Measuring Complexity in Cantor Dynamics

Karl Petersen

cantorsalta2015: Dynamics on Cantor Sets
CIMPA Research School, 2-13 November 2015

November 26, 2015
## Contents

1 Introduction 5
  1.1 Preface ........................................ 5
  1.2 Complexity and entropy 5
  1.3 Some definitions and notation 5
  1.4 Realizations of systems 7

2 Asymptotic exponential growth rate 9
  2.1 Topological entropy .................................. 9
  2.2 Ergodic-theoretic entropy ................................ 9
  2.3 Measure-theoretic sequence entropy ..................... 12
  2.4 Topological sequence entropy ......................... 13
  2.5 Slow entropy ........................................ 14
  2.6 Entropy dimension ................................... 17
  2.7 Permutation entropy .................................. 18
  2.8 Independence entropy ................................ 19
  2.9 Sofic, Rokhlin, naïve entropies .......................... 20
  2.10 Kolmogorov complexity ................................ 22

3 Counting patterns 25
  3.1 The complexity function in one-dimensional symbolic dynamics .... 25
  3.2 Sturmian sequences .................................... 26
  3.3 Episturmian sequences .................................. 28
  3.4 The Morse sequence .................................... 29
  3.5 In higher dimensions, tilings, groups, etc. .................. 29
  3.6 Topological complexity .................................. 31
  3.7 Low complexity, the number of ergodic measures, automorphisms .... 31
  3.8 Palindrome complexity .................................. 33
  3.9 Nonrepetitive complexity and Eulerian entropy ............... 35
  3.10 Mean topological dimension .................................. 36
  3.11 Amorphic complexity via asymptotic separation numbers .... 36
  3.12 Inconstancy ........................................... 37
  3.13 Measure-theoretic complexity ............................ 38
  3.14 Pattern complexity ..................................... 39

4 Balancing freedom and interdependence 41
  4.1 Neurological intricacy .................................. 41
  4.2 Topological intricacy and average sample complexity ............. 43
  4.3 Ergodic-theoretic intricacy and average sample complexity ......... 46
  4.4 The average sample complexity function .......................... 46
  4.5 Computing measure-theoretic average sample complexity .......... 48
  4.6 The search for maximizing measures on subshifts .................. 48
Part 1

Introduction

1.1 Preface

In these notes we discuss several quantitative definitions of the broad and vague notion of complexity, especially from the viewpoint of dynamical systems, focusing on transformations on the Cantor set, in particular shift dynamical systems. After this introduction, the second part reviews dynamical entropy, the asymptotic exponential growth rate of the number of patterns or available information, uncertainty, or randomness in a system as the window size grows. The third part treats the more precise complexity function, which counts the actual number of patterns of each size, and then several of its variations. In the fourth part we present a new quantity that measures the balance within a system between coherent action of the whole and independence of its parts. There is a vast literature on these matters, new papers full of new ideas are appearing all the time, and there are plenty of questions to be asked and investigated with a variety of approaches. (Our list of references is in no way complete.) Some of the attractiveness of the subject is due to the many kinds of mathematics that it involves: combinatorics, number theory, probability, and real analysis, as well as dynamics. For general background and useful surveys, see for example [4, 19, 75, 78, 133, 137, 152, 181].

Parts of these notes are drawn from earlier writings by the author or from theses of his students Kathleen Carroll and Benjamin Wilson (see also [45, 153]). Thanks to Valérie Berthé, Francesco Dolce, and Bryna Kra, among others, for pointing out references. I thank the participants, organizers, and supporters of the CIMPA Research School cantorsalta2015, Dynamics on Cantor Sets, for the opportunity to participate and the incentive to produce these notes.

1.2 Complexity and entropy

An elementary question about any phenomenon under observation is, how many possibilities are there. A system that can be in one of a large but finite number of states may be thought to be more complex than one that has a choice among only a few. Then consider a system that changes state from time to time, and suppose we note the state of the system at each time. How many possible histories, or trajectories, can there be in a time interval of fixed length? This is the complexity function, and it provides a quantitative way to distinguish relatively simple systems (for example periodic motions) from more complicated (for example “chaotic”) ones. In systems with a lot of freedom of motion the number of possible histories may grow very rapidly as the length of time it is observed increases. The exponential growth rate of the number of histories is the entropy. While it may seem to be a very crude measure of the complexity of a system, entropy has turned out to be the single most important and useful number that one can attach to a dynamical system.

1.3 Some definitions and notation

A topological dynamical system is a pair \((X, T)\), where \(X\) is a compact Hausdorff space (usually metric) and \(T : X \to X\) is a continuous mapping. In these notes \(X\) is usually the Cantor set, often in a specific representation as a subshift or as the set of infinite paths starting at the root in a Bratteli diagram—see
A measure-preserving system \((X, \mathcal{B}, T, \mu)\) consists of a measure space \((X, \mathcal{B}, \mu)\) and a measure-preserving transformation \(T : X \to X\). Often no generality is lost in assuming that \((X, \mathcal{B}, \mu)\) is the Lebesgue measure space of the unit interval, and in any case usually we assume that \(\mu(X) = 1\). A standard measure space is one that is measure-theoretically isomorphic to a Borel subset of a complete metric space with its \(\sigma\)-algebra of Borel sets and a Borel measure. \(T\) is assumed to be defined and one-to-one a.e., with \(T^{-1}\mathcal{B} \subset \mathcal{B}\) and \(\mu T^{-1} = \mu\). The system is called ergodic if for every invariant measurable set (every \(B \in \mathcal{B}\) satisfying \(\mu(B \triangle T^{-1}B) = 0\)) either \(\mu(B) = 0\) or \(\mu(X \setminus B) = 0\).

A homomorphism or factor mapping between topological dynamical systems \((X, T)\) and \((Y, S)\) is a continuous onto map \(\phi : X \to Y\) such that \(\phi T = S \phi\). We say \(Y\) is a factor of \(X\), and \(X\) is an extension of \(Y\). If \(\phi\) is also one-to-one, then \((X, T)\) and \((Y, M)\) are topologically conjugate and \(\phi\) is a topological conjugacy. A homomorphism or factor mapping between measure-preserving systems \((X, \mathcal{B}, T, \mu)\) and \((Y, \mathcal{B}, S, \nu)\) is a map \(\phi : X \to Y\) such that \(\phi^{-1} \mathcal{C} \subset \mathcal{B}\), \(\phi T = S \phi\) a.e., and \(\mu \phi^{-1} = \nu\). If in addition \(\phi\) is one-to-one a.e., equivalently \(\phi^{-1} \mathcal{C} = \mathcal{B}\) up to sets of measure 0, then \(\phi\) is an isomorphism.

We focus on these notes especially on topological dynamical systems which are shift dynamical systems. Let \(A\) be a finite set called an alphabet. The elements of this set are letters and shall be denoted by digits. A sequence is a one-sided infinite string of letters and a bisequence is an infinite string of letters that extends in two directions. The full \(A\)-shift, \(\Sigma(A)\), is the collection of all bisequences of symbols from \(A\). If \(A\) has \(n\) elements,

\[
\Sigma(A) = \Sigma_n = A^\mathbb{Z} = \{x = (x_i)_{i \in \mathbb{Z}} \mid x_i \in A \text{ for all } i \in \mathbb{Z}\}.
\]

The one-sided full \(A\)-shift is the collection of all infinite sequences of symbols from \(A\) and is denoted

\[
\Sigma^+(A) = \Sigma_n^+ = A^\mathbb{N} = \{x = (x_i)_{i \in \mathbb{N}} \mid x_i \in A \text{ for all } i \in \mathbb{N}\}.
\]

We also define the shift transformation \(\sigma : \Sigma(A) \to \Sigma(A)\) and \(\Sigma^+(A) \to \Sigma^+(A)\) by

\[
(\sigma x)_i = x_{i+1} \text{ for all } i.
\]

The pair \((\Sigma_n, \sigma)\) is called the \(n\)-shift dynamical system.

We give \(A\) the discrete topology and \(\Sigma(A)\) and \(\Sigma^+(A)\) the product topology. Furthermore, the topologies on \(\Sigma(A)\) and \(\Sigma^+(A)\) are compatible with the metric \(d(x, y) = 1/2^n\), where \(n = \inf \{|k| \mid x_k \neq y_k\}\) [133]. Thus two elements of \(\Sigma(A)\) are close if and only if they agree on a long central block. In a one-sided shift, two elements are close if and only if they agree on a long initial block.

A subshift is a pair \((X, \sigma)\) (or \((X^+, \sigma)\)), where \(X \subset \Sigma_n\) (or \(X^+ \subset \Sigma^+_n\)) is a nonempty, closed, shift-invariant set. We will be concerned primarily with subshifts of the 2-shift dynamical system.

A finite string of letters from \(A\) is called a block and the length of a block \(B\) is denoted \(|B|\). Furthermore, a block of length \(n\) is an \(n\)-block. A formal language is a set \(\mathcal{L}\) of blocks, possibly including the empty block \(\epsilon\), on a fixed finite alphabet \(A\). The set of all blocks on \(A\), including the empty block \(\epsilon\), is denoted by \(A^*\). Given a subshift \((X, \sigma)\) of a full shift, let \(\mathcal{L}_n(X)\) denote the set of all \(n\)-blocks that occur in points in \(X\). The language of \(X\) is the collection

\[
\mathcal{L}(X) = \bigcup_{n=0}^{\infty} \mathcal{L}_n(X).
\]

A shift of finite type (SFT) is a subshift determined by excluding a finite set of blocks.

Let \(A\) be a finite alphabet. A map \(\theta : A \to A^*\) is called a substitution. A substitution \(\theta\) is extended to \(A^*\) and \(A^0\) by \(\theta(b_1 b_2 \ldots) = \theta(b_1)\theta(b_2)\ldots\). A substitution \(\theta\) is called primitive if there is \(m\) such that for all \(a \in A\) the block \(\theta^m(a)\) contains every element of \(A\).
### 1.4. Realizations of systems

There is natural dynamical system associated with any sequence. Given a one-sided sequence $u$, we let $X_u^+$ be the closure of $\{\sigma^n u | n \in \mathbb{N}\}$, where $\sigma$ is the usual shift. Then $(X_u^+, \sigma)$ is the dynamical system associated with $u$. For any sequence $u$, denote by $\mathcal{L}(u)$ the family of all subblocks of $u$.

**Exercise 1.3.1.** Show that $X_u^+$ consists of all the one-sided sequences on $\mathcal{A}$ all of whose subblocks are subblocks of $u$: $X_u^+ = \{ x : \mathcal{L}(x) \subset \mathcal{L}(u) \}$.

In a topological dynamical system $(X, T)$, a point $x \in X$ is called *almost periodic* or *syndetically recurrent* if for every $\epsilon > 0$ there is some $N = N(\epsilon)$ such that the set $\{ n \geq 0 : d(T^n x, x) < \epsilon \}$ has gaps of size at most $N$. If $X$ is a subshift, then $x \in X$ is almost periodic if and only if every allowed block in $x$ appears in $x$ with bounded gaps.

A topological dynamical system $(X, T)$ is *minimal* if one of the following equivalent properties holds:

1. $X$ contains no proper closed invariant set;
2. $X$ is the orbit closure of an almost periodic point;
3. every $x \in X$ has a dense orbit in $X$.

The *complexity function* of a language $\mathcal{L}$ is the function $p_\mathcal{L}(n) = \text{card}(\mathcal{L} \cap \mathcal{A}^n)$, $n \geq 0$. This is an elementary, although possibly complicated and informative, measure of the size or complexity of a language and, if the language is that of a subshift, of the complexity of the associated symbolic dynamical system. Properties of this function (for example its asymptotic growth rate, which is the topological entropy of the associated subshift) and extensions and variations comprise most of the subject matter of these notes.

A *Bratteli diagram* is a graded graph whose set $\mathcal{V}$ of vertices is the disjoint union of finite sets $\mathcal{V}_n, n = 0, 1, 2, \ldots$; $\mathcal{V}_0$ consists of a single vertex $v_0$, called the *root*; and the set $\mathcal{E}$ of edges is also the disjoint union of sets $\mathcal{E}_n, n = 1, 2, \ldots$ such that the source vertex of each edge $e \in \mathcal{E}_n$ is in $\mathcal{V}_{n-1}$ and its range vertex is in $\mathcal{V}_n$. Denote by $X$ the set of all infinite directed paths $x = (x_n), x_n \in \mathcal{E}_n$ for all $n \geq 1$, in this graph that begin at the root. $X$ is a compact metric space when we agree that two paths are close if they agree on a long initial segment. Except in some degenerate situations the space $X$ is infinite, indeed uncountable, and homeomorphic to the Cantor set.

Suppose we fix a linear order on the set of edges into each vertex. Then the set of paths $X$ is partially ordered as follows: two paths $x$ and $y$ are *comparable* if they agree from some point on, in which case we say that $x < y$ if at the last level $n$ where they are different, the edge $x_n$ of $x$ is smaller than the edge $y_n$ of $y$. A map $T$, called the *Vershik map*, is defined by letting $Tx$ be the smallest $y$ that is larger than $x$, if there is one. There may be maximal paths $x$ for which $Tx$ is not defined, as well as minimal paths. In nice situations, $T$ is a homeomorphism after the deletion of perhaps countably many maximal and minimal paths and their orbits. If the diagram is *simple*—which means that for every $n$ there is $m > n$ such that there is a path in the graph from every $v \in \mathcal{V}_n$ to every $w \in \mathcal{V}_m$—and if there are exactly one maximal path $x_{\text{max}}$ and exactly one minimal path $x_{\text{min}}$, then one may define $Tx_{\text{max}} = x_{\text{min}}$ and arrive at a minimal homeomorphism $T : X \rightarrow X$. See [23, 70] for surveys on Bratteli-Vershik systems.

### 1.4 Realizations of systems

There are several results concerning the realization of measure-preserving systems as topological dynamical systems up to measure-theoretic isomorphism.

1. The Jewett-Krieger Theorem states that every non-atomic ergodic measure-preserving system on a Lebesgue space is measure-theoretically isomorphic to a system $(X, \mathcal{B}, \mu, T)$ in which $X$ is the Cantor set, $\mathcal{B}$ is the completion of the Borel $\sigma$-algebra of $X$, $T$ is a minimal homeomorphism (every orbit is dense), and $\mu$ is a unique $T$-invariant Borel probability measure on $X$. 

2. The Krieger Generator Theorem says that every ergodic measure-preserving system \((X, \mathcal{B}, \mu, T)\) of finite entropy (see below) is measure-theoretically isomorphic to a subsystem of any full shift which has strictly larger (topological) entropy—see below—with a shift-invariant Borel probability measure. Thus full shifts are “universal” in this sense. The proof is accomplished by producing a finite measurable partition of \(X\) such that coding orbits according to visits of the members of the partition produces a map to the full shift that is one-to-one a.e. Recently Seward [172, 173] has extended the theorem to actions of countable groups, using his definition of Rokhlin entropy (see Section 2.9).

3. Krieger proved also that such an embedding is possible into any mixing shift of finite type (see below) that has strictly larger topological entropy than the measure-theoretic entropy of \((X, \mathcal{B}, \mu, T)\). Moreover, he gave necessary and sufficient conditions that an expansive homeomorphism of the Cantor set be topologically conjugate to a subshift of a given mixing shift of finite type. “There is a version of the finite generator theorem for ergodic measure preserving transformations of finite entropy, that realizes such a transformation by means of an invariant probability measure of any irreducible and aperiodic topological Markov chain, whose topological entropy exceeds the entropy of the transformation ([4], [2 28]). One can say that a corollary of theorem 3 achieves for minimal expansive homeomorphisms of the Cantor discontinuum what the finite generator theorem does for measure preserving transformations.” [126]

4. Lind and Thouvenot [132] proved that hyperbolic toral automorphisms (the matrix has no eigenvalue of modulus 1) are universal. This was extended by Quas and Soo [158] to quasi-hyperbolic toral automorphisms (no roots of unity among the eigenvalues), and they also showed that the time-1 map of the geodesic flow on a compact surface of constant negative curvature is universal [159].

5. Every minimal homeomorphism of the Cantor set is topologically conjugate to the Vershik map on a simple Bratteli diagram with unique maximal and minimal paths [83, 95].

6. Every ergodic measure-preserving system is measure-theoretically isomorphic to a minimal Bratteli-Vershik system with a unique invariant Borel probability measure [178, 179].
Part 2

Asymptotic exponential growth rate

2.1 Topological entropy

Let $X$ be a compact metric space and $T : X \to X$ a homeomorphism.

