal{F}}$ of $\mathcal{A}^{\mathbb{Z}}$. In fact, every subshift is of this type, as an easy topological argument will show.

If $X_{\mathcal{F}}$ is determined by a finite set \mathcal{F} of “forbidden” words, we call $X_{\mathcal{F}}$ a *(sub)shift of finite type*, or SFT for short. This is the most fully studied class of symbolic dynamical systems, and the one that has been exploited most in the analysis of general dynamical systems. The systems originally considered by Hadamard were of this type.

For any subshift X we will denote the set of words of length n that appear in some element of X by $\mathcal{B}_n(X)$, the *allowed n -blocks* of X . If \mathcal{F} is a set of words of length not exceeding m , then the SFT $X_{\mathcal{F}}$ is characterized by its set of allowed m -blocks: $\mathcal{B}_m(X)$ is the set of m -blocks over the alphabet \mathcal{A} that do not contain a word of \mathcal{F} , and a sequence $x \in \mathcal{A}^{\mathbb{Z}}$ is in $X_{\mathcal{F}}$ if and only if all of its m -blocks are in $\mathcal{B}_m(X)$. A shift of finite type that is determined by its m -blocks is an $(m-1)$ -step SFT. The idea behind this terminology is that we must look back $m-1$ steps in our symbolic sequence to see which symbols we are allowed to write next. The golden mean shift is a 1-step SFT, with allowed 2-blocks 00, 01 and 10. If we allow consecutive 1’s, but no strings of three in a row, we get a 2-step SFT.

A simple example of a subshift that is not of finite type is the *even shift* first studied by B. Weiss [We]. It consists of all binary strings in which two 1’s are always separated by an even number of 0’s. Its set of forbidden words is $\mathcal{F} = \{101, 10001, 1000001, \dots\}$. A variation on this theme is the *prime gap shift*, in which two 1’s are separated by a prime number of 0’s.

A point of notation is in order before we close this section. We have been speaking of “the” shift map σ on an arbitrary subshift X . In careful parlance, two maps are not the same if they have different domains. A shift space is really a pair (X, σ_X) , where X is a closed subset of some $\mathcal{A}^{\mathbb{Z}}$ invariant under the coordinate shift on that particular full shift, and σ_X is the restriction of that coordinate shift to X . In these notes we use X as a shorthand for the pair (X, σ_X) , but often the map σ_X is singled out instead.

4. Coding and isomorphism

The term *code* is variously used in symbolic dynamics and related fields for maps of different sorts between symbolic systems, or from a general dynamical system to a symbolic one. For example, the one-sided golden mean shift might be described as a coding of the map $x \rightarrow \{\gamma x\}$ on the interval. It should be noted that in coding and information theory, a mapping may be called an *encoder*, and its image a *code*.

Within symbolic dynamics we are naturally interested in maps that preserve, at least to some extent, the topology and dynamics of the shift space. We want nearby points to go to nearby points, and if x is sent to y then its shift σx should go to σy . A *homomorphism* from one subshift to another is a continuous map ϕ that commutes with the shift, that is, for which $\phi \circ \sigma = \sigma \circ \phi$. An onto homomorphism is traditionally called a *factor map*, a term also used in ergodic theory, although the term *quotient map* would be more consistent with usage in other areas of mathematics. An isomorphism (invertible homomorphism) from one subshift to another is also called

a *topological conjugacy*, or simply a *conjugacy*, as is usual in the general theory of dynamical systems.

We can define a factor map ϕ from the golden mean shift to the even shift as follows: map $x = (x_i)$ to $y = (y_i)$ where $y_i = 1 - (x_i + x_{i+1})$ for all i . Since each 1 in x is immediately preceded and followed by a 0, the 0's in y are produced in pairs. Clearly $\sigma\phi(x) = \phi(\sigma x)$. The map is continuous because it is given by a local rule, so that a central block of y is determined by a slightly longer central block of x . In general, a *sliding block code* from a subshift X to a subshift Y is a map ϕ given by a local rule $(\phi(x))_i = \Phi(x_{i-m} \dots x_{i+a})$, where m and a are integers with $-m \leq a$ and Φ is a map from the $(m+a+1)$ -blocks of X to the symbols of Y . The numbers m and a , usually taken to be nonnegative, are called respectively the *memory* and *anticipation* of the code. An argument using the compactness of X yields the Curtis-Hedlund-Lyndon theorem: every homomorphism between subshifts is given by a sliding block code.