First definition \cite{2}: For an open cover $\mathcal{U}$ of $X$, let $N(\mathcal{U})$ denote the minimum number of elements in a subcover of $\mathcal{U}$, $H(\mathcal{U}) = \log N(\mathcal{U})$,
\[
h(\mathcal{U},T) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{U} \vee T^{-1} \mathcal{U} \vee \ldots \vee T^{-n+1} \mathcal{U}),
\]
and
\[
h(T) = \sup_{\mathcal{U}} h(\mathcal{U},T).
\]

Second definition \cite{37}: For $n \in \mathbb{N}$ and $\epsilon > 0$, a subset $A \subset X$ is called $n, \epsilon -$separated if given $a, b \in A$ with $a \neq b$, there is $k \in \{0, \ldots, n-1\}$ with $d(T^k a, T^k b) \geq \epsilon$. We let $S(n, \epsilon)$ denote the maximum possible cardinality of an $n, \epsilon$-separated set. Then
\[
h(T) = \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log S(n, \epsilon).
\]

Third definition \cite{37}: For $n \in \mathbb{N}$ and $\epsilon > 0$, a subset $A \subset X$ is called $n, \epsilon$-spanning if given $x \in X$ there is $a \in A$ with $d(T^k a, T^k x) \leq \epsilon$ for all $k = 0, \ldots, n-1$. We let $R(n, \epsilon)$ denote the minimum possible cardinality of an $n, \epsilon$-spanning set. Then
\[
h(T) = \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log R(n, \epsilon).
\]

Exercise 2.1.1. If $(X, T)$ is a subshift ($X$ = a closed shift-invariant subset of the set of all doubly infinite sequences on a finite alphabet, $T = \sigma$ = shift transformation), then
\[
h(\sigma) = \lim_{n \to \infty} \frac{\log \text{(number of } n\text{-blocks seen in sequences in } X)}{n}.
\]

Theorem 2.1.1 ("Variational Principle"). $h(T) = \sup \{h_\mu(T) : \mu \text{ is an invariant (ergodic) Borel probability measure on } X\}$.

2.2 Ergodic-theoretic entropy

2.2.1 Definition

A finite (or sometimes countable) measurable partition
\[
\alpha = \{A_1, \ldots, A_r\}
\]
Part 2. Asymptotic exponential growth rate

of $X$ is thought of as the set of possible outcomes of an experiment (performed at time 0) or as an
alphabet of symbols used to form messages (the experiment could consist of receiving and reading one
symbol). The entropy of the partition is

$$H_\mu(\alpha) = \sum_{A \in \alpha} -\mu(A) \log \mu(A)$$

(the logs can be base $e$, 2, or $r$);

it represents the amount of information gained=amount of uncertainty removed when the experiment
is performed or one symbol is received (averaged over all possible states of the world—the amount of
information gained if the outcome is $A$ (i.e., we learn to which cell of $\alpha$ the world actually belongs) is $-\log \mu(A)$). (Note that this is large when $\mu(A)$ is small.) Notice that the information gained when we
learn that an event $A$ occurred is additive for independent events.

The partition

$$T^{-1} \alpha = \{ T^{-1} A : A \in \alpha \}$$

represents performing the experiment $\alpha$ (or reading a symbol) at time 1, and $\alpha \lor T^{-1} \alpha \lor \ldots \lor T^{-n+1} \alpha$
represents the result of $n$ repetitions of the experiment (or the reception of a string of $n$ symbols). Then
$$H(\alpha \lor T^{-1} \alpha \lor \ldots \lor T^{-n+1} \alpha) / n$$
is the long-term time average of the information gained per unit time. (This limit exists because of the
subadditivity of $H$: $H(\alpha \lor \beta) \leq H(\alpha) + H(\beta)$.)

The entropy of the system $(X, \mathcal{B}, \mu, T)$ is defined to be

$$h_\mu(T) = \sup_{\alpha} h(\alpha, T),$$

the maximum information per unit time available from any finite- (or countable-) state stationary process
generated by the system.

Theorem 2.2.1 (Kolmogorov-Sinai). If $T$ has a finite generator $\alpha$—a partition $\alpha$ such that the smallest
$\sigma$-algebra that contains all $T^j \alpha, j \in \mathbb{Z}$, is $\mathcal{B}$—then $h_\mu(T) = h(\alpha, T)$. (Similarly if $T$ has a countable
generator with finite entropy.)

Theorem 2.2.2. If $\{\alpha_k\}$ is an increasing sequence of finite partitions which generates $\mathcal{B}$ up to sets of
measure 0, then $h(\alpha_k, T) \to h(T)$ as $k \to \infty$.

2.2.2 Conditioning

For a countable measurable partition $\alpha$ and sub-$\sigma$-algebra $\mathcal{F}$ of $\mathcal{B}$, we define the conditional information
function of $\alpha$ given $\mathcal{F}$ by

$$I_{\alpha|\mathcal{F}}(x) = -\sum_{A \in \alpha} \log \mu(A|\mathcal{F})(x) \chi_A(x);$$

this represents the information gained by performing the experiment $\alpha$ (if the world is in state $x$) after
we already know for each member of $\mathcal{F}$ whether or not it contains the point $x$. The conditional entropy
of $\alpha$ given $\mathcal{F}$ is

$$H(\alpha|\mathcal{F}) = \int_X I_{\alpha|\mathcal{F}}(x) d\mu(x);$$

this is the average over all possible states $x$ of the information gained from the experiment $\alpha$. When $\mathcal{F}$
is the $\sigma$-algebra generated by a partition $\beta$, we often just write $\beta$ in place of $\mathcal{F}$.

Proposition 2.2.3. 1. $H(\alpha \lor \beta|\mathcal{F}) = H(\alpha|\mathcal{F}) + H(\beta|\mathcal{F} \lor \mathcal{F})$.

2. $H(\alpha|\mathcal{F})$ is increasing in its first variable and decreasing in its second.

Theorem 2.2.4. For any finite (or countable finite-entropy) partition $\alpha$,

$$h(\alpha, T) = H(\alpha|\mathcal{B}(T^{-1} \alpha \lor T^{-2} \alpha \lor \ldots)),$$
2.2. Ergodic-theoretic entropy

2.2.3 Examples

1. Bernoulli shifts: \( h = - \sum p_i \log p_i \). Consequently \( B(1/2, 1/2) \) is not isomorphic to \( B(1/3, 1/3, 1/3) \).
2. Markov shifts: \( h = - \sum p_i \sum P_{ij} \log P_{ij} \).
3. Discrete spectrum (the span of the eigenfunctions is dense in \( L^2 \)): \( h = 0 \). (Similarly for rigid systems—ones for which there is a sequence \( n_k \to \infty \) with \( T^{n_k} f \to f \) for all \( f \in L^2 \).) Similarly for any system with a one-sided generator, for then \( h(\alpha, T) = H(\alpha|\alpha^\infty) = H(\alpha|B) = 0 \). It’s especially easy to see for an irrational rotation of the circle, for if \( \alpha \) is the partition into two disjoint arcs, then \( \alpha_0^n \) only has \( 2(n+1) \) sets in it.
4. Products: \( h(T_1 \times T_2) = h(T_1) + h(T_2) \).
5. Factors: If \( \pi : T \to S \), then \( h(T) \geq h(S) \).
6. Bounded-to-one factors: \( h(T) = h(S) \). See [149, p. 56].
7. Skew products: \( h(T \times \{S_x\}) = h(T) + h_T(S) \). Here the action is \( (x, y) \to (T x, S_x y) \), with each \( S_x \) a m.p.t. on \( Y \), and the second term is the fiber entropy \( h_T(S) = \sup (\int_X H(\beta|S_x^{-1} \beta \vee S_x^{-1} S_T \beta \vee \ldots) d\mu(x) : \beta \text{ is a finite partition of } Y) \).
8. Automorphism of the torus: \( h = \sum_{|\lambda_i| > 1} \log |\lambda_i| \) (the \( \lambda_i \) are the eigenvalues of the integer matrix with determinant \( \pm 1 \)).
9. Pesin’s Formula: If \( \mu \ll m \) (Lebesgue measure on the manifold), then
\[
h_\mu(f) = \int \sum_{\lambda_k(x) > 0} q_k(x) \lambda_k(x) \, d\mu(x),
\]
where the \( \lambda_k(x) \) are the Lyapunov exponents and \( q_k(x) = \dim(V_k(x) \setminus V_{k-1}(x)) \) are the dimensions of the corresponding subspaces.
10. Induced transformation (first-return map): For \( A \subset X \), \( h(T_A) = h(T)/\mu(A) \).
11. Finite rank ergodic: \( h = 0 \).

Proof. Suppose rank = 1, let \( P \) be a partition into two sets (labels 0 and 1), let \( \epsilon > 0 \). Take a tower of height \( L \) with levels approximately \( P \)-constant (possible by rank 1; we could even take them \( P \)-constant) and \( \mu(\text{junk}) < \epsilon \). Suppose we follow the orbit of a point \( N \gg L \) steps; how many different \( P, N \)-names can we see? Except for a set of measure \( < \epsilon \), we hit the junk \( n \sim \epsilon N \) times. There are \( L \) starting places (levels of the tower): \( C(N, n) \) places with uncertain choices of 0, 1; and \( 2^n \) ways to choose 0 or 1 for these places. So the sum of \( \mu(A) \log \mu(A) \) over \( A \) in \( P_0^{n-1} \) is \( \leq \) the log of the number of names seen in the good part minus the log of \( 2^N(\epsilon/2^N) \log(\epsilon/2^N) \), and dividing by \( N \) gives
\[
\frac{\log L}{N} + NH(\epsilon, 1-\epsilon) + \frac{N\epsilon}{N} + \frac{\epsilon(-\log \epsilon + N)}{N} \sim 0.
\]
Similarly for any finite partition \( P \). Also for rank \( r \)—then we have to take care (not easily) about the different possible ways to switch columns when spilling over the top. \( \square \)

Exercise 2.2.1. Prove the statements in 1–5 above.

2.2.4 Ornstein’s Isomorphism Theorem

Two Bernoulli shifts are isomorphic if and only if they have the same entropy.
2.2.5 Shannon-McMillan-Breiman Theorem

For a finite measurable partition $\alpha$, and $x \in X$, let $\alpha(x)$ denote the member of $\alpha$ to which $x$ belongs, and let $\alpha_0^{-1} = \alpha \lor T^{-1} \alpha \lor \ldots \lor T^{-n+1} \alpha$. If $T$ is an ergodic m.p.t. on $X$, then

$$-\frac{1}{n} \log \mu(\alpha_0^{-1}(x)) \to h(\alpha, T) \text{ a.e. and in } L^1.$$ 

2.2.6 Brin-Katok local entropy formula [39]

Let $(X, T)$ be a topological dynamical system (with $X$ compact metric as usual). For $x \in X$, $\delta > 0$, $n \in \mathbb{N}$, define

$$(2.2.1) B(x, \delta, n) = \{ y : d(T^j x, T^j y) < \delta \text{ for all } j = 1, \ldots, n \},$$

$$h^+(x) = \lim_{\delta \to 0} \limsup_{n \to \infty} -\frac{\log \mu(B(x, \delta, n))}{n}, \text{ and }$$

$$h^-(x) = \lim_{\delta \to 0} \liminf_{n \to \infty} -\frac{\log \mu(B(x, \delta, n))}{n}.$$ 

Then $h^+(x) = h^-(x)$ a.e., and

$$(2.2.2) h_\mu(T) = \int_X h^+(x) \, d\mu.$$ 

If $(X, B, \mu, T)$ is ergodic, then $h^+(x) = h^-(x) = h_\mu(T)$ a.e. with respect to $\mu$. Thus the Bowen-Dinaburg ball $B(x, \delta, n)$ plays the same role as the partition $\alpha_0^{-1}$ in the Shannon-McMillan-Breiman Theorem. The functions $h^+$ and $h^-$ may be called the upper and lower local entropies of $(X, B, \mu, T)$.

2.2.7 Wyner-Ziv-Ornstein-Weiss entropy calculation algorithm

For a stationary ergodic sequence $\{\omega_1, \omega_2, \ldots\}$ on a finite alphabet and $n \geq 1$, let $R_n(\omega) =$ the first place to the right of 1 at which the initial $n$-block of $\omega$ reappears (not overlapping its first appearance). Then

$$\frac{\log R_n(\omega)}{n} \to h \text{ a.e.}.$$ 

This is also related to the Lempel-Ziv parsing algorithm, in which a comma is inserted in a string $\omega$ each time a block is completed (beginning at the preceding comma) which has not yet been seen in the sequence.

2.3 Measure-theoretic sequence entropy

A. G. Kushnirenko [129] defined the sequence entropy of a measure-preserving system $(X, B, \mu, T)$ with respect to a sequence $S = (t_i)$ of integers as follows. Given a (finite or countable) measurable partition $\alpha$ of $X$, as usual

$$H(\alpha) = -\sum_{A \in \alpha} \mu(A) \log \mu(A).$$

Then we let

$$(2.3.1) h^S_\mu(\alpha, T) = \limsup_{n \to \infty} \frac{1}{n} H(\bigvee_{i=1}^n T^{-t_i} \alpha) \quad \text{and} \quad h^S_\mu(T) = \sup_{\alpha} h^S_\mu(\alpha, T),$$

the supremum being taken over all countable measurable partitions of finite entropy. $h^S_\mu(\alpha, T)$ is the asymptotic rate of information gain if a measurement $\alpha$ is made on a dynamical system at the times in the sequence $S$. $h^S_\mu(T)$ is an isomorphism invariant. Kushnirenko

1. calculated the $(2^n)$ entropy for some skew product maps on the torus;
2. used $(2^n)$ entropy to prove that the time-1 map of the horocycle flow on a two-dimensional orientable manifold of constant negative curvature is not isomorphic to its Cartesian square, although both have countable Lebesgue spectrum and entropy zero; and
3. proved that a system \((X, \mathcal{B}, \mu, T)\) has purely discrete spectrum if and only if \(h^S_\mu(T) = 0\) for every sequence \(S\).

D. Newton [146] (see also [127]) showed that for each sequence \(S\) there is a constant \(K(S)\) such that for every system \((X, \mathcal{B}, \mu, T)\), \(h^S_\mu(T) = K(S)h(T)\), unless \(K(S) = \infty\) and \(h(T) = 0\); if \(K(S) = 0\) and \(h(T) = \infty\), then \(h^S_\mu(T) = 0\). Thus sequence entropy is no better than ordinary entropy for distinguishing positive entropy systems up to isomorphism. The constant \(K(S)\) is defined as follows:

\[
K(S) = \lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n} \text{card} \bigcup_{i=1}^{n} [-k + t_i, k + t_i].
\]

Exercise 2.3.1. Calculate \(K(S)\) for the sequences \((i), (2^i), (3i), (i^2), (p_i)\) \((p_i = i\text{'th prime})\).

Saleski [165, 166] established connections among sequence entropies and various mixing properties. In particular, he showed that (1) \((X, \mathcal{B}, \mu, T)\) is strongly mixing if and only if for every \((\text{countable})\) measurable partition \(\alpha\) with \(H(\alpha) < \infty\), and every increasing sequence \(S\) in \(\mathbb{N}\), there is a subsequence \(S_0\) for which \(h_{S_0}(\alpha, T) = H(\alpha)\); and (2) \((X, \mathcal{B}, \mu, T)\) is weakly mixing if and only if for every countable measurable partition of finite entropy there is an increasing sequence \(S\) in \(\mathbb{N}\) for which \(h^S_\mu(\alpha, T) = H(\alpha)\).

These results were improved by Hulse [105] as follows. (1) \((X, \mathcal{B}, \mu, T)\) is strongly mixing if and only if every infinite sequence \(S\) has a subsequence \(S_0\) such that for every measurable partition \(\alpha\) of finite entropy, \(h_{S_0}(\alpha, T) = H(\alpha)\). (2) \((X, \mathcal{B}, \mu, T)\) is weakly mixing if and only if there exists an increasing sequence \(S\) in \(\mathbb{N}\) such that for every measurable partition \(\alpha\) of finite entropy, \(h^S_\mu(\alpha, T) = H(\alpha)\).

Garcia-Ramos [82] has recently clarified, in both the measure-theoretic and topological contexts, the connections among discrete spectrum, zero sequence entropy, and some new concepts of equicontinuity.

### 2.4 Topological sequence entropy

T. N. T. Goodman [87] defined sequence entropy for topological dynamical systems \(X, T\), where \(X\) is a compact Hausdorff space and \(T : X \to X\) is a continuous map. As in the definition of topological entropy, for any open cover \(\mathcal{U}\) of \(X\), \(N(\mathcal{U})\) denotes the minimum cardinality of subcovers of \(\mathcal{U}\) and \(H(\mathcal{U}) = \log N(\mathcal{U})\). For an open cover \(\mathcal{U}\) and sequence \(S = (t_i)\) of integers,

\[
h^S_{\text{top}}(\mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=1}^{n} T^{-t_i} \alpha\right),
\]

and \(h^S_{\text{top}}(T) = \sup_{\mathcal{U}} h^S_{\text{top}}(\mathcal{U}, T)\). Goodman gives equivalent Bowen-type definitions and establishes some basic properties. He proves the topological analogue of Newton’s result with the same constant \(K(S)\) in case \(X\) has finite covering dimension and shows that topological sequence entropy does not satisfy the variational principle—it is possible that \(h^S_{\text{top}}(T) > \sup\mu h^S_\mu(X, \mathcal{B}, \mu, T)\).

Huang, Li, Shao, and Ye [99] studied “sequence entropy pairs” and “weak mixing pairs”, in the spirit of the study of entropy pairs initiated by F. Blanchard in his investigation of topological analogues of the \(K\) property [24–26]. For sets \(U, V \subset X\) in a topological dynamical system \((X, T)\), we define \(\mathcal{N}(U, V) = \{n \geq 0 : U \cap T^{-n}V \neq \emptyset\}\). Then \((x_1, x_2) \in X \times X\) is defined to be a weak mixing pair if for any open neighborhoods \(U_1, U_2\) of \(x_1, x_2\), respectively, \(\mathcal{N}(U_1, U_1) \cap \mathcal{N}(U_1, U_2) \neq \emptyset\) (cf. [151]). Denote the set of weak mixing pairs by \(WM(X, T)\).

We say that \((x_1, x_2) \in X \times X \setminus \triangle\) is a sequence entropy pair if whenever \(U_1, U_2\) are closed mutually disjoint neighborhoods of \(x_1, x_2\), respectively, there exists a sequence \(S\) in \(\mathbb{N}\) such that \(h^S_\mu(\{U_1^c, U_2^c\}, T) > 0\). The system \((X, T)\) is said to have uniform positive sequence entropy (for short s.u.p.e.) if every pair \((x_1, x_2) \in X \times X \setminus \triangle\) is a sequence entropy pair.

Exercise 2.4.1. \((X, T)\) has s.u.p.e. if and only if for any cover \(\mathcal{U}\) of \(X\) by two non-dense open sets, there is a sequence \(S\) in \(\mathbb{N}\) such that \(h^S_\mu(\mathcal{U}, T) > 0\).
Huang, Li, Shao, and Ye show that \((X, T)\) is topologically weakly mixing if and only if \(WM(X, T) = X \times X \setminus \triangle\), and that topological weak mixing is also equivalent to s.u.p.e.

Huang, Shao and Ye [101] gave new proofs of many known results about measure-theoretic and topological sequence entropy and characterized mild mixing and rigidity in terms of sequence entropy. Furstenberg and Weiss [81] defined the stronger property of mild mixing: a finite measure-preserving system is mildly mixing if and only if its product with any finite measure-preserving system is ergodic. Furstenberg and Weiss [81] defined the stronger property of mild mixing:

A system is mildly mixing if and only if it has no proper rigid factors, i.e., no proper factors for which there exists a sequence \((n_i)\) such that \(T^{n_i} f \to f\) in \(L^2\) for all \(f \in L^2\). Zhang [188, 189] provided sequence entropy characterizations of mild mixing and relative mild mixing. An IP set is the set of all finite sums of a sequence of natural numbers. Huang, Shao, and Ye prove that an invertible measure-preserving system \((X, \mathcal{B}, \mu, T)\) (with \(\mu(X) = 1\)) is mildly mixing if and only if for every measurable set \(A\) with measure strictly between 0 and 1 and every IP set \(F\) there is an infinite sequence \(S \subset F\) such that \(h_S^\mu(\{B, B^c\}, T) > 0\); and that it is rigid if and only if there is an IP set \(F\) such that \(h_S^\mu(T) = 0\) for every infinite \(S \subset F\).

A topological dynamical system \((X, T)\) is called topologically mildly mixing if for every transitive system \((X', T')\) the product system \((X \times X', T \times T')\) is transitive [84, 102]. Huang, Shao, and Ye show that a topological dynamical system \((X, T)\) is topologically weakly mixing if and only if for each finite cover \(\mathcal{U}\) of \(X\) consisting of non-dense open sets there is an infinite sequence \(S\) in \(\mathbb{N}\) such that \(h_{\text{top}}(\mathcal{U}, T) > 0\); and that it is topologically mildly mixing if and only if for each finite cover \(\mathcal{U}\) of \(X\) consisting of non-dense open sets and IP set \(F\) there is an infinite sequence \(S \subset F\) such that \(h_{\text{top}}^S(T) > 0\).

### 2.5 Slow entropy

There have been several efforts to attach to a dynamical system measures of the rate of growth of \(H_\mu(\alpha_0^{n-1})\), in the measure-preserving case, or \(N(\mathcal{U}_0^{n-1})\), in the topological case, that are finer than the exponential growth rate, which gives the ordinary entropies. The quantities may be defined so as to be isomorphism invariants, and then they can be used to try to distinguish systems of zero or infinite entropy up to isomorphism. It seems that more has been done in the zero entropy case, where there are many classes of familiar and interesting examples still requiring further study. The definitions of **entropy dimension** and **power entropy** are designed especially to detect polynomial growth rates. See [93, p. 37 ff., p. 92 ff.] and [123].

F. Blume [28–31] defined a variety of rates of entropy convergence as follows. Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system and \(\mathcal{P}\) a fixed class of finite measurable partitions of \(X\). Let \(a = (a_n)\) be an increasing sequence of positive numbers with limit \(\infty\), and let \(c > 0\). Then \((X, \mathcal{B}, \mu, T)\) is said to be of type \(LS \geq c\) for \((a, \mathcal{P})\) if

\[
\limsup_{n \to \infty} \frac{H_\mu(\alpha_0^{n-1})}{a_n} \geq c \quad \text{for all } \alpha \in \mathcal{P},
\]

and of type \(LI \geq c\) for \((a, \mathcal{P})\) if

\[
\liminf_{n \to \infty} \frac{H_\mu(\alpha_0^{n-1})}{a_n} \geq c \quad \text{for all } \alpha \in \mathcal{P}.
\]

The types \(LS \leq c, LI \leq c, LS < \infty, LS = \infty, LI < \infty, LI = \infty, LS > 0, LI > 0\) are defined analogously.

When focusing on zero-entropy systems, it is natural to consider the class \(\mathcal{P}\) of all partitions of \(X\) into two sets of positive measure, since every zero-entropy system has a two-set generator [125]. Then the quantities defined above are invariants of measure-theoretic isomorphism. If \(h_\mu(X, T, \alpha) = \lim_{n \to \infty} H_\mu(\alpha_0^{n-1})/n = 0\), one should restrict attention to sequences \(a\) for which \(a_n/n \to 0\).

Pointwise entropy convergence rates can be defined by using the information function

\[
I_\alpha(x) = -\log \mu(\alpha(x)),
\]
2.5. Slow entropy

where \( \alpha(x) \) denotes the cell of \( \alpha \) to which the point \( x \) belongs. Then \((X,\mathcal{B},\mu,T)\) is defined to be of pointwise type \( \text{LS} \geq c \) for \((a,\mathcal{P})\) if

\[
\limsup_{n \to \infty} \frac{I_{\alpha^{n-1}}(x)}{a_n} \geq c \text{ a.e. for all } \alpha \in \mathcal{P},
\]

and the other pointwise types are defined analogously.

Here are some of Blume’s results in this area.

1. If \( a_n/n \to 0 \) and \( \mathcal{P} \) is the class of two-set partitions, then there are no aperiodic measure-preserving systems of types \( LI < \infty, \text{LS} < \infty, \text{LS} \leq c, \) or \( LI \leq c. \)

2. Suppose that \( g: [0, \infty) \to \mathbb{R} \) is a monotone increasing function for which

\[
\limsup_{n \to \infty} \int_1^\infty \frac{g(x)}{x^2} dx < \infty
\]

and \( a_n = g(\log_2 n) \) for all \( n = 1, 2, \ldots \). If \( T^k \) is ergodic for all \( k \in \mathbb{N} \), then \((X,\mathcal{B},\mu,T)\) is of type \( \text{LS} = \infty \) for \( a,\mathcal{P} \).

3. Every rank-one mixing transformation is of type \( \text{LS} > 0 \) for \( a_n = \log_2 n \) and \( \mathcal{P} \).

4. No rank-one system can have a convergence rate of type \( LI \geq c \) for any sequence \( a_n \) that grows faster than \( \log_2 n \): If \((X,\mathcal{B},\mu,T)\) is rank one, then there is a partition \( \alpha \in \mathcal{P} \) for which

\[
\liminf_{n \to \infty} \frac{H_\mu(\alpha^{n-1})}{\log_2 n} \leq 2.
\]

5. Every totally ergodic system is of pointwise type \( \text{LS} \geq 1 \) for \((\log_2 n),\mathcal{P})\).

**Exercise 2.5.1.** Show that every measure-preserving system is of pointwise type \( LI < \infty \) and \( \text{LS} < \infty \) for \((\langle n \rangle,\mathcal{P})\). (Hint: SMB.)

**Exercise 2.5.2.** Show that every \( K \)-system (all nontrivial factors have positive entropy) is of pointwise type \( \text{LS} > 0 \) and \( LI > 0 \) for \((\langle n \rangle,\mathcal{P})\).

In order to study the problem of the realizability of measure-preserving actions of amenable groups by diffeomorphisms of compact manifolds, Katok and Thouvenot [117] defined entropy-like measure-theoretic isomorphism invariants in terms of subexponential growth rates. One fixes a sequence \( (\alpha_n) \) of positive numbers increasing to infinity, or a family \( (a_n(t)) \), \( 0 < t < \infty \), such as \( a_n(t) = n^t \) or \( a_n(t) = e^{nt} \) (the latter will produce the usual entropy). We describe the case when the acting group is \( \mathbb{Z}^2 \), i.e. the action of two commuting measure-preserving transformations \( S \) and \( T \) on a probability space \((X,\mathcal{B},\mu)\).

Let \( \alpha \) be a finite measurable partition of \( X \) and \( \epsilon > 0 \). Let

\[
\alpha_n = \bigvee_{i=0}^{n-1} \bigvee_{j=0}^{n-1} S^{-i} \alpha \lor T^{-j} \alpha
\]

denote the refinement of \( \alpha \) by the action of \([0, n-1]^2 \subset \mathbb{Z}^2 \). The elements of each \( \alpha_n \) are given a fixed order, so that they become vectors of length \( n^2 \), and the \( \tilde{d} \) distance between two of them is defined to be the proportion of places at which they disagree. We define \( S(\alpha, n, \epsilon) \) to be the minimum number of elements of \( \alpha_n \) such that the union of \( \tilde{d}, \epsilon \) balls centered at them has measure at least \( 1 - \epsilon \).

Denote by \( A(\alpha, \epsilon) \) any one of

\[
\limsup_{n \to \infty} \frac{S(\alpha, n, \epsilon)}{a_n}, \quad \liminf_{n \to \infty} \frac{S(\alpha, n, \epsilon)}{a_n}, \quad \sup \{ t : \limsup_{n \to \infty} S(\alpha, n, \epsilon) a_n(t) > 0 \}, \quad \sup \{ t : \liminf_{n \to \infty} S(\alpha, n, \epsilon) a_n(t) > 0 \},
\]
and then define

\[(2.5.9) \quad A(\alpha) = \lim_{\epsilon \to 0} A(\alpha, \epsilon) \quad \text{and} \quad A = \sup_{\alpha} A(\alpha),\]

so that \(A\) is an isomorphism invariant.

The authors show that for an ergodic measure-preserving action by diffeomorphisms of \(\mathbb{Z}^2\) on a compact manifold (or even bi-Lipschitz homeomorphisms on a compact metric space) there is a sequence \((a_n)\) for which \(S(\alpha, n, \epsilon)\) grows at most exponentially. Then they give a cutting and stacking construction that alternates periodic and independent concatenations to produce an action which violates this growth condition and therefore cannot be realized as a smooth action preserving a Borel (not necessarily smooth) measure.

Hochman [96] used this type of argument to produce examples of infinite measure-preserving \(\mathbb{Z}^d\) actions for \(d > 1\) that are not isomorphic to actions by diffeomorphisms on a compact manifold preserving a \(\sigma\)-finite Borel measure. (Because of Krengel’s theorem [124] stating that every \(\sigma\)-finite measure-preserving action of \(\mathbb{Z}\) has a two-set generator, there cannot be such examples when \(d = 1\)—the given measure can be carried to a horseshoe inside a differentiable system.)

**Power entropy** is useful as a measure of complexity in zero-entropy systems if the relevant growth rate is polynomial. For example, if, as in Section 2.1, for a topological dynamical system \((X, T)\) one denotes by \(S(n, \epsilon)\) the maximum possible cardinality of an \((n, \epsilon)\)-separated set, one may define

\[(2.5.10) \quad h^+_\text{pow}(X, T) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log S(n, \epsilon)}{\log n} \quad \text{and} \quad h^-\text{pow}(X, T) = \lim_{\epsilon \to 0} \liminf_{n \to \infty} \frac{\log S(n, \epsilon)}{\log n},\]

and define \(h_{\text{pow}}\) to be their common value when they are equal. In this case

\[(2.5.11) \quad S(n, \epsilon) \sim n^{h_{\text{pow}}} \quad \text{for small} \ \epsilon.\]

If one replaces in the preceding definition the Bowen-Dinaburg metric

\[(2.5.12) \quad d_n(x, y) = \sup \{d(T^k x, T^k y) : k = 0, 1, \ldots, n - 1\}\]

\((d\ \text{denotes the metric on} \ X)\) by the Hamming metric

\[(2.5.13) \quad \overline{d}_n(x, y) = \frac{1}{n} \sum_{k=0}^{n-1} d(T^k x, T^k y),\]

the resulting entropies are called **modified power entropies**. The recent paper [88] shows the following (see also Section 3.11).

1. Neither power entropy nor modified power entropy satisfies a variational principle with respect to any real-valued isomorphism invariant defined for measure-preserving systems.

2. Recall that a minimal system \((X, T)\) is called **almost automorphic** if it is an almost one-to-one extension of a minimal equicontinuous system \((Y, S)\): there is a factor map \(\pi : X \to Y\) such that \(\Omega = \{y \in Y : \text{card} \ \pi^{-1}(y) = 1\}\) is nonempty (in which case it is residual [176]—see Exercise 2.5.3). Such a system \((X, T)\) has a unique invariant measure, \(\mu\). If \(\mu(\Omega) = 1\), the extension is called **regular**. It is proved in [88] that the modified power entropy cannot tell the difference between equicontinuous and almost automorphic systems, since it is 0 for both—and in fact \(\sup_n \overline{S}(n, \epsilon) < \infty\) for all \(\epsilon\).

3. If \((X, T)\) is almost automorphic (but not equicontinuous), then there is \(\epsilon > 0\) such that \(\sup_n S(n, \epsilon) = \infty\). (Cf. Theorem 3.6.1.) Consequently “power entropy” redefined for some other growth scale can distinguish between equicontinuous and almost automorphic systems.
Exercise 2.5.3. Prove the statement from above that if \( \pi : X \rightarrow Y \) is a factor map between minimal systems for which there exists a singleton fiber, then \( \{ y \in Y : \text{card } \pi^{-1}y = 1 \} \) is residual. (Hint: Show that the function \( g(y) = \sup \{ d_X(x, x') : \pi x = \pi x' \} \) is upper semicontinuous on \( Y \).

2.6 Entropy dimension

In 1997 de Carvalho [63] proposed an invariant for topological dynamical systems that is possibly finer than topological entropy. For a topological dynamical system \((X, T)\), define

\[
d_{\text{top}}(X, T) = \inf \left\{ s > 0 : \sup_{U} \limsup_{n \to \infty} \frac{1}{n^s} \log N(U_0^{n-1}) = 0 \right\},
\]

the supremum being taken over all open covers \( U \) of \( X \). Recall that \( U_0^{n-1} = U \lor T^{-1}U \lor \cdots \lor T^{-n+1}U \), and \( N(U) \) denotes the minimal number of elements in any subcover of \( U \).

The author also defines a version for measure-theoretic systems \((X, \mathcal{B}, \mu, T)\):

\[
d_{\text{mt}}(X, T) = \inf \left\{ s > 0 : \sup_{\alpha} \limsup_{n \to \infty} \frac{1}{n^s} H_{\mu}(\alpha_0^{n-1}) = 0 \right\},
\]

the supremum being taken over all finite measurable partitions \( \alpha \) of \( X \).

The author shows that if the topological entropy is finite, then the topological entropy dimension is less than or equal to 1; while if the topological entropy is finite and positive, then the topological entropy dimension equals 1. Thus \( d_{\text{top}}(X, T) \) is of interest mainly for zero-entropy systems and could perhaps be used to show that some zero-entropy systems are not topologically conjugate.

Related definitions were given by Dou, Huang, and Park in 2011 [66]. The upper entropy dimension of \((X, T)\) with respect to an open cover is

\[
\mathcal{D}(U) = \inf \left\{ s \geq 0 : \limsup_{n \to \infty} \frac{1}{n^s} \log N(U_0^{n-1}) = 0 \right\}
\]

\[
= \sup \left\{ s \geq 0 : \limsup_{n \to \infty} \frac{1}{n^s} \log N(U_0^{n-1}) = \infty \right\}.
\]

The lower entropy dimension of \((X, T)\) with respect to an open cover \( U \) is defined by replacing \( \limsup \) by \( \liminf \). The upper and lower entropy dimensions of \((X, T)\) are defined to be

\[
\mathcal{D}(X, T) = \sup_{U} \mathcal{D}(U) \quad \text{and} \quad \underline{D}(X, T) = \inf_{U} \mathcal{D}(U),
\]

respectively. These are invariants of topological conjugacy. When they are equal, they are denoted by \( D(X, T) \) and called the entropy dimension of \((X, T)\). If \( h_{\text{top}}(X, T) > 0 \), then again \( D(X, T) = 1 \).