Of particular interest are the *higher block codes*. We can define a homomorphism from any subshift X into the full $\mathcal{B}_n(X)$ -shift by the sliding block code $\Phi(x_0x_1 \dots x_{n-1}) = [x_0x_1 \dots x_{n-1}]$. Here we use the square brackets to emphasize that the enclosed block is being treated as a single symbol in a new alphabet. Thus when $m = 2$, the sequence $\dots x_{-1}x_0x_1x_2 \dots$ is sent to

$$\dots [x_{-1}x_0][x_0x_1][x_1x_2] \dots$$

This homomorphism is clearly one-to-one. Its image is the *n-block presentation* of X , denoted $X^{[n]}$.

Higher block presentations provide an important technical tool in symbolic dynamics. If X is an m -step shift of finite type, then $X^{[m]}$ is a 1-step SFT: the allowed 2-blocks of $X^{[m]}$ are the blocks $[x_0 \dots x_{m-1}][x_1 \dots x_m]$ where $x_0 \dots x_m$ is an allowed $(m+1)$ -block of X . Thus every SFT is conjugate to a 1-step SFT. Also, if ϕ is a sliding block code with memory m and anticipation l as described above, it induces a sliding block code ψ of zero memory and anticipation from $X^{[m+l]}$ to Y given by the block map $\Psi([x_{i-m} \dots x_{i+l}]) = \Phi(x_{i-m} \dots x_{i+l})$ from symbols (1-blocks) of $X^{[m+l]}$ to symbols of Y . By this device we are often able to reduce general arguments about sliding block codes to the case of one-block codes $(\phi(x))_i = \Phi(x_i)$.

5. Graphs and matrices

Recall that a 1-step SFT X is characterized by its set of allowed 2-blocks, that is, by a list of which symbols may follow which in our symbol sequences. We may represent such a system by a directed graph: the vertices are the symbols of the alphabet \mathcal{A} and there is an edge from a to b if and only if the word ab is allowed. Each element (x_i) of X corresponds to a biinfinite walk on the graph, following edges from one vertex to the next. Conversely, a (finite) directed graph G with no parallel edges determines a 1-step SFT \hat{X}_G with alphabet equal to the set of vertices of G . We call this the *vertex shift* associated with G . The graph of the golden mean shift is shown in figure 1. From here on, *graph* will always mean a directed graph.

A graph G with n vertices is conveniently described by giving its *adjacency matrix*, the $n \times n$ matrix $A = (a_{ij})$ where a_{ij} is the number of edges from the i th vertex to the j th. Thus every vertex shift corresponds to a square matrix A of 0's

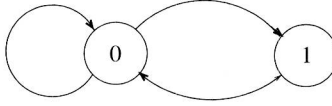


FIGURE 1. Vertex graph of the golden mean shift

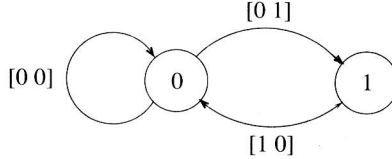


FIGURE 2. 2-block presentation of the golden mean shift

and 1's, sometimes called the *transition matrix* of the vertex shift. The transition matrix for the golden mean shift is

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

Vertex shifts capture the constraints of a 1-step SFT in an appealingly simple way. Graphs of this sort are used in the field of stochastic processes to describe Markov chain models. The vertices are states of a system, and an edge from a to b is labeled by the probability of transition from state a to state b , which is assumed to be stationary and independent of previous states. The absence of an edge from a to b indicates zero transition probability. Then the vertex shift given by the graph is the underlying topological space supporting the Markov chain. For this reason 1-step shifts of finite type are also called *topological Markov chains*.