Ferenczi and Park [77] defined another concept of measure-theoretic entropy dimension. Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system, \( \alpha \) a finite measurable partition of \( X \), and \( \epsilon > 0 \). The \((d_n, \epsilon)\)-ball around a point \( x \in X \) is

\[
B(x, n, \epsilon) = \{ y \in X : d(\alpha_0^{n-1}(y), \alpha_0^{n-1}(x)) < \epsilon \},
\]

\( K(n, \epsilon) \) is the minimum cardinality of a set of \((d_n, \epsilon)\)-balls that covers a set of measure at least \( 1 - \epsilon \). Because of the Shannon-McMillan-Breiman Theorem, the entropy of the system coded by the partition \( \alpha \) is given by

\[
h_{\mu}(\alpha, T) = \lim_{\epsilon \to 0^+} \lim_{n \to \infty} \frac{\log K(n, \epsilon)}{n};
\]

see [115] for the statement in metric spaces.
Thinking in terms of dimension theory suggests the definitions
\begin{equation}
\mathcal{D}(\alpha, \epsilon) = \sup \{ s \in [0, 1] : \limsup_{n \to \infty} \frac{\log K(n, \epsilon)}{n^s} > 0 \},
\end{equation}
\begin{equation}
\mathcal{D}(\alpha) = \lim_{\epsilon \to 0} \mathcal{D}(\alpha, \epsilon), \quad \text{and} \quad \mathcal{D}(X, B, \mu, T) = \sup_{\alpha} \mathcal{D}(\alpha).
\end{equation}
The versions with \( D \) use \( \lim \inf \) instead of \( \lim \sup \).

\( \mathcal{D} \) and \( D \) are called the upper and lower entropy dimensions of \((X, B, \mu, T)\) If they are equal, they are denoted by \( D(X, B, \mu, T) \) and called the entropy dimension. The definitions extend easily to measure-preserving actions of other groups, especially \( \mathbb{Z}^d \).

Ahn, Dou, and Park [3] showed that while the topological entropy dimension of a topological dynamical system \((X, T)\) always dominates its measure-theoretic entropy dimension for any invariant measure, the variational principle does not hold: For each \( t \in (0, 1) \) they construct a uniquely ergodic system \((X, T)\) with the measure-theoretic entropy dimension \( D(X, B, \mu, T) = 0 \) for the unique invariant measure \( \mu \), while the topological entropy dimension \( D(X, T) = t \).

See [128] and its references for some further developments about entropy dimensions.

### 2.7 Permutation entropy

The permutation entropy of a real-valued time series was introduced in [16] and studied in the dynamical setting, especially for interval maps, in [15]. It may be regarded as a measure of the amount of information obtained by observing the ordering of an initial orbit piece, and thus of the complexity of the orderings or orbits (or sequences of measurements). Instead of counting the number of \( n \)-blocks that arise from coding orbits with respect to a partition, one counts the number of different order patterns that can appear (or computes the Shannon entropy, with respect to an invariant measure, of the partition according to order patterns).

Let \( T : X \to X \) be a continuous map on a compact interval \( X \subset \mathbb{R} \). For a fixed \( n \in \mathbb{N} \) one partitions \( X \) according to the order patterns of the first \( n \) points in the orbit as follows. Denote \( x_0 = T^k x \) for \( k = 0, 1, \ldots, n-1 \). We define a permutation \( \tau(x) = (k_1 k_2 \ldots k_{n-1}) \) of \((0, 1, \ldots, n-1)\) by
\begin{equation}
x_{k_1} \leq x_{k_2} \leq \cdots \leq x_{k_{n-1}}
\end{equation}
and, to break ties, \( \tau(i) < \tau(i+1) \) in case \( x_i = x_{i+1} \).

Denote by \( \rho_n \) the partition of \( X \) into the nonempty sets \( X_\tau \) among all the \( n! \) permutations \( \tau \).

Let \( \mu \) be an invariant probability measure for \((X, T)\). The topological and measure-theoretic permutation entropies of order \( n \) are defined to be
\begin{equation}
\begin{aligned}
&h_{top}^P(n) = \frac{1}{n-1} \log \text{card} \rho_n \quad \text{and} \\
&h_{\mu}^P(n) = \frac{1}{n-1} \sum_{A \in \rho_n} -\mu(A) \log \mu(A).
\end{aligned}
\end{equation}

**Theorem 2.7.1.** [15] If \( T : X \to X \) is a piecewise monotone map on an interval \( X \subset \mathbb{R} \), then the limits of the permutation entropies of order \( n \) exist as \( n \) tends to infinity and equal the usual entropies:
\begin{equation}
\lim_{n \to \infty} h_{top}^P(n) = h_{top} \quad \text{and} \quad \lim_{n \to \infty} h_{\mu}^P(n) = h_{\mu}(X, T).
\end{equation}

This theorem, along with its extensions mentioned just below, is analogous to Theorems 4.2.10 and 4.3.3, providing further evidence for a topological version of Theorem 4.3.4.

The definitions extend to arbitrary systems once one assigns to the elements of a finite partition distinct real numbers or fixes a real or vector-valued function (observable) on \( X \) to follow along orbits. To obtain
complete information (and reach the ordinary entropy) one should consider a generating sequence of partitions or observables. See [8, 92, 118, 120] and their references.

The assignment of order patterns to strings from a totally ordered set is a potentially many-to-one function. For example, the observation strings 321233 and 421333 both determine the permutation $\tau = (324561)$. Haruma and Nakajima [92] clarified and extended previous results by showing (by counting) that for finite-state stationary ergodic processes the permutation entropy coincides with the Kolmogorov-Sinai entropy.

For a recent survey of permutation entropies and their relation to ordinary entropies, see [119]. These quantities may be useful in numerical studies of actual systems, since they can be calculated from observed orbits, their behavior as a parameter determining the system changes may reflect the actual entropy even if slow to approach the limit, and since one observes only order relations they may reduce computation time.

2.8 Independence entropy

Louidor, Marcus, and Pavlov [138] introduce a measure of how freely one may change symbols at various places in sequences or higher-dimensional configurations in a symbolic dynamical system and still remain in the system. Part of the motivation comes from practical questions in coding, where one may want to insert check bits without violating desirable constraints on the structure of signals, and another part from questions about the entropy of higher-dimensional subshifts, even shifts of finite type.

We give the definitions for a one-dimensional subshift $(X, \sigma)$ over a finite alphabet $\mathcal{A}$; the extensions to $\mathbb{Z}^d$ subshifts are immediate. The set of nonempty subsets of $\mathcal{A}$ is denoted by $\hat{\mathcal{A}}$. For a configuration $\hat{x} \in \hat{\mathcal{A}}^\mathbb{Z}$, the set of fillings of $\hat{x}$ is

\begin{equation}
\Phi(\hat{x}) = \{ x \in \mathcal{A}^\mathbb{Z} : \text{for all } k \in \mathbb{Z}, x_k \in \hat{x}_k \},
\end{equation}

and the multi-choice shift corresponding to $(X, \sigma)$ is

\begin{equation}
\hat{X} = \{ \hat{x} \in \hat{\mathcal{A}}^\mathbb{Z} : \Phi(\hat{x}) \subseteq X \}.
\end{equation}

Thus $\hat{X}$ consists of all the sequences on $\hat{\mathcal{A}}$ that display the freedom to flex sequences in $X$: for those $k \in \mathbb{Z}$ for which $\hat{x}_k$ is not a singleton, we may freely choose symbols from the sets $\hat{x}_k$ and generate sequences that are always in the original subshift $X$.

**Exercise 2.8.1.** Describe $\hat{X}$ when $X$ is the golden mean shift of finite type (no adjacent 1s).

**Exercise 2.8.2.** Show that if $X$ is a shift of finite type, then so is $\hat{X}$.

It can be shown that in dimension $d = 1$, if $X$ is sofic, then so is $\hat{X}$ [155]; but in higher dimensions the question is open.

To define independence entropy, we consider finite blocks $\hat{B}$ on the alphabet $\hat{\mathcal{A}}$, considered as elements of $\hat{\mathcal{A}}^{|\hat{B}|}$, and their fillings

\begin{equation}
\Phi(\hat{B}) = \{ x \in \mathcal{A}^{|\hat{B}|} : x_k \in \hat{x}_k \text{ for all } k = 0, 1, \ldots, |\hat{B}| - 1 \},
\end{equation}

and define the independence entropy of $(X, \sigma)$ to be the exponential growth rate

\begin{equation}
\h_{\text{ind}}(X, \sigma) = \lim_{n \to \infty} \log \max \left\{ \frac{\text{card} \Phi(\hat{B}) : \hat{B} \in \mathcal{L}_n(\hat{X}, \sigma)}{n} \right\}
\end{equation}

of the maximal number of different fillings of $n$-blocks on the alphabet $\hat{\mathcal{A}}$. It is a measure of the maximal independence among entries in $n$-blocks, rather than of the number of possible $n$-blocks.

**Exercise 2.8.3.** Show that $\h_{\text{ind}}(X) \leq \h_{\text{top}}(X)$ for every subshift $(X, \sigma)$. 

Exercise 2.8.4. Show that the independence entropy of the golden mean shift of finite type is \( \log 2/2 \).

Louidor and Marcus [139] described a natural way to form the axial product of \( d \) one-dimensional shift spaces \( X_1, \ldots, X_d \) on a common alphabet \( \mathcal{A} \). It is the set \( X_1 \otimes \cdots \otimes X_d \) with all configurations \( x \in \mathcal{A}^{Z_d} \) such that for every \( i = 1, \ldots, d \) and every choice of \( n_1, \ldots, n_i-1, n_i+1, \ldots, n_d \in \mathbb{Z} \), the “row”

\[
y_k = x(n_1, \ldots, n_i-1, k, n_i+1, \ldots, n_d)
\]
is in \( X_i \), together with the \( \mathbb{Z}_d \) action by the shifts in each coordinate. If all \( X_i = X \), the resulting product is the \( d \)-fold axial power \( X^{\otimes d} \) of \( X \).

**Theorem 2.8.1.** [138] For one-dimensional subshifts \( (X, \sigma) \),

1. \( h_{\text{ind}}(X^{\otimes d}) = h_{\text{ind}}(X) \)
2. \( h_{\text{top}}(X^{\otimes d}) \) is nonincreasing in \( d \) and hence has a limit, \( h_\infty(X, \sigma) \geq h_{\text{ind}}(X, \sigma) \).

**Theorem 2.8.2.** [141] For every one-dimensional subshift \( (X, \sigma) \), \( h_\infty(X, \sigma) = h_{\text{ind}}(X, \sigma) \).

### 2.9 Sofic, Rokhlin, naïve entropies

Lewis Bowen [33, 35, 36] investigated the possibility of extending the Kolmogorov-Ornstein results (two Bernoulli shifts are measure-theoretically isomorphic if and only if they have the same entropy) to possibly infinite alphabets and possibly nonamenable acting groups. Let \((\Omega, \mathcal{F}, P)\) be a standard Borel probability space and \( G \) a countable discrete group. \( G \) acts on the product probability space \((\Omega, \mathcal{F}, P)^G\) by \( gx(h) = x(g^{-1} h) \) for all \( x \in \Omega^G \) and \( g, h \in G \). First one needs a definition of entropy. If there is a finite or countable subset \( \Omega_0 \subset \Omega \) of full measure, then the entropy of the measure \( P \) is defined to be

\[
H(P) = -\sum_{\omega \in \Omega_0} P(\omega) \log P(\omega),
\]
and otherwise \( H(P) = \infty \). The theorems of Kolmogorov and Ornstein state that two Bernoulli shifts \((\Omega, \mathcal{F}, P)^\mathbb{Z}\) and \((\Omega', \mathcal{F}', P')^\mathbb{Z}\) are measure-theoretically isomorphic if and only if \( H(P) = H(P') \).

Ornstein and Weiss [148] extended these results to countably infinite amenable groups. They also noted that there were some potential problems for nonamenable groups, by showing that for Bernoulli actions of the free group on two generators the 2-shift factors onto the 4-shift (each with the product measure of its uniform measure).

L. Bowen [33] gave a definition of entropy for actions by sofic groups (defined by Gromov [89] and Weiss [182]) and showed that it is an isomorphism invariant for Bernoulli actions of sofic groups for countable alphabets with finite entropies, and a complete isomorphism invariant for such Bernoulli actions by countably infinite linear groups (groups of invertible matrices with entries from a field). In [36], using his \( f \)-invariant (a special case of the sofic entropy), he showed that two Bernoulli actions by a free group on a finite number of generators on countable alphabets of finite entropy are isomorphic if and only if the alphabet entropies are equal.

A sofic group [89, 182] is a countable group \( G \) that can be approximated by finitely many permutations acting on subsets, as described in the following. (Recall that an amenable group is one that can be approximated by nearly invariant subsets.) For \( \epsilon > 0 \) and \( F \subset G \), a map \( \phi: G \to S_n = \text{the group of permutations of } \{1, \ldots, n\} \) is called an \((F, \epsilon)\)-approximation to \( G \) if the set

\[
V(F) = \{ v \in \{1, \ldots, n\} : \phi(f_1)\phi(f_2)v = \phi(f_1f_2)v \text{ for all } f_1, f_2 \in F \}
\]
satisfies

\[
\text{card } V(F) \geq (1-\epsilon)n.
\]
2.9. Sofic, Rokhlin, naïve entropies

Then $G$ is defined to be sofic if there is a sequence of $(F_k, \epsilon_k)$-approximations to $G$ such that $\epsilon_k \rightarrow 0$, $F_k \subset F_{k+1}$ for all $k$, and $\bigcup_{k=1}^\infty F_k = G$. Amenable groups and residually finite groups (the intersection of all the finite index normal subgroups is trivial) are sofic. It seems to be an open question whether every countable group is sofic.

The sofic entropy of a measure-preserving action of a sofic group $G$ on a probability space $(X, \mathcal{B}, \mu)$ is defined with respect to a fixed sofic approximation $\Phi = \{\phi_k\}$ of $G$ and a generating finite or countable ordered measurable partition $\alpha$ of $X$. According to L. Bowen, it is a measure of the exponential growth rate of the number of partitions of $\{1,\ldots,n_k\}$ that approximate $\alpha$. A main theorem is that the definition does not depend on the choice of the partition $\alpha$.

Let $\alpha = \{A_1, A_2, \ldots\}$ be an ordered finite measurable partition of $X$, $\phi : G \rightarrow S_n$ a map, $\beta = \{B_1, B_2, \ldots\}$ a partition of $\{1,\ldots,n\}$, $\nu$ the uniform probability measure on $\{1,\ldots,n\}$, and $F \subset G$ a finite subset. For any function $g : F \rightarrow \mathbb{N}$, let

\[(2.9.4) \quad A_g = \cap_{f \in F} A_{g(f)} \quad \text{and} \quad B_g = \cap_{f \in F} \phi(f) B_{g(f)}.
\]

Define the $F$-distance between $\alpha$ and $\beta$ to be

\[(2.9.5) \quad d_F(\alpha, \beta) = \sum_{g : F \rightarrow \mathbb{N}} |\mu(A_g) - \nu(B_g)|.
\]

Then for $\epsilon > 0$ $AP(\phi, \alpha; F, \epsilon)$ is defined to be the set of all ordered partitions $\beta$ of $\{1,\ldots,n\}$ with the same number of atoms as $\alpha$ for which $d_F(\alpha, \beta) \leq \epsilon$.

Fix the sofic approximation $\Phi = \{\phi_k\}$ of $G$. For a finite ordered measurable partition $\alpha$ of $X$, $\epsilon > 0$, and finite $F \subset G$, define

\[(2.9.6) \quad H(\Phi, \alpha; F, \epsilon) = \limsup_{k \rightarrow \infty} \frac{1}{n_k} \log \text{card} \ AP(\phi_k, \alpha; F, \epsilon),
\]

\[h(\Phi, \alpha) = \inf_{\text{finite } F \subset G} H(\Phi, \alpha; F).
\]

The definitions extend also to countable partitions $\alpha$.

**Theorem 2.9.1.** [33] If $G$ is a countable sofic group with sofic approximation $\Phi = \{\phi_k\}$ acting by measure-preserving transformations on a probability space $(X, \mathcal{B}, \mu)$, then $h(\Phi, \alpha) = h(\Phi, \beta)$ for any two generating partitions $\alpha, \beta$ of $X$.

If $G, (X, \mathcal{B}, \mu), \Phi$ are as in the Theorem and there is a generating partition $\alpha$ (the smallest $\sigma$-algebra containing all sets $g\alpha, g \in G$ is, up to sets of measure 0, the full $\sigma$-algebra $\mathcal{B}$) with $H_\mu(\alpha) < \infty$, then $h_{\text{sofic}}(X, \mathcal{B}, \mu, G, \Phi)$ is defined to be $h(\Phi, \alpha)$; otherwise it is undefined. In principle $h_{\text{sofic}}(X, \mathcal{B}, \mu, G, \Phi) \in [−\infty, \infty]$. Notice that there is not a single sofic entropy of the system, but a family depending on the sofic approximation $\Phi$ to $G$.