Even if the graph G has parallel edges, we can still view it as representing a 1-step SFT if we take the set of edges of G , instead of the vertices, as the symbol set. The *edge shift* X_G is the set of biinfinite walks on the edges of G , that is, the set of edge sequences (x_i) such that the terminal vertex of x_i is the initial vertex of x_{i+1} for all i . The edge shift is also denoted by X_A , where A is the adjacency matrix of G as before.

Edge shifts allow a more efficient representation in many cases. For example, the full n -shift is given as a vertex shift by the $n \times n$ matrix of 1's, and as an edge shift by the 1×1 matrix (n) . (The full shift is a shift of finite type, with empty set of forbidden words.) Not every 1-step SFT is an edge shift. For example, there is no graph with two edges that represents the golden mean shift. However, if X is the vertex shift with graph G then its 2-block presentation $X^{[2]}$ can be naturally identified with the edge shift of G by identifying an edge from vertex a to vertex b with the 2-block $[ab]$. Hence every shift of finite type is conjugate to an edge shift. In figure 2 we show the 2-block presentation of the golden mean shift represented as an edge shift.

It can be seen that a subshift that is conjugate to a shift of finite type is itself of finite type. However, the homomorphic image of an SFT may not be an SFT, as we can see by the factor map from the golden mean shift to the even shift. The class of subshifts that are factors of SFT are called the *sofic* shifts. This term, derived from the Hebrew word for *finite*, was coined by B. Weiss [We]. Suppose

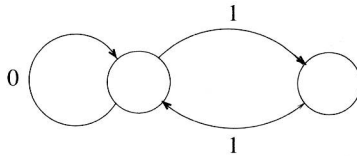


FIGURE 3. Graph of even shift

Y is a factor of an SFT X . By replacing X with a higher block presentation if necessary, we can assume that X is an edge shift X_G and that the factor map is a 1-block code. That is, it takes an edge sequence $x = (x_i) \in X$ to $y = (\Phi(x_i)) \in Y$ where Φ is a map from edges of G to symbols of Y . If we label each edge e of G with the symbol $\Phi(e)$, then the image of an edge sequence under the sliding block code is the sequence of edge labels. Hence any sofic shift may be represented by a labeling of the edges of a directed graph with symbols (not necessarily distinct) from some alphabet \mathcal{A} . As with SFT, the elements of Y correspond to biinfinite walks in this graph.

The 2-block code from the golden mean shift to the even shift described in the previous section produces the edge labeling in figure 3. It is easy to see that the prime gap system cannot be represented by a finite labeled graph, so it is not a sofic shift.

Sofic systems have a natural connection to automata theory. We think of the vertices of the graph as internal states of a machine, and the label on an edge from v to v' as the instruction that the machine, when in state v , should go to state v' if that label is read as input. A *nondeterministic finite-state automaton* (NFA) is just such a directed, edge-labeled graph, with one or more designated *initial states* and *accepting states*. The automaton is said to *accept* or *recognize* a word $b_1 \dots b_n$ if this word labels a path from an initial state to an accepting state. The set of all words accepted by an NFA is a *regular language*. In this terminology, the set of allowed blocks of all lengths in a sofic system Y is a regular language given by a NFA in which all the states are both initial and accepting. For more on connections between automata theory and symbolic dynamics, see [BP] or [BMN].

6. Invariants

One of the most basic questions we may ask about symbolic dynamical systems is how we can tell when two subshifts are conjugate. Even for the simplest class of systems, the shifts of finite type, a complete and effective classification remains elusive.

On the other hand, we have many ways of telling that two subshifts are different. By an *invariant* of conjugacy we mean any quantity or mathematical object that we can assign to subshift that remains unchanged when we replace the subshift by a conjugate one. There are several well-known invariants that can be defined for dynamical systems in general, and others that apply only to shift spaces, or to the smaller class of SFT. We list a few:

6.1. Periodic point count. If f is a homeomorphism of a compact space X , we denote by $\text{Fix}_n(X)$ the set of all $x \in X$ with $f^n(x) = x$. We will call these the period n points of the dynamical system. Note that with this terminology, a period

n point is also a period m point if m is a multiple of n . We distinguish the smallest such n by calling it the least period of x . A conjugacy between dynamical systems preserves periods of points, so conjugate shifts have the same cardinality of period n points for every n .