L. Bowen has also proved [35] that if $G$ is amenable, then for every sofic approximation $\Phi$ the sofic entropy $h_{\text{sofic}}(X, \mathcal{B}, \mu, G, \Phi)$ coincides with the ordinary entropy [148]. Greatly generalizing the Ornstein-Weiss example mentioned above, L. Bowen also showed [34] that any two nontrivial Bernoulli actions by a countable group that contains a nonabelian free subgroup are weakly isomorphic—each is a measure-preserving factor of the other.

Kerr and Li (see [122] and its references) have studied sofic entropy for both measure-preserving and topological sofic group actions as well as the relation of a kind of combinatorial independence with mixing properties and positive entropy. (Maybe there is a connection with the ideas in [139]?)

P. Burton [42] picked up on a suggestion of L. Bowen and pursued a definition of entropy for actions of nonamenable groups that at first seems mybe useless, because for such group actions it is always
either 0 or \( \infty \). However, the quantity does thus serve to divide systems into two disjoint classes, and its relation to sofic entropy is of interest. There are both measure-theoretic (already defined by L. Bowen) and topological (due to P. Burton) versions.

Let \((X, B, \mu, G)\) be a measure-preserving system, with \((X, B, \mu)\) a standard probability space and \(G\) a countable discrete group. For a finite measurable partition \(\alpha\) of \(X\) and finite \(F \subset G\), \(\alpha^F = \bigvee_{g \in F} (g\alpha)\). Then one may define

\[
\begin{align*}
  h_{\text{tv}}(X, \mu, \alpha) &= \inf \left\{ \frac{H_\mu(\alpha^F)}{\text{card } F} : F \text{ a finite nonempty subset of } G \right\} \quad \text{and} \\
  h_{\text{tv}}(X, \mu, \alpha) &= \sup \{ h_{\text{tv}}(X, \mu, \alpha) : \alpha \text{ a finite partition of } X \}. 
\end{align*}
\]

Similarly in the topological case. If \(G\) is a countable group acting by homeomorphisms on a compact metric space \(X\), then for an open cover \(\mathcal{U}\) of \(X\) and finite \(F \subset G\) one defines \(\mathcal{U}^F = \bigvee_{g \in F} (g\mathcal{U})\).

\[
\begin{align*}
  h_{\text{tv}}(X, G, \mathcal{U}) &= \inf \left\{ \frac{\log \#(\mathcal{U}^F)}{\text{card } F} : F \text{ a finite nonempty subset of } G \right\} \quad \text{and} \\
  h_{\text{tv}}(X, G) &= \sup \{ h_{\text{tv}}(X, G, \mathcal{U}) : \mathcal{U} \text{ an open cover of } X \}. 
\end{align*}
\]

L. Bowen proved that if \(G\) is nonamenable, then \(h_{\text{tv}}(X, \mu, G)\) is always either 0 or \(\infty\), and P. Burton noted that for \(G\) nonamenable also always \(h_{\text{tv}}(X, G)\) is 0 or \(\infty\).

Verifying conjectures of L. Bowen, P. Burton proved that for finitely generated \(G\), if \(h_{\text{tv}}(X, G) = 0\), then the topological sofic entropy [121] of \((X, G)\) is nonnegative for every sofic approximation \(\Phi\) for \(G\). The analogous statement in the measure-preserving case has been proved by Abert, Austin, Seward, and Weiss (see [42]).

Seward [172,173] defined the Rokhlin entropy of a measure-preserving system \((X, B, \mu, G)\), with \((X, B, \mu)\) a standard probability space and \(G\) a countably infinite group, to be

\[
\begin{align*}
  h_{\text{Rok}}(X, B, \mu, G) &= \inf \{ H_\mu(\alpha) : \alpha \text{ a countable Borel generating partition} \}. 
\end{align*}
\]

Rokhlin [163] proved that for a measure-preserving \(\mathbb{Z}\) action the Rokhlin entropy coincides with the ordinary Kolmogorov-Sinai entropy: \(h_{\text{Rok}}(X, B, \mu, G) = h(X, B, \mu, \mathbb{Z})\). Now it is known that the Rokhlin entropy and ordinary entropy coincide for all free ergodic measure-preserving actions of amenable groups. With this definition Seward was able to extend Krieger’s finite generator theorem to actions of arbitrary countable groups.

**Theorem 2.9.2.** [172,173] Let \(G\) be a countably infinite group, suppose that the system \((X, B, \mu, G)\), with \((X, B, \mu)\) a non-atomic standard probability space, is ergodic, and \(p = (p_i)\) is a finite or countable probability vector. If

\[
\begin{align*}
  h_{\text{Rok}}(X, B, \mu, G) &< H(p) = \sum -p_i \log p_i, 
\end{align*}
\]

then there is a generating partition \(\alpha = \{ A_i \} \) with \(\mu(A_i) = p_i\) for all \(i\). In particular, if \(h_{\text{Rok}}(X, B, \mu, G) \leq \log k\) for some integer \(k\), then there is a generator \(\alpha\) with \(k\) elements, and \((X, B, \mu, G)\) embeds in the \(k\)-shift \((\{0,1,\ldots,k-1\}^G, G)\).

**2.10 Kolmogorov complexity**

The (Kolmogorov) complexity \(K(w)\) of a finite sequence \(w\) on a finite alphabet is defined to be the length of the shortest program that when input to a fixed universal Turing machine produces output \(w\) (or at least a coding of \(w\) by a block of 0’s and 1’s). For a topological dynamical system \((X, T)\), open cover \(\mathcal{U} = \{ U_0, \ldots, U_{r-1} \}\) of \(X\), \(x \in X\), and \(n \geq 1\), we consider the set of codings of the initial \(n\) points in the orbit of \(x\) according to the cover \(\mathcal{U}\): let \(C(x; n)\) be the set of \(n\)-blocks \(w\) on \(\{0, \ldots, r - 1\}\) such that
T^j x \in U_{w_j}, j = 0, \ldots, n. Then we define the upper and lower complexity of the orbit of a point \( x \in X \) to be

\[ K^+(x, T) = \sup_{\mathcal{U}} \limsup_{n \to \infty} \min \left\{ \frac{K(w)}{n} : w \in C(x, n) \right\} \]

and

\[ K^-(x, T) = \sup_{\mathcal{U}} \liminf_{n \to \infty} \min \left\{ \frac{K(w)}{n} : w \in C(x, n) \right\}. \]

**Theorem 2.10.1** (Brudno [41], White [183, 184]). If \( \mu \) is an ergodic invariant measure on \((X, T)\), then

\[ K^+(x, T) = K^-(x, T) = h_{\mu}(X, T) \text{ a.e. } d\mu(x). \]

Brudno showed that \( K^+(x, T) = h_{\mu}(X, T) \) a.e. for every ergodic measure \( \mu \) on \((X, T)\). He also noted that attempting to use finite measurable partitions \( \alpha \) instead of open covers,

\[ \sup_{\alpha} \limsup_{n \to \infty} \frac{K(\alpha^{-1}_n(x))}{n}, \]

gives \( \infty \) for all nonperiodic points \( x \in X \).

White also explains how ingredients of the proofs, some going back to Brudno and even Kolmogorov, produce a universal coding scheme: an effective (achievable by a Turing machine) algorithm for coding any finite-state ergodic stationary source that almost surely achieves a long-term data compression rate equal to the entropy of the source (the best possible). Lempel-Ziv, Wyner-Ziv, Ornstein-Shields, etc. have also produced universal coding schemes.
Part 3

Counting patterns

A complexity function \( p_x(n) \) counts the number of patterns of “size” \( n \) that appear in an object \( x \) under investigation. One of the simplest situations (one might suppose) is that of a one-dimensional sequence \( u \) on a finite alphabet \( A \). If \( u \) is a sequence or bisequence, the complexity function of \( u \), denoted \( p_u \), maps \( n \) to the number of blocks of length \( n \) that appear in \( u \). If \( X \) is a subshift, then \( p_X(n) \) is the number of blocks of length \( n \) that appear in \( \mathcal{L}(X) \). In higher-dimensional symbolic dynamical systems one may count the number of configurations seen in rectangular regions, and in tilings one may count the number of patches of tiles of a fixed size that are equivalent under translations, or, if preferred, under translations and rotations. The asymptotic exponential growth rate of the complexity function,

\[
\limsup_{n \to \infty} \frac{\log p_x(n)}{n},
\]

is a single number that measures the complexity of \( x \) in one sense, while the function \( p_x \) itself is a precise measurement of how the complexity or richness of the object grows with size. There is a huge literature on complexity functions of various kinds for various structures; see for example [4, 19, 75, 78, 137]. Here we look at a few representative examples.

**Exercise 3.0.1.** [75] Show that the growth rate of the complexity function \( p_X \) of a subshift \((X, \sigma)\) is an invariant of topological conjugacy by proving that if \((X, \sigma)\) and \((Y, \sigma)\) are topologically conjugate subshifts on finite alphabets, then there is a constant \( c \) such that

\[
p_X(n-c) \leq p_Y(n) \leq p_X(n+c) \text{ for all } n > c.
\]

**3.1 The complexity function in one-dimensional symbolic dynamics**

Let \( u \) be a one or two-sided infinite sequence on a finite alphabet \( A \), and let \( p_u(n) \) denote the number of \( n \)-blocks in \( u \). Since every block continues in at least one direction, \( p_u(n+1) \geq p_u(n) \) for all \( n \).

**Exercise 3.1.1.** Find the complexity functions of the bisequence \( u = \ldots 121232121 \ldots \) and the one-sided sequence \( v = 321212121 \ldots \).

Hedlund and Morse [143, Theorems 7.3 and 7.4] showed that a two-sided sequence \( u \) is periodic if and only if there is a \( k \) such that \( p_u(k+1) = p_u(k) \), equivalently if and only if there is an \( n \) such that \( p_u(n) \leq n \).

**Exercise 3.1.2.** Show that for one-sided sequences \( u \) the following conditions are equivalent: (1) there is an \( n \) such that \( p_u(n) \leq n \); (2) there is a \( k \) such that \( p_u(k+1) = p_u(k) \); (3) \( u \) is eventually periodic; (4) \( p_u \) is bounded.

(Hint: For (2) implies (3), note that each \( k \)-block in \( u \) must have a unique right extension to a \((k+1)\)-block, and that some \( k \)-block must appear at least twice in \( u \).)
Exercise 3.1.3. Show that for a two-sided sequence \( u \), if there is an \( n \) such that \( p_u(n) \leq n \), then \( u \) is periodic.

3.2 Sturmian sequences

Hedlund and Morse [144] defined Sturmian sequences as those that have the smallest possible complexity among non-eventually-periodic sequences.

Definition 3.2.1. A sequence \( u \) is called Sturmian if it has complexity \( p_u(n) = n + 1 \) for all \( n \).

If \( u \) is Sturmian, then \( p_u(1) = 2 \). This implies that Sturmian sequences are over a two-letter alphabet. For the duration of this discussion on Sturmian systems, we fix the alphabet \( \mathcal{A} = \{0, 1\} \).

Exercise 3.2.1. The Fibonacci substitution is defined by:

\[
\phi : 0 \mapsto 01, 1 \mapsto 0.
\]

The fixed point \( f = 01001001010010010100100101 ... \) of the Fibonacci substitution is called the Fibonacci sequence. Show that \( f \) is a Sturmian sequence.

Definition 3.2.2. A set \( S \) of blocks is balanced if for any pair of blocks \( u, v \) of the same length in \( S \), \(|u| - |v| \leq 1\), where \(|u|\) is the number of occurrences of 1 in \( u \) and \(|v|\) is the number of occurrences of 1 in \( v \).

It immediately follows that if a sequence \( u \) is balanced and not eventually periodic then it is Sturmian. This is a result of the fact that if \( u \) is aperiodic, then \( p_u(n) \geq n + 1 \) for all \( n \), and if \( u \) is balanced then \( p_u(n) \leq n + 1 \) for all \( n \). In fact, it can be proved that a sequence \( u \) is balanced and aperiodic if and only if it is Sturmian [137]. Furthermore, it immediately follows that any shift of a Sturmian sequence is also Sturmian.

Sturmian sequences also have a natural association to lines with irrational slope. To see this, we introduce the following definitions.

Definition 3.2.3. Let \( \alpha \) and \( \beta \) be real numbers with \( 0 \leq \alpha, \beta \leq 1 \). We define two infinite sequences \( x_{\alpha,\beta} \) and \( x'_{\alpha,\beta} \) by

\[
(x_{\alpha,\beta})_n = [\alpha(n+1) + \beta] - [\alpha n + \beta]
\]

\[
(x'_{\alpha,\beta})_n = [\alpha(n+1) + \beta] - [\alpha n + \beta]
\]

for all \( n \geq 0 \). The sequence \( x_{\alpha,\beta} \) is the lower mechanical sequence and \( x'_{\alpha,\beta} \) is the upper mechanical sequence with slope \( \alpha \) and intercept \( \beta \).

The use of the words slope and intercept in the above definitions stems from the following graphical interpretation. Consider the line \( y = \alpha x + \beta \). The points with integer coordinates that sit just below this line are \( F_n = (n, [\alpha n + \beta]) \). The straight line segment connecting two consecutive points \( F_n \) and \( F_{n+1} \) is horizontal if \( x_{\alpha,\beta} = 0 \) and diagonal if \( x_{\alpha,\beta} = 1 \). Hence, the lower mechanical sequence can be considered a coding of the line \( y = \alpha x + \beta \) by assigning to each line segment connecting \( F_n \) and \( F_{n+1} \) a 0 if the segment is horizontal and a 1 if the segment is diagonal. Similarly, the points with integer coordinates that sit just above this line are \( F'_n = (n, [\alpha n + \beta]) \). Again, we can code the line \( y = \alpha x + \beta \) by assigning to each line segment connecting \( F'_n \) and \( F'_{n+1} \) a 0 if the segment is horizontal and a 1 if the segment is diagonal. This coding yields the upper mechanical sequence [137].
3.2. Sturmian sequences

The sequence \( x_{\alpha,\beta} \) codes the forward orbit of \( \beta \in [0,1) \) under the translation \( x \to x + \alpha \mod 1 \) with respect to the partition \( A = [0,1-\alpha), A^c = [1-\alpha,1) \): \( x_{\alpha,\beta} = 0 \) if \( \beta + n\alpha \in A \), \( x_{\alpha,\beta} = 1 \) if \( \beta + n\alpha \in A^c \); and the sequence \( x'_{\alpha,\beta} \) codes the orbit of \( \beta \) with respect to the partition \( B = (0,1-\alpha), B^c = (1-\alpha,1) \).

A mechanical sequence is rational if the line \( y = \alpha x + \beta \) has rational slope and irrational if \( y = \alpha x + \beta \) has irrational slope. In [137] it is proved that a sequence \( u \) is Sturmian if and only if \( u \) is irrational mechanical. In the following example we construct a lower mechanical sequence with irrational slope, thus producing a Sturmian sequence.

Example 3.2.4. Let \( \alpha = 1/\tau^2 \), where \( \tau = (1 + \sqrt{5})/2 \) is the golden mean, and \( \beta = 0 \). The lower mechanical sequence \( x_{\alpha,\beta} \) is constructed as follows:

\[
\begin{align*}
(x_{\alpha,\beta})_0 &= [1/\tau^2] = 0 \\
(x_{\alpha,\beta})_1 &= [2/\tau^2] - [1/\tau^2] = 0 \\
(x_{\alpha,\beta})_2 &= [3/\tau^2] - [2/\tau^2] = 1 \\
(x_{\alpha,\beta})_3 &= [4/\tau^2] - [3/\tau^2] = 0 \\
(x_{\alpha,\beta})_4 &= [5/\tau^2] - [4/\tau^2] = 0 \\
(x_{\alpha,\beta})_5 &= [6/\tau^2] - [5/\tau^2] = 1 \\
&\vdots
\end{align*}
\]

Further calculation shows that \( x_{\alpha,\beta} = 0010010100... = 0f \). Note that a similar calculation gives \( x'_{\alpha,\beta} = 1010010100... = 1f \), hence the Fibonacci sequence is a shift of the lower and upper mechanical sequences with slope \( 1/\tau^2 \) and intercept 0.

Exercise 3.2.2. Show that while Sturmian sequences are aperiodic, they are syndetically recurrent: every block that occurs in a Sturmian sequence occurs an infinite number of times with bounded gaps.

As a result of the preceding Exercise, any block in \( L_n(u) \) appears past the initial position and can thus be extended on the left. Since there are \( n + 1 \) blocks of length \( n \), it must be that exactly one of them can be extended to the left in two ways.

Definition 3.2.5. In a Sturmian sequence \( u \), the unique block of length \( n \) that can be extended to the left in two different ways is called a left special block, and is denoted \( L_n(u) \). The sequence \( l(u) \) which has the \( L_n(u) \)'s as prefixes is called the left special sequence or characteristic word of \( X_u^+ \) [78,137].

In a similar fashion, we define the right special blocks of \( L_n(u) \).

Definition 3.2.6. In a Sturmian sequence \( u \), the unique block of length \( n \) that can be extended to the right in two different ways is called a right special block, and is denoted \( R_n(u) \). The block \( R_n(u) \) is precisely the reverse of \( L_n(u) \) [78].

We now address how to determine the left special sequence in a Sturmian system.

Since every Sturmian sequence \( u \) is irrational mechanical, there is a line with irrational slope \( \alpha \) associated to \( u \). We use this \( \alpha \) to determine the left special sequence of \( X_u^+ \).

Let \( (d_1, d_2, \ldots, d_n, \ldots) \) be a sequence of integers with \( d_1 \geq 0 \) and \( d_n > 0 \) for \( n > 1 \). We associate a sequence \( (s_n)_{n\geq-1} \) of blocks to this sequence by

\[
s_{-1} = 1, \quad s_0 = 0, \quad s_n = s_{n-1}d_n s_{n-2}.
\]

The sequence \( (s_n)_{n\geq-1} \) is a standard sequence, and \( (d_1, d_2, \ldots, d_n, \ldots) \) is its directive sequence. We can then determine the left special sequence of \( X_u^+ \) with the following proposition stated in [137].
Proposition 3.2.7. Let $\alpha = [0, 1 + d_1, d_2, ...]$ be the continued fraction expansion of an irrational $\alpha$ with $0 < \alpha < 1$, and let $(s_n)$ be the standard sequence associated to $(d_1, d_2, ...)$. Then every $s_n$, $n \geq 1$, is a prefix of $l$ and

$$l = \lim_{n \to \infty} s_n.$$ 

This is illustrated in the following two examples.

Example 3.2.8. Let $\alpha = 1/\tau^2$, where $\tau = (1 + \sqrt{5})/2$ is the golden mean. The continued fraction expansion of $1/\tau^2$ is $[0, 2, 1]$. By the above proposition $d_1 = 1, d_2 = 1, d_3 = 1, d_4 = 1, ...$. The standard sequence associated to $(d_1, d_2, ...)$ is constructed as follows:

- $s_1 = s_0^{d_1} s_{-1} = 01$
- $s_2 = s_1^{d_2} s_0 = 010$
- $s_3 = s_2^{d_3} s_1 = 01001$
- $s_4 = s_3^{d_4} s_2 = 01001010$

Continuing this process, the left special sequence of $X_u^+$, where $u$ is a coding of a line with slope $1/\tau^2$, is $l = 01001010010010100101001... = f$. It follows that the left special sequence of $X_f^+$ is $f$.

3.3 Episturman sequences

Sturmian sequences have many equivalent definitions and characterizations as well as many amazing properties. Some of the properties allow generalization to sequences on alphabets of more than two symbols. Episturman sequences were defined in [68] as follows. The right palindrome closure of a block $B$ is the unique shortest palindrome which has $B$ as a prefix.

Definition 3.3.1. A one-sided sequence $u$ is standard episturman if the right palindrome closure of every prefix of $u$ is also a prefix of $u$. A sequence $v$ is episturman if there is a standard episturman sequence $u$ such that $\mathcal{L}(v) = \mathcal{L}(u)$.

Exercise 3.3.1. Prove that every Sturmian sequence is episturman.

Exercise 3.3.2. Prove that every episturman sequence is syndetically recurrent.

We say that an infinite sequence $u$ is closed under reversals if for each subblock $B = b_1 ... b_n$ of $u$, its reversal $B' = b_n ... b_1$ is also a subblock of $u$. Recall that a right-special subblock of an infinite sequence $u$ is a subblock $B$ of $u$ that has two distinct continuations: there are symbols $b, c \in A$ with $b \neq c$ and both $Bb$ and $Bc$ appearing in $u$ as subblocks.

Theorem 3.3.2. [68] A one-sided sequence $u$ is episturman if and only if it is closed under reversals and for each $n \geq 1$ it has at most one right-special subblock of length $n$.

For a review of properties of episturman sequences, including their complexity, see [85].
3.4 The Morse sequence

The Morse sequence, more properly called the Prouhet-Thue-Morse sequence, is the fixed point

\[ \omega = 0110100110010110 \ldots \]

of the substitution \( 0 \to 01, 1 \to 10 \). The complexity function of the Morse sequence is more complicated than that of the Fibonacci sequence. For the Morse sequence, \( p_\omega(1) = 2, p_\omega(2) = 4 \), and, for \( n \geq 3 \), if \( n = 2^r + q + 1, r \geq 0, 0 < q \leq 2^r \), then

\[
p_\omega(n) = \begin{cases} 
6(2^{r-1}) + 4q & \text{if } 0 < q \leq 2^{r-1} \\
8(2^{r-1}) + 2q & \text{if } 2^{r-1} < q \leq 2^r.
\end{cases}
\]

The complexity function of the Morse sequence is discussed in more detail in [78, Chapter 5].

3.5 In higher dimensions, tilings, groups, etc.

The complexity of configurations and tilings in higher-dimensional spaces and even groups is an area of active investigation. A central question has been the possibility of generalizing the observation of Hedlund and Morse (Exercise 3.1.2) to higher dimensions: any configuration of low enough complexity, in some sense, should be eventually periodic, in some sense. A definite conjecture in this direction was stated in 1997 by M. Nivat in a lecture in Bologna (see [72]): For a \( d \)-dimensional configuration \( x : \mathbb{Z}^d \to A \) on a finite alphabet \( A \), define its rectangular complexity function to be the function \( P_x(m_1, \ldots, m_d) \) which counts the number of different \( m_1 \times \cdots \times m_d \) box configurations seen in \( x \). The Nivat Conjecture posits that if \( x \) is a two-dimensional configuration on a finite alphabet for which there exist \( m_1, m_2 \geq 1 \) such that \( P_x(m_1, m_2) \leq m_1 m_2 \), then \( x \) is periodic: there is a nonzero vector \( w \in \mathbb{Z}^2 \) such that \( x(v + w) = x(v) \) for all \( v \in \mathbb{Z}^2 \).

Cassaigne [48] characterized all two-dimensional configurations with complexity function \( P_x(m_1, m_2) = m_1 m_2 + 1 \).

Vuillon [180] considered tilings of the plane generated by a cut-and-project scheme. Recall (see [9]) that Sturmian sequences code (according to the two possible tile=interval lengths) tilings of a line obtained by projecting onto it the points in the integer lattice that are closest to it along another, transverse, line. Vuillon formed tilings of the plane with three types of diamonds by projecting onto a plane points of the cubical lattice in \( \mathbb{Z}^3 \) and proved that the number of different \( m \times n \) parallelograms is \( mn + m + n \).

Berthé and Vuillon [21] showed that these two-dimensional configurations code the \( \mathbb{Z}^2 \) action of two translations on the circle. By applying a one-block code from the three-letter alphabet to a two-letter alphabet, they produced for each \( m \) and \( n \) a two-dimensional configuration which is syndetically recurrent and is not periodic in any rational direction but has the relatively low rectangular complexity function \( P(m, n) = mn + n \). Two-dimensional configurations with this complexity function were characterized in [20].

Sander and Tijdeman [169–171] studied the complexities of configurations \( x : \mathbb{Z}^d \to \{0, 1\} \) in terms of the number of distinct finite configurations seen under a sampling window. Let \( A = \{a_1, \ldots, a_n\} \), each \( a_i \in \mathbb{Z}^d \), be a fixed non-empty sampling window, and define

\[
P_x(A) = \text{card}\{ (x(v + a_1), \ldots, x(v + a_n)) : v \in \mathbb{Z}^d \}
\]

(3.5.1)

to be the number of distinct \( A \)-patterns in \( x \) (written here as ordered \( |A| \)-tuples). A natural extension of (3.1.2) might be that if there is a nonempty set \( A \subset \mathbb{Z}^d \) for which \( P_x(A) \leq |A| \), then \( x \) must be periodic: there is a \( w \in \mathbb{Z} \setminus \{0\} \) such that \( x(v + w) = x(v) \) for all \( v \in \mathbb{Z} \). Sander and Tijdeman proved the following.
1. If \( P_x(A) \leq |A| \) for some \( A \subset \mathbb{Z}^d \) with \( |A| \leq 3 \), then \( x \) is periodic.

2. In dimension 1, the observation of Hedlund and Morse generalizes from sampling windows that are intervals to arbitrary sampling windows: If \( x \in \{0,1\}^\mathbb{Z} \) satisfies \( P_x(A) \leq |A| \) for some (non-empty) sampling window \( A \), then \( x \) is periodic.

3. There are a non-periodic two-dimensional configuration \( x : \mathbb{Z}^2 \to \{0,1\} \) and a sampling window \( A \subset \mathbb{Z}^2 \) of size \( |A| = 4 \) such that \( P_x(A) = 4 = |A| \).

4. Conjecture: If \( A \subset \mathbb{Z}^2 \) is a (non-empty) sampling window that is the restriction to \( \mathbb{Z}^2 \) of a convex subset of \( \mathbb{R}^2 \) and \( x : \mathbb{Z}^2 \to \{0,1\} \) satisfies \( P_x(A) \leq |A| \), then \( x \) is periodic.

5. If there is a sampling window \( A \subset \mathbb{Z}^2 \) that consists of all points in a rectangle (with both sides parallel to the coordinate axes) with one side of length 2, and \( P_x(A) \leq |A| \), then \( x \) is periodic.

6. In dimension 3 and above, there are rectangular box sampling windows \( A \) with sides parallel to the coordinate axes and nonperiodic configurations \( x \) with \( P_x(A) \leq |A| \).

The last statement above was recently improved by Cyr and Kra [60]: If there is a sampling window \( A \) that consists of all points in a rectangle (with both sides parallel to the coordinate axes) with one side of length 3, and \( P_x(A) \leq |A| \), then \( x \) is periodic.

Kari and Szabados [114] (see also [113]) represented configurations in \( \mathbb{Z}^d \) as formal power series in \( d \) variables with coefficients from \( A \) and used results from algebraic geometry to study configurations in \( \mathbb{Z}^d \) which have low complexity in the sense that for some sampling windows \( A \) they satisfy \( P_x(A) \leq |A| \). They proved that in dimension two any non-periodic configuration \( x \) can satisfy such an estimate for only finitely many rectangular sampling windows \( A \).

Epifanio, Koskas, and Mignosi [72] made some progress on the Nivat Conjecture by showing that if \( x \) is a configuration on \( \mathbb{Z}^2 \) for which there exist \( m, n \geq 1 \) such that \( P_x(m,n) < mn/144 \), then \( x \) is periodic. The statement was improved by Quas and Zamboni [160] by combinatorial and geometrical arguments to replace 1/144 by 1/16, and by Cyr and Kra [58] by arguments involving subdynamics to replace it by 1/2.

We do not give definitions of all the terminology associated with tilings and tiling dynamical systems—see for example [79,162,174] for background. For a tiling \( x \) of \( \mathbb{R}^d \) that has finite local complexity, one may define its complexity function \( P_x(r) \) to be the number of different patches (identical up to translation, or perhaps translation and rotation) seen in \( x \) within spheres of radius \( r \). In analogy with Exercise 3.0.1 for subshifts, Frank and Sadun [156] and A. Julien [108] (see also [107]) showed that if two minimal tiling dynamical systems are aperiodic and have finite local complexity, then their complexity functions are equivalent—within bounded multiples of each other up to bounded translations (or dilations)—see the cited papers for precise statements.

The investigation of the complexity function and the calculation or even estimation of entropy are extending to subshifts on groups (see for example [10,154]) and even on trees [11,12,14].

Analogues of the Nivat Conjecture for general Delaunay sets in \( \mathbb{R}^d \) were proved by Lagarias and Pleasants [130,131]. Huck and Richard [104] estimate the pattern entropy of “model sets” (certain point sets that result from cut and project schemes) in terms of the size of the defining window.

Durand and Rigo [71] proved a reformulation of Nivat’s Conjecture by redefining periodicity and using a different complexity function: A subset \( E \subset \mathbb{Z}^d \) is “definable in Presburg arithmetic \((\mathbb{Z}; <, +)\)” if and only if the number \( R_E(n) \) of different blocks that appear infinitely many times in \( E \) is \( O(n^{d-1}) \) and “every section is definable in \((\mathbb{Z}; <, +)\)”. We do not attempt to explain the terminology here, but just note that the subsets of \( \mathbb{N} \) definable in \((\mathbb{N}; <, +)\) correspond exactly to the eventually periodic sequences, so the theorem of Durand and Rigo may be regarded as an extension of the Hedlund-Morse theorem to all dimensions.
3.6 Topological complexity

Let $(X,T)$ be a topological dynamical system. If $(X,T)$ is a subshift and $U$ is the time-0 cover (also partition) consisting of the cylinder sets determined by fixing a symbol at the origin, then the complexity function $p_X(n)$ (which by definition is the number of distinct $n$-blocks in all sequences in the system) is the minimal possible cardinality of any subcover of $U_0^{n-1} = U \cup T^{-1}U \cup \cdots \cup T^{-n+1}U$; i.e., in this case $p_X(n)$ equals the $N(U_0^{n-1})$ of the definition of topological entropy (see Sections 3.1 and 2.1). Blanchard, Host, and Maass [25] took this as the definition of the topological complexity function: $p_U(n) = N(U_0^{n-1})$ is the minimum possible cardinality of a subcover of $U_0^{n-1}$.

**Theorem 3.6.1.** [25] A topological dynamical system is equicontinuous if and only if every finite open cover has bounded complexity function. (Cf. Sections 3.1 and 3.13.)

**Exercise 3.6.1.** Discuss this theorem in relation to a Sturmian subshift and the irrational translation on $[0,1]$ of which it is an almost one-to-one extension.

They also related the complexity function to concepts of mixing and chaos.

**Definition 3.6.2.** A topological dynamical system is scattering if every covering by non-dense open sets has unbounded complexity function. It is 2-scattering if every covering by two non-dense open sets has unbounded complexity function.

The following results are from [25].

1. Every topologically weakly mixing system is scattering.
2. For minimal systems, 2-scattering, scattering, and topological weak mixing are equivalent.
3. If every non-trivial closed cover $U$ of $X$ has complexity function satisfying $p_U(n) \geq n + 2$ for all $n$, then $(X,T)$ is topologically weakly mixing.
4. If $(X,T)$ has a point of equicontinuity, then there exists an open cover $U$ of $X$ with $p_U(n) \leq n + 1$ for all $n$.
5. A system is scattering if and only if its Cartesian product with every minimal system is transitive.
6. Every scattering system is disjoint from every minimal distal system. (Recall that $(X,T)$ and $(Y,S)$ are disjoint if the only closed invariant subset of their Cartesian product that projects onto both $X$ and $Y$ is all of $X \times Y$.)

The topological complexity of nilsystems has been studied for example in [64,98,157].

3.7 Low complexity, the number of ergodic measures, automorphisms

Irrational translations on the unit interval and their generalizations to interval exchanges have natural codings as subshifts on finite alphabets which have low complexity and a small number of ergodic measures. A subshift $(X,\sigma)$ is said to have minimal block growth if there is a constant $c < \infty$ such that its complexity function $p_X(n) = |L_n(X)|$ satisfies $p_X(n) - n < c$ for all $n$. Such subshifts were studied and characterized by Coven and Hedlund [53], Coven [52], and Paul [150].

Cassaigne [46,47] proved that a sequence $u$ satisfies $p_u(n) \leq Kn$ for some constant $K$ and all $n$ if and only if the sequence $p_u(n+1) - p_u(n)$ is bounded. Rote [164] constructed sequences with $p_u(n) = 2n$ for all $n$ and proved that a sequence $u$ on $\{0,1\}$ has complexity function $p_u(n) = 2n$ and language closed under switching 0 and 1 if and only if its difference sequence $u_{n+1} - u_n \mod 2$ is Sturmian.

The coding of an exchange of $r$ intervals satisfies $p_X(n) = (r-1)n + 1$ for all $n$ (generalizing the Sturmian case, when $r = 2$). Veech [177] and Katok [116] showed independently that an exchange of $r$
Part 3. Counting patterns

intervals has at most \( \lfloor r/2 \rfloor \) ergodic measures. Boshernitzan [32] showed that a general minimal subshift with linear block growth has a bounded number of ergodic measures. Denoting by \( \mathcal{E}(X, \sigma) \) the set of ergodic invariant Borel probability measures on \((X, \sigma)\), his main results are as follows:

1. If \( \liminf \frac{p_X(n)}{n} = \alpha \), then \( |\mathcal{E}(X, \sigma)| \leq \alpha \).

2. If there is an integer \( K \geq 3 \) such that \( \limsup \frac{p_X(n)}{n} < K \), then \( |\mathcal{E}(X, \sigma)| \leq K - 2 \).