As we observed before, for a shift space the period n points are those sequences (x_i) with $x_{i+n} = x_i$ for all i . If the alphabet \mathcal{A} has r symbols then there are at most r^n period n points.

It is a well-known fact that in a graph with adjacency matrix A , the number of walks of length n from vertex i to vertex j along the edges of the graph is just the ij entry of A^n . For an edge shift X_A the period n points correspond to walks of length n with the same initial state and final state i . Hence the number of period n points is easily computed as the trace, or sum of the diagonal entries, of A^n .

If the number $p_n(X)$ of period n points of X is finite for all n , the periodic point count can be encoded in the Artin-Mazur *zeta function* of X . This is the power series

$$\zeta_X(t) = \exp\left(\sum_{n=1}^{\infty} \frac{p_n(X)}{n} t^n\right).$$

For the SFT X_A given by an $r \times r$ transition matrix A , this function takes the remarkably simple form

$$(6.1) \quad \begin{aligned} \zeta_A(t) &= \frac{1}{\det(I - tA)} \\ &= \frac{1}{t^r \chi_A(t^{-1})}, \end{aligned}$$

where χ_A is the characteristic polynomial of A . This formula, known as the Bowen-Lanford formula, can be verified by putting A in its Jordan normal form and using the trace result cited above. More generally, it follows from a theorem of Manning [Ma] that every sofic system has rational zeta function.

6.2. Topological entropy. Topological entropy was defined in [AKM] for general compact dynamical systems, in analogy with the concept of measure theoretic entropy developed earlier by Shannon and by Kolmogorov and Sinai. We will not give the general definition, which is fairly involved, since there is a much simpler formulation for the special case of symbolic dynamical systems. This formulation can be motivated in terms of information theory.

Think of an allowed n -block of a subshift X as the information we would gain by observing our symbolic dynamical system for n ticks of the clock. If X is the full 2-shift, the n -block may be any one of the 2^n binary strings of length n , so by recording what particular n -block occurs we gain n bits of information, or one bit per symbol. However, if X is the golden mean shift then not all binary strings can occur. We gain less information from observing a particular block since there are many blocks that we could have ruled out in advance. The number N_n of binary strings of length n with no consecutive 1's satisfies the Fibonacci recurrence relation

$$N_{n+2} = N_{n+1} + N_n,$$

since we can form an allowed $(n+2)$ -block either by tacking a 0 onto an allowed $(n+1)$ -block or by putting 01 after an allowed n -block. The number N_n is the $(n+2)$ -th Fibonacci number, and grows asymptotically as $C\gamma^n$ where C is a constant and $\gamma = (1 + \sqrt{5})/2$ is the golden mean. We may say that the amount of

information we gain by observing a particular n -block of the golden mean shift is about $\log_2(C\gamma^n) = n \log_2 \gamma + \log_2 C$ bits, or roughly $\log_2(\gamma)$ bits per symbol.

We define the (topological) *entropy* of a shift space X to be the limit

$$h(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N_n$$

where $N_n = N_n(X)$ is the number of allowed n -blocks of X . We can describe $h(X)$ as the per-symbol information rate of the shift, or as the exponential growth rate of the number of n -blocks. That this limit exists (and is equal to the infimum of the sequence) can be established from the observation that $N_{m+n} \leq N_m \cdot N_n$. Whether we use natural or base 2 logarithms is a matter of personal proclivity; with the natural log we are measuring information in *nats* instead of bits.

If Y is a factor of X given as the image of a sliding m -block code, then the number of n -blocks of Y cannot exceed the number of $(n+m)$ -blocks of X . Thus

$$\begin{aligned} h(Y) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log N_n(Y) \\ (6.2) \quad &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log N_{n+m}(X) \\ &= \left(\lim_{n \rightarrow \infty} \frac{n+m}{n} \right) h(X) = h(X). \end{aligned}$$

Since conjugate shifts are factors of one another, they have equal entropy.