3. If there is a real \( \alpha \geq 2 \) such that \( \limsup \frac{p_X(n)}{n} = \alpha \), then \( |\mathcal{E}(X, \sigma)| \leq \lfloor \alpha \rfloor - 1 \).

**Exercise 3.7.1.** Show that (2) implies (3).

**Exercise 3.7.2.** Show that for a (primitive, proper) substitution \( \theta : A \to A^* \) (\( A^* \) is the set of all blocks on the alphabet \( A \)) that defines a minimal subshift \((X_\theta, \sigma)\) as the orbit closure of a sequence fixed under \( \theta \), the complexity function \( p_X \) satisfies \( p_X(n) \leq Kn \) for all \( n \), where \( K = \max\{|K(a)| : a \in A\} \).

Cyr and Kra [55] showed that these results of Boshernitzan are sharp by constructing for each integer \( d \geq 3 \) a minimal subshift \((X, \sigma)\) for which

\[
\liminf \frac{p_X(n)}{n} = d, \quad \limsup \frac{p_X(n)}{n} = d + 1, \quad \text{and} \quad |\mathcal{E}(X, \sigma)| = d.
\]

Damron and Fickenscher [62] considered minimal subshifts with eventually constant block growth: there are constant \( K \) and \( n_0 \) such that

\[
p_X(n + 1) - p_X(n) = K \quad \text{for all} \quad n \geq n_0, \quad \text{equivalently} \quad p_X(n) = Kn + C \quad \text{for all} \quad n \geq n_0.
\]

This condition is satisfied by codings of interval exchanges but is stronger than linear block growth. They improved Boshernitzan’s bound (of \( K - 1 \)) for this smaller class of subshifts by showing that if a minimal subshift \((X, \sigma)\) has eventually constant block growth with constant \( K \geq 4 \), then \( |\mathcal{E}(X, \sigma)| \leq K - 2 \). The proof involved careful study of the frequencies of blocks and the relation between Rauzy graphs for \( n \)-blocks and for \((n + 1)\)-blocks.

There is recent activity showing that if the complexity function \( p_X(n) \) of a minimal subshift \((X, \sigma)\) grows slowly, then there are not very many automorphisms (shift-commuting homeomorphisms) of the system. This is in contrast to nontrivial mixing shifts of finite type, for which the automorphism group is huge [38,94]. The papers of Coven [51], Olli [147], Salo-Törmä [168], Coven-Quas-Yassawi [54], Donoso-Durand-Maass-Petite [65], and Cyr-Kra [57] contain such results. In the latter two papers it is proved that if

\[
\liminf_{n \to \infty} \frac{p_X(n)}{n} < \infty,
\]

then the automorphism group of \((X, \sigma)\) is virtually \( \mathbb{Z} \), meaning that its quotient by the subgroup generated by \( \sigma \) is a finite group.

The recent paper of Cyr and Kra [59] shows that if \( p_X \) has stretched exponential growth in the sense that there is \( \beta < 1/2 \) such that

\[
\limsup_{n \to \infty} \frac{\log p_X(n)}{n^\beta} = 0,
\]

then the automorphism group of \((X, \sigma)\) is amenable. (The linear and quadratic complexity cases are discussed in [56,57]).

The recent paper of Salo [167] gives an example of a minimal subshift \((X, \sigma)\) for which \( p_X(n) = O(n^{1.757}) \) (subquadratic complexity), and the automorphism group is not finitely generated.
3.8 Palindrome complexity

A palindrome is a block $B = b_1 \ldots b_n$ which reads the same forwards as backwards: denoting by $B' = b_n \ldots b_1$ the reversal of $B$, we have $B' = B$. For a (one or two-sided) sequence $u$ and each $n \geq 0$ we denote by $\text{Pal}_u(n)$ the number of $n$-palindromes (palindromes of length $n$) in $u$. (We may define $\text{Pal}_u(0) = \text{Pal}_u(1) = 1$.) For a language $\mathcal{L}$, $\text{Pal}_\mathcal{L}(n)$ denotes the number of $n$-palindromes in $\mathcal{L}$, and for a subshift $(X, \sigma)$, $\text{Pal}_X(n)$ denotes the number of $n$-palindromes in $\mathcal{L}(X)$, i.e. the number of palindromes of length $n$ found among all sequences in $X$. For any $\alpha$, denote by $\text{Pal}_\alpha$ the sum of $\text{Pal}_\alpha(n)$ over all $n$.

Besides the inherent combinatorial interest of finding palindromes and estimating how many occur, such knowledge may have applications elsewhere, for example in mathematical physics, as explained in [5]. Let $A$ be a finite alphabet and $f : A \to \mathbb{R}$ a one-to-one “potential function”. Let $u$ be a sequence on $A$. Then the one-dimensional discrete Schrödinger operator $H_u$ acting on $l^2(\mathbb{Z})$ is defined by

$$
(H_u\phi)(n) = \phi(n-1) + \phi(n+1) + f(u_n)\phi(n) \text{ for all } n \in \mathbb{Z}.
$$

It is thought that the nature of the spectrum of $H_u$ can indicate behavior of a material (perhaps a quasicrystal) described by the sequence $u$. Absolutely continuous spectrum (with respect to Lebesgue measure) might indicate behavior similar to a conductor, pure point spectrum might indicate insulating behavior, and singular continuous spectrum might be more interesting than either of the others.

**Theorem 3.8.1.** [97] Let $u$ be a sequence on a finite alphabet whose orbit closure $(\overline{O}(u), \sigma)$ is strictly ergodic and infinite and which contains arbitrarily long palindromes, i.e. $\text{Pal}_u(n)$ is not eventually $0$. Then for uncountably many $x \in \overline{O}(u)$, the operator $H_x$ has purely singular continuous spectrum.

**Exercise 3.8.1.** Show that the fixed points of the Fibonacci substitution $0 \to 01, 1 \to 0$ and Morse substitution $0 \to 01, 1 \to 10$ contain arbitrarily long palindromes.

It is interesting that Sturmian sequences can be characterized not just by their complexity function (it is as small as possible for a sequence that is not eventually periodic: $p_u(n) = n + 1$ for all $n$), but also by their palindrome complexity functions. Droubay [67] showed that the fixed point of the Fibonacci substitution satisfies the conclusion of the following theorem of Droubay and Pirillo [69]. A generalization to two dimensions was given in [22].

**Theorem 3.8.2.** [69] A one-sided sequence $u$ on a finite alphabet is Sturmian if and only if $\text{Pal}_u(n) = 1$ for all even $n \geq 2$ and $\text{Pal}_u(n) = 2$ for all odd $n$.

Damanik and Zare [61] studied the palindromic complexity of fixed points of primitive substitutions. Recall that the orbit closure of any such sequence is strictly ergodic, so that every block $B$ has a uniform frequency of occurrence $\text{Freq}_u(B)$ in $u$, which equals the measure of the cylinder set consisting of all sequences that have $B$ occurring at the origin.

**Theorem 3.8.3.** [61] Let $u$ be a one-sided sequence that is a fixed point of a primitive substitution on a finite alphabet. Then

1. $\text{Pal}_u(n)$ is bounded;
2. There are constants $c_1$ and $c_2$ such that if $B$ is a palindrome of length $n$ in $u$, then

$$
\frac{c_1}{n} \leq \text{Freq}_u(B) \leq \frac{c_2}{n} \text{ for all } n.
$$

The palindrome complexity function is studied for many examples in [5], where it is also proved that if a sequence $u$ has linear complexity function (there is a constant $C$ such that $p_u(n) \leq Cn$ for all $n$), then $\text{Pal}_u$ is bounded.
Part 3. Counting patterns

**Theorem 3.8.4.** [5] Let $u$ be a one-sided sequence on a finite alphabet that is not eventually periodic. Then for all $n \geq 1$,

\[(3.8.3) \quad \text{Pal}_u(n) < \frac{1}{n}16p_u \left(n + \left\lfloor \frac{n}{4} \right\rfloor \right).\]

These results were generalized in [13] as follows. Recall that for a sequence $u$, $\mathcal{L}(u)$ denotes the family of all subblocks of $u$. A formal language $\mathcal{L}$ is called **closed under reversal** if $\mathcal{L}' \subset \mathcal{L}$.

**Theorem 3.8.5.** [13] Let $u$ be a syndetically recurrent sequence on a finite alphabet.

1. If $\mathcal{L}(u)$ is not closed under reversal, then $u$ does not contain arbitrarily long palindromes: $\text{Pal}_u(n) = 0$ for all sufficiently large $n$.

2. If $\mathcal{L}(u)$ is closed under reversal, then

\[(3.8.4) \quad \text{Pal}_u(n) + \text{Pal}_u(n + 1) \leq p_u(n + 1) - p_u(n) + 2 \text{ for all } n.\]

3. Suppose that $u$ is the natural coding of a “nondegenerate” exchange of $r$ intervals, so that $p_u(n) = n(r-1)+1$ for all $n$. If $\mathcal{L}(u)$ is closed under reversal, then equality holds in the preceding estimate, and in fact

\[(3.8.5) \quad \text{Pal}_u(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ r & \text{if } n \text{ is odd}. \end{cases}\]

**Exercise 3.8.2.** Check that Equation 3.8.5 implies that equality holds in Equation 3.8.4

The paper [86] studies blocks and sequences that contain a lot of palindromes. The following observation was made in [68].

**Exercise 3.8.3.** Let $B$ be a block of length $|B| = n$. Show that the number of different palindromes that are subblocks of $B$ (including the empty block) is at most $n + 1$. (Hint: Define a block $B$ to have **Property J** if there is a suffix of $B$ that is a palindrome and appears only once in $B$. For $x \in A$, what are the possibilities for $\text{Pal}_{Bx}$, depending on whether $B$ has property $J$ or not? Then use induction on the lengths of prefixes of $B$ to show that $\text{Pal}_B$ is the cardinality of the set of prefixes of $B$ that have property $J$.)

In view of the preceding observation, the authors of [86] define a finite block $B$ to be **rich** if it contains the maximum possible number, $|B| + 1$, of palindromes as subblocks. An infinite sequence is defined to be rich if every one of its subblocks is rich. (See also [7] and [40].)

**Exercise 3.8.4.** Prove that a block $B$ is rich if and only if each of its prefixes has a suffix that is a palindrome and appears exactly once in $B$.

**Exercise 3.8.5.** If $B$ is rich then each of its subblocks is also rich.

**Theorem 3.8.6.** [86] A finite block or infinite sequence $w$ is rich if and only if for every subblock $B$ of $w$, if $B$ contains exactly two occurrences of a palindrome as a prefix and a suffix, then $B$ is itself a palindrome.

In [68] it is proved that every episturmian sequence is rich.

**Exercise 3.8.6.** Let $B$ be a block of length $n = |B|$ on an alphabet $A$ of cardinality $|A| = r$. What is the maximum possible cardinality of the family $\mathcal{L}_B$ of all (different) subblocks of $B$?

**Exercise 3.8.7.** What is the minimum length $L(k, r)$ of a block on an alphabet $A$ of $r$ symbols that contains all $k$-blocks on $A$? What about configurations on $r$ symbols in $\mathbb{Z}^2$?
3.9 Nonrepetitive complexity and Eulerian entropy

Exercise 3.8.8. How many rich words of length 5 are there on an alphabet of 3 letters? How many are there of length \( n \) on an alphabet of \( r \) letters?

Exercise 3.8.9. Formulate a definition of palindrome in \( \mathbb{Z}^2 \). Does the support of a “palindrome” in \( \mathbb{Z}^2 \) have to be a rectangle? How about a definition in \( \mathbb{Z}^d \)?

3.9 Nonrepetitive complexity and Eulerian entropy

T. K. S. Moothathu [142] defined the nonrepetitive complexity function \( P_u^N \) of a sequence \( u \) on a finite alphabet \( A \) in terms of how long an initial block of \( u \) could be before it contained a repeat of a subblock of length \( n \):

\[
P_u^N(n) = \max \{ m \in \mathbb{N} : u_i^{i+n-1} \neq u_j^{j+n-1} \text{ for all } 0 \leq i < j \leq m-1 \}.
\]

The exponential growth rate may be called the Eulerian entropy of \( u \), because there is a connection with Eulerian circuits in directed graphs, and denoted by \( h_E(u) \):

\[
h_E(u) = \limsup_{n \to \infty} \frac{\log P_u^N(n)}{n}.
\]

Referring to Bowen’s definition of topological entropy in terms of separated sets (2.1), the concept of Eulerian entropy extends to points in arbitrary topological dynamical systems \((X, T)\) when \( X \) is a metric space. For a point \( x \in X \), denote by \( \beta(x, n, \epsilon) \) the maximum \( m \in \mathbb{N} \) for which the initial orbit segment of \( x \) of length \( m \) is \((n, \epsilon)\)-separated, so that

\[
\beta(x, n, \epsilon) = \max \{ m \in \mathbb{N} : \{x, Tx, \ldots, T^{m-1}x\} \text{ is } (n, \epsilon)\text{-separated} \},
\]

define the Eulerian entropy at \( x \) to be

\[
h_E(X, T, x) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log \beta(x, n, \epsilon)}{n},
\]

and define the Eulerian entropy of the system to be

\[
h_E(X, T) = \sup_{x \in X} h_E(X, T, x).
\]

The exact relationships among \( h_E(X, T, x) \), \( h_E(X, T) \), and \( h_{\text{top}}(X, T) \) are not clear, but here are some results from [142].

1. \( h_E(X, T) \) is an invariant of topological conjugacy.
2. Let \((X, \sigma)\) be a one-step mixing shift of finite type on a finite alphabet \( A \) which has a safe symbol, a symbol \( s \in A \) such that for all \( a \in A \) both blocks \( as, sa \in \mathcal{L}(X) \). Then there is a sequence \( x \in X \) such that \( h_E(X, \sigma, x) = h_{\text{top}}(X, \sigma) \).
3. If \((X, \sigma)\) is a mixing shift of finite type such that for all large enough \( n \) the de Bruijn graph of \((X, \sigma)\) is irreducible, then \( h_E(X, \sigma, x) = h_{\text{top}}(X, \sigma) \) for a residual set of \( x \in X \). (Recall that the de Bruijn graph of \((X, \sigma)\) has for its vertices the set of \((n-1)\)-blocks, and there is an edge from \( B \) to \( C \) if and only if there is an \( n \)-block that has prefix \( B \) and suffix \( C \).)
4. There is a (nontransitive) homeomorphism \( T \) on the Cantor set \( X \) such that \( h_E(X, T, x) < h_{\text{top}}(X, T) \) for all \( x \in X \).

Exercise 3.9.1. Construct some de Bruijn graphs for several mixing shifts of finite type. How can one tell whether they will be irreducible for all large enough \( n \)?

Exercise 3.9.2. Determine \( P_u^N(n) \) and \( h_E(u) \) to the extent possible when \( u \) is the fixed point of the Fibonacci substitution, the Prouhet-Thue-Morse sequence, the Champernowne sequence, etc.
3.10 Mean topological dimension

The complexity function \( p_X \) and most of its variants are especially useful in the case of systems that have topological entropy zero. In a study of holomorphic and harmonic maps, M. Gromov [90] proposed an invariant, which he called \textit{mean topological dimension}, that is especially useful in the study of systems with infinite entropy, such as the shift on \([0, 1] \mathbb{Z}\). This invariant was studied in [136] (based on Lindenstrauss’ Ph.D. dissertation) and used there and in subsequent papers [91, 134, 135] to settle questions about the possibility of embedding a given topological dynamical system \((X, T)\) in a standard target such as \((([0, 1]^d) \mathbb{Z}, \sigma)\).

We deal with a compact metric space \(X\) and finite open covers \(\mathcal{U}, \mathcal{V}\) of \(X\). Recall that \(\mathcal{V}\) refines \(\mathcal{U}\), written \(\mathcal{V} \succ \mathcal{U}\), if every member of \(\mathcal{V}\) is contained in some member of \(\mathcal{U}\), and the join of \(\mathcal{U}\) and \(\mathcal{V}\) is \(\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}\).

The \textit{dimension} of a finite open cover \(\mathcal{U}\) is defined to be

\[
D(\mathcal{U}) = \min_{\mathcal{V} \succ \mathcal{U}} \max_{x \in X} \sum_{V \in \mathcal{V}} 1_V(x) - 1,
\]

the smallest \(n\) such that for every refinement \(\mathcal{V}\) of \(\mathcal{U}\) no point of \(X\) is in more than \(n + 1\) elements of \(\mathcal{V}\). The \textit{cover dimension} of the compact metric space \(X\) is

\[
D(X) = \sup \{ D(\mathcal{U}) : \mathcal{U} \text{ is a finite open cover of } X \}.
\]

The \textit{mean topological dimension} of \((X, T)\) is defined to be

\[
\text{mdim}(X, T) = \sup_{\mathcal{U}} D(\mathcal{U}, T)
\]

and is an invariant under topological conjugacy.

Here are a few results from [134, 136].

1. If \(D(X) < \infty\) or \(h_{\text{top}}(X, T) < \infty\), then \(\text{mdim}(X, T) = 0\).

2. \(\text{mdim}([0, 1]^2, \sigma) = D([0, 1]) = 1\).