We see immediately that the entropy of the full r -shift is $\log r$, since there are r^n words of length n . For a shift of finite type X_A given by a transition matrix A , the number of allowed n -blocks is the sum of the entries of A^n . By a result known as the Perron-Frobenius theorem, every square nonnegative matrix A has a nonnegative real eigenvalue λ_A (the *Perron-Frobenius eigenvalue*) that is greater than or equal to the modulus of every other eigenvalue of A . For the golden mean shift $\lambda_A = \gamma$. It can be shown that in general, $h(X_A) = \log \lambda_A$.

6.3. Algebraic invariants for shifts of finite type. As we have seen, for a shift of finite type X_A the number of period n points, the zeta function and the entropy can all be simply expressed in terms of algebraic invariants of the transition matrix A . We can see from the invariance of the zeta function that if X_A and X_B are conjugate SFT then A and B must have the same characteristic polynomial, up to some factor t^k . In fact a stronger statement is true: the Jordan forms of the invertible parts of A and B must be the same.

Another useful algebraic invariant is the *Bowen-Franks group*. If A is an $r \times r$ transition matrix, its Bowen-Franks group is the quotient of \mathbb{Z}^r by its image under the matrix $I - A$:

$$BF(A) = \mathbb{Z}^r / \mathbb{Z}^r(I - A).$$

If X_A is conjugate to X_B then the Bowen-Franks groups of A and B must be isomorphic [BF]. This condition is easily checked by computing the elementary divisors of A and B . The Bowen-Franks group is, in fact, invariant under flow equivalence of SFT, a weaker equivalence than conjugacy.

One of the most sought-after goals in symbolic dynamics has been a complete and effective classification of shifts of finite type in terms of their transition matrices. In 1973 R. Williams [Wi] introduced two important equivalence relations on the set of square matrices over the non-negative integers, *shift equivalence* and *strong shift equivalence*. He showed that finite type shifts X_A and X_B are conjugate if

and only A and B are strong shift equivalent. Strong shift equivalence implies shift equivalence; the converse statement became known as the shift equivalence conjecture or Williams conjecture. The importance of the conjecture lies in the fact that latter equivalence is more tractable: in fact, it is known to be decidable [KR1]. Jack Wagoner's chapter relates the developments that led in 1997 to a counterexample to the shift equivalence conjecture by K. H. Kim and F. Roush [KR2] following joint work with Wagoner, and outlines the current state of affairs.

Crucial to the solution of the shift equivalence problem was the study of the group of automorphisms, or self-conjugacies, of a shift of finite type. Bob Devaney's chapter will take us a step beyond Hadamard's inspiration by showing how the automorphism group of a shift can encode information from other dynamical settings, in this case families of complex polynomial maps.

7. Wider vistas

There are several ways in which we can relax our notion of symbolic dynamical system to get a larger class of systems. One is to allow a countable alphabet in place of a finite one. This makes the shift space noncompact, which introduces some complications, for example in finding an appropriate definition of entropy. The theory of countable state topological Markov chains—vertex shifts on a graph with countably many vertices but only finitely many edges entering or leaving each edge—is of particular interest. A good introduction to this topic is Chapter 7 of [Ki].

We could instead choose our symbol set to be a compact group such as the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Although this takes us far from the original idea of a symbolic dynamical system as a discretization of space, if our map is still a coordinate shift map then some of the spirit and techniques remain. We will encounter shift spaces with alphabet \mathbb{T} in Doug Lind's chapter.

The study of a dynamical system (X, σ) is really the study of the behavior of X under the iterates σ^n of σ . We may describe this as an action of the group \mathbb{Z} on X : for every $n \in \mathbb{Z}$ we have a coordinate shift map σ^n that shifts all coordinates by n , with $\sigma^{m+n} = \sigma^m \circ \sigma^n$. In general, an *action* of a group G by homeomorphisms on a space X is a map that takes each $g \in G$ to a homeomorphism f_g of X in such a way that $f_{gh} = f_g \circ f_h$. Another way to broaden our notion of symbolic dynamical systems is to consider actions by other discrete infinite groups in place of \mathbb{Z} . One of the most exciting currents in symbolic dynamics is development of the theory of \mathbb{Z}^d -actions. Elements of a \mathbb{Z}^d symbolic dynamical system are d -dimensional arrays $(x_n)_{n \in \mathbb{Z}^d}$ of symbols, and for each $\mathbf{m} \in \mathbb{Z}^d$ there is a shift map $\sigma^{\mathbf{m}}$ that shifts all coordinates by \mathbf{m} . Doug Lind's chapter is a survey of these multidimensional systems. As you will see, there are very nice results for special class of multidimensional systems with algebraic structure, but the study of general \mathbb{Z}^d -actions involves substantial complications not found in the one-dimensional case.

We can think of an element of a two-dimensional symbolic dynamical system as a tiling of the plane by unit square tiles of different colors, where the colors are simply our alphabet of symbols. A translation of the tiling by an integer vector \mathbf{m} gives another tiling that represents a coordinate shift of the original. To bring the techniques of symbolic dynamics to bear on the general problem of tiling the plane with tiles of various shapes we need to allow general translations in the plane. In Robbie Robinson's chapter, which examines not just planar but d -dimensional

tilings, we will be working in the setting of \mathbb{R}^d actions by translation of Euclidean space.

References

- [AKM] R. Adler, A. Konheim & M. McAndrew, *Topological Entropy* **114** (1965), 309–319.
- [BF] R. Bowen & J. Franks, *Homology for zero-dimensional basic sets*, *Annals of Math.* **106** (1977), 37–92.
- [BP] M.-P. Béal & D. Perrin, *Symbolic dynamics and finite automata*, *Handbook of Formal Languages* (G. Rozenberg & A. Salomaa, eds.), vol. 2, Springer-Verlag, Berlin, 1997, pp. 463–505.
- [BMN] F. Blanchard, A. Maass & A. Nogueira, eds., *Topics in Symbolic Dynamics and Applications*, Cambridge Univ. Press, Cambridge, U.K., 2000.
- [Bo] M. Boyle, *Symbolic dynamics and matrices*, *Combinatorial and Graph Theoretic Problems in Linear Algebra*, IMA Volumes in Math. and its Appl. (R. Brualdi et al., eds.), vol. 50, 1993, pp. 1–38.
- [Ha] J. Hadamard, *Les surfaces à courbures opposées et leur lignes géodésiques*, *J. Maths. Pures et Appliqués* **4** (1898), 27–73.
- [Ki] B. Kitchens, *Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Chains*, Springer-Verlag, Berlin, Heidelberg, 1998.
- [KR1] K.H. Kim & F.W. Roush, *Decidability of shift equivalence*, in J.W. Alexander, ed., *Dynamical systems*, *Lecture Notes in Mathematics*, vol. 1342, Springer-Verlag, Heidelberg, 1988.
- [KR2] K.H. Kim & F.W. Roush, *Williams’s conjecture is false for irreducible subshifts*, *Elec. Res. Announc. Amer. Math. Soc.* **3** (1997), 105–109.
- [MH] M. Morse & G.A. Hedlund, *Symbolic Dynamics*, *Amer. J. Math.* **60** (1938), 815–866.
- [LM] D. Lind & B. Marcus, *Symbolic Dynamics and Coding*, Cambridge Univ. Press, Cambridge, U.K., 1995.
- [Ma] A. Manning, *Axiom A diffeomorphisms have rational zeta functions*, *Bull. London Math. Soc.* **3** (1971), 215–220.
- [Sh] C.E. Shannon, *A mathematical theory of communication*, *Bell Syst. Tech. J.* **27** (1948), 379–423, 623–656.
- [We] B. Weiss, *Subshifts of finite type and sofic systems*, *Monats. Math.* **77** (1973), 462–474.
- [Wi] R.F. Williams, *Classification of subshifts of finite type*, *Annals of Math.* **98** (1973), 120–153; *erratum*, *Annals of Math.* **99** (1974), 380–381.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF SOUTH ALABAMA, MOBILE, AL 36688

E-mail address: swilliam@jaguar1.usouthal.edu