3. If \(Y \subset X\) is closed and \(T\)-invariant \((TY = Y)\), then \(\text{mdim}(Y, T) \leq \text{mdim}(X, T)\).

4. It was an open question of J. Auslander whether every minimal topological dynamical system \((X, T)\) can be embedded in \(([0, 1]^2, \sigma)\) (in analogy with Bebutov’s theorem on the embedding of \(\mathbb{R}\) actions in \(C(\mathbb{R})\) with the translation action—see [109]). Because of the preceding result the answer is negative, since for any \(r \in [0, \infty)\) there is a minimal system \((X, T)\) with \(\text{mdim}(X, T) = r\).

5. If \(d \in \mathbb{N}\) and \((X, T)\) is a topological dynamical system with \(\text{mdim}(X, T) < d/36\), then \((X, T)\) embeds in \((([0, 1]^d)^2, \sigma)\).

Exercise 3.10.1. Prove that \(D([0, 1]) = 1\).

3.11 Amorphic complexity via asymptotic separation numbers

To measure the complexity of zero-entropy systems, in [80] the authors propose the concepts of \textit{separation numbers} and \textit{amorphic complexity}. Let \((X, T)\) be a topological dynamical system, with \(X\) a compact metric space. For \(\delta > 0\), \(n \in \mathbb{N}\), and points \(x, y \in X\), let

\[
S_n(x, y, \delta) = \text{card}\{k = 0, 1, \ldots, n - 1 : d(T^k x, T^k y) > \delta\}
\]
be the number of times that the first $n$ iterates of $x$ and $y$ are separated by a distance greater than $\delta$; and, for $\nu \in (0,1]$, define $x$ and $y$ to be $(\delta, \nu)$-separated if

$$
\limsup_{n \to \infty} \frac{S_n(x,y,\delta)}{n} \geq \nu.
$$

A subset $E \subset X$ is called $(\delta, \nu)$-separated if every pair $x, y \in E$ with $x \neq y$ is $(\delta, \nu)$-separated. The maximum cardinality of a $(\delta, \nu)$-separated subset of $X$ is called the \textit{(asymptotic) separation number for distance $\delta$ and frequency $\nu$} and is denoted by Sep$(\delta, \nu)$. (See also Section 2.5.)

**Exercise 3.11.1.** Prove that if $(X,T)$ is equicontinuous, then Sep$(\delta, \nu)$ is uniformly bounded in $\delta$ and $\nu$.

For a fixed $\delta > 0$, the upper and lower scaling rates of the separation numbers as the frequency tends to 0 are defined to be

$$
A^+(\delta) = \limsup_{\nu \to 0} \frac{\log \text{Sep}(\delta, \nu)}{\log \nu} \quad \text{and} \quad A^-(\delta) = \liminf_{\nu \to 0} \frac{\log \text{Sep}(\delta, \nu)}{\log \nu}.
$$

The polynomial scaling rate has been chosen as possibly the most interesting one to apply to zero-entropy examples. And notice also that the rate is in terms of shrinking frequency instead of increasingly long time interval as usual, since the time asymptotics were already included before.

The \textit{upper and lower amorhic complexities} are defined to be

$$
A^+(X,T) = \sup_{\delta > 0} A^+(\delta) \quad \text{and} \quad A^-(X,T) = \sup_{\delta > 0} A^-(\delta).
$$

If $A^+(X,T) = A^-(X,T)$, then their common value is called the \textit{amorphic complexity} of $(X,T)$ and is denoted by $A(X,T)$.

Here are some of the results in [80].

1. If $(X,T)$ has positive topological entropy or is weakly mixing for some invariant measure with nontrivial support, then Sep$(\delta, \nu)$ is infinite for some values of $\delta$, $\nu$.

2. Recall that a minimal system $(X,T)$ is called almost automorphic if it is an almost one-to-one extension (meaning that there exists a singleton fiber) of a minimal equicontinuous system [176]. If $(X,T)$ is almost automorphic but not equicontinuous, then Sep$(\delta, \nu)$ is not uniformly bounded in $\delta$ and $\nu$.

3. If $(X,T)$ is an extension of an equicontinuous system $(Y,S)$ with respect to a factor map $\pi : (X,T) \to (Y,S)$ such that $\mu\{y \in Y : \text{card } \pi^{-1}\{y\} > 1\} = 0$ for every invariant measure $\mu$ on $Y$, then Sep$(\delta, \nu) < \infty$ for all $\delta, \nu$. (Regular Toeplitz systems provide such examples.)

4. If $(X,T)$ is a Sturmian subshift, then $A(X,T) = 1$.

5. The set of values of $A(X,T)$ for regular Toeplitz systems $(X,T)$ is dense in $[1, \infty)$.

### 3.12 Inconstancy

Inspired by nineteenth-century ideas of Crofton and Cauchy on the fluctuations of curves and of M. Mendès France [140] on the entropy or temperature of a curve, Allouche and Maillard-Teyssier [6] define the \textit{inconstancy} of a plane curve to be twice its length divided by the perimeter of its convex hull. Mendès-France had suggested the logarithm of this quantity as the \textit{entropy} of the curve. (According to Fechner's law in psychophysics, the magnitude of a sensation is proportional to the logarithm of the intensity of the stimulus.) The inconstancy $I(u)$ of a sequence $u$ of real numbers is defined to be the
inconstancy of the piecewise linear curve obtained by connecting the points on its graph. See [6] for the references, interesting examples, and discussion of possible applications of inconstancy.

For sequences \( u_0u_1 \ldots \) on the alphabet of two symbols 0 and \( h > 0 \) for which the frequencies \( \mu[ab] \) exist for all 2-blocks \( ab \), the authors provide the following results.

1. \( I(u) = 1 + (\sqrt{h^2 + 1} - 1)(\mu[0h] + \mu[h0]) \).

2. If \( d \in \mathbb{N} \), the inconstancy of the periodic sequence \( (0d1)^\infty \) is
   \[
   I((0d1)^\infty) = \frac{d - 1 + 2\sqrt{2}}{d + 1}.
   \]
   This is largest \( (\sqrt{2}) \) for \( d = 1 \) and tends to 1 as \( d \to \infty \) and the curve becomes flat.

3. For a random binary sequence, the inconstancy is \( (1 + \frac{\sqrt{2}}{2}) = 1.207 \ldots \).

4. For the Prouhet-Thue-Morse sequence, the inconstancy is \( (1 + 2\sqrt{2})/3 = 1.276 \ldots \) This follows from the fact that \( \mu[00] = \mu[11] = 1/6, \mu[01] = \mu[10] = 1/3 \). The inconstancy of this sequence is relatively high because it does not contain long strings (length more than two) of 0 or 1.

5. If \( u \) is a Sturmian sequence for which the frequency of 1 is \( \alpha \) and which does not contain the block 11, then
   \[
   I(u) = 1 + 2(\sqrt{2} - 1)\alpha.
   \]

Exercise 3.12.1. Show that no Sturmian sequence contains both blocks 00 and 11.

Exercise 3.12.2. Prove statements (3) and (4) above.

3.13 Measure-theoretic complexity

Ferenczi [74] has proposed upper and lower measure-theoretic complexities, \( P_+T \) and \( P_-T \), as asymptotic growth rates of the number of \( d \)-n-\( \epsilon \) balls needed to cover \((1 - \epsilon)\) of the space. A similar idea is used in [117] for the construction of some \( \mathbb{Z}^2 \) actions and in [161], with the \( f \) rather than \( d \) metric, to show that different Cartesian powers of the horocycle flow are not Kakutani equivalent.

Let \( T : X \to X \) be an ergodic measure-preserving transformation on a probability space \((X, B, \mu)\). For a finite (or sometimes countable) measurable partition of \( X \) and a point \( x \in X \), denote by \( \alpha(x) \) the cell of \( \alpha \) to which \( x \) belongs. For integers \( i \leq j \) we denote by \( \alpha_j \) the partition \( T^{-i}\alpha \vee \ldots \vee T^{-j}\alpha \). The \( d \), or Hamming, distance between two blocks \( B = b_1 \ldots b_n \) and \( C = c_1 \ldots c_n \) is defined to be

\[
\overline{d}(B, C) = \frac{1}{n} \text{card}\{i = 1, \ldots, n : b_i \neq c_i\}.
\]

For \( x \in X, n \in \mathbb{N}, \) and \( \epsilon > 0 \), the \( d \)-\( n \)-\( \epsilon \) ball centered at \( x \) is defined to be

\[
B(\alpha, x, n, \epsilon) = \{y \in X : \overline{d}(\alpha_0^{n-1}(x), \alpha_0^{n-1}(y)) < \epsilon\}.
\]

Define

\[
K(\alpha, n, \epsilon, T) = \min\{K : \text{there are } x_1, \ldots, x_K \in X \text{ such that }\}
\]

\[
\mu(\cup_{i=1}^K B(\alpha, x_i, n, \epsilon)) \geq 1 - \epsilon\}.
\]

Let \( g : \mathbb{N} \to \mathbb{N} \) be an increasing function. Define

\[
P^+_{\alpha, T} < g \text{ to mean } \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{K(\alpha, n, \epsilon, T)}{g(n)} \leq 1 \text{ and}
\]

\[
P^+_{\alpha, T} > g \text{ to mean } \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{K(\alpha, n, \epsilon, T)}{g(n)} \geq 1.
\]
If both relations hold, we write $P_{\alpha,T}^+ \sim g$. Similar notation defines $P_{\alpha,T}^-$ when the lim sups are replaced by lim infs.

The suprema over all partitions $\alpha$ are defined as follows. For an increasing function $g : \mathbb{N} \to \mathbb{N}$, $P_T^+ \prec g$ means that $P_{\alpha,T}^+ \prec g$ for every partition $\alpha$. And $P_T^+ \succ g$ means that for every increasing $h : \mathbb{N} \to \mathbb{N}$ such that $h(n) \leq g(n)$ for all large enough $n$ and with $\lim \sup h(n)/g(n) < 1$ there is a partition $\alpha$ such that $P_{\alpha,T}^+ \succ h$. Similar careful defining produces $P_T^-$. These asymptotic growth rate equivalence classes, $P_T^+$ and $P_T^-$, are called the upper and lower measure-theoretic complexities of $(X, \mathcal{B}, \mu, T)$, respectively.

Ferenczi establishes the following properties of these measure-theoretic complexity “functions”.

1. If $\alpha$ is a generating partition, then $P_T^+ \sim P_{\alpha,T}^+$ and $P_T^- \sim P_{\alpha,T}^-$. 
2. $P_T^+$ and $P_T^-$ are invariants of measure-theoretic isomorphism.

3. $\lim \log P_T^+(n)/n = \lim \log P_T^-(n)/n = h_\mu(T)$, in the sense that $P_T^+$ and $P_T^-$ are dominated by $e^{n(h+\epsilon)}$ for every $\epsilon > 0$ and dominate $e^{n(h-\epsilon)}$ for every $\epsilon > 0$.

4. $(X, \mathcal{B}, \mu, T)$ is measure-theoretically isomorphic to a translation on a compact group if and only if $P_T^+ \prec g$ for every unbounded increasing $g : \mathbb{N} \to \mathbb{N}$. (Same for $P_T^-$.) This is to be compared with the characterization of topological dynamical systems with bounded topological complexity functions as the translations on compact groups—see Sections 3.1 and 3.6.

5. For the Morse system, $P_T^+ \sim 10n/3$ and $P_T^- \sim 3n$.
6. For rank one systems, $P_T^- \prec an^2$ for every $a > 0$.

7. For the Chacon system (the orbit closure of the fixed point of the substitution $0 \to 0010, 1 \to 1$), $P_T^+ \sim P_T^- \sim 2n$.

### 3.14 Pattern complexity

Kamae and Zamboni [111, 112] have defined and studied maximal pattern complexity for infinite sequences, which shares some features with sequence entropy (Sections 2.3 and 2.4) and average sample complexity (Section 4.4). A pattern is defined to be a finite increasing sequence $\tau = \tau(0)\tau(1)\tau(2)\ldots\tau(k-1)$ of integers with $\tau(0) = 0$. If $u = u_0u_1\ldots$ is an infinite one-sided sequence on a finite alphabet, one considers the set of words seen at places in $u$ along the pattern $\tau$, namely

\[(3.14.1) \quad F_\tau(u) = \{u_iu_{i+\tau(1)}\ldots u_{i+\tau(k-1)} : i = 0, 1, 2, \ldots\}.\]

The maximal pattern complexity function of $u$ is defined to be

\[(3.14.2) \quad P_u^*(k) = \sup\{|F_\tau(u)| : |\tau| = k\}, k = 1, 2, \ldots.\]

The authors establish the following results.

1. Let $u$ be an infinite sequence on a finite alphabet, and suppose that the weak* limit of $\sum_{i=0}^{n-1} \delta_{\tau_iu_i}/n$ exists and so defines a shift-invariant measure $\mu$ on the orbit closure $X$ of $u$. Then for any sequence $S = (t_i)$, the sequence entropy (see Section 2.3) is bounded in terms of the maximal pattern complexity: with $\alpha$ the usual time-0 partition of $X$,

\[(3.14.3) \quad h^S_{\mu}(\alpha, \sigma) \leq \lim \sup_{k \to \infty} \frac{\log P_u^*(k)}{k}.\]

Hence, if $(X, \sigma, \mu)$ does not have purely discrete spectrum, the maximal pattern complexity function $P_u^*(k)$ has exponential growth rate.

2. If there exists $k$ such that $P_u^*(k) < 2k$, then $u$ is eventually periodic (cf. Exercise 3.1.2).
3. For Sturmian (and generalized Sturmian) sequences, $P^*_u(k) = 2k$ for all $k$.

4. For the Prouhet-Thue-Morse sequence, $P^*_u(k) = 2^k$ for all $k$.

Further developments concerning pattern complexity, including possible applications to image recognition, can be found in [76, 110, 186] and their references.

**Exercise 3.14.1.** The *window complexity function* $P^w_u(n)$ of an infinite sequence $u$ on a finite alphabet $A$, studied in [49], counts the number of different blocks of length $n$ seen in the sequence $u$ beginning at coordinates that are multiples of $n$:

$$P^w_u(n) = \text{card}\{u_{kn}u_{kn+1}\ldots u_{(k+1)n-1} : k \geq 0\}.$$  

In the spirit of pattern complexity, one could also consider counting blocks of length $n$ that appear in a sequence $u$ along an arithmetic progression of coordinates:

$$P^a_u(n) = \text{card}\{u_iu_{i+k}u_{i+2k}\ldots u_{i+(n-1)k} : i \geq 0, k \geq 1\}.$$  

Discuss the relationships among the ordinary complexity function $p_u$ (Section 3.1), $P^w_u$, $P^a_u$, and $P^*_u$. Compute them for Sturmian sequences and the Prouhet-Thue-Morse sequence.

**Exercise 3.14.2.** Consider the *Oldenburger-Kolakoski sequence*

$$\omega = 122112122122112\ldots,$$

which coincides with its sequence of run lengths. Can you say anything about any measure of complexity for this sequence? Anything about its orbit closure as a subshift? (It seems to be an open problem whether the frequency of 1 exists, although numerical studies show that it seems to be 1/2. So finding any invariant measure that it defines could be difficult...)

What about a sequence on the alphabet $\{2, 3\}$ that has the same self-similarity property?

Or consider all such sequences on $\{1, 2, 3\}$? Since now some choices seem possible, the orbit closure should have positive topological entropy?
Part 4

Balancing freedom and interdependence

4.1 Neurological intricacy

In any system that consists of individual elements there is a tension between freedom of action of the individuals and coherent action of the entire system. There is maximum complexity, information, disorder, or randomness when all elements act independently of one another. At the other extreme, when the elements of the system are strongly linked and the system acts essentially as a single unit, there is maximum order. In the first situation, it seems that there is little advantage to the individual elements in being part of a larger system, and the system does not benefit from possible concerted action by its constituents. In the second situation, most individual elements could be superfluous, and the system does not benefit from any diversity possibly available from its parts. This tension between the one and the many, the individual and the state, the soloist and the orchestra, the part and the whole, is ancient and well known.

It is natural to think that evolving organisms or societies may seek a balance in which the benefits of diversity and coherence are balanced against their disadvantages. Abrams and Panaggio [1] constructed a differential equations model to describe the balance between competitive and cooperative pressures to attempt to explain the prevalence of right-handedness in human populations. (Left-handers may have an advantage in physical competitions, where opponents are accustomed to face mostly right-handers, and the population as a whole may benefit from diversity. But left-handers will be at a disadvantage when faced with objects and situations designed for the comfort of the prevalent right-handers.) Blei [27] defined measures of interdependence among families of functions in terms of functional dependence among subfamilies using combinatorics, functional analysis, and probability.

Neuroscientists G. Edelman, O. Sporns, G. Tononi [175] proposed a measure, which they called “neural complexity”, of the balance between specific and mass action in the brains of higher vertebrates. High values of this quantity are associated with non-trivial organization of the network; when this is the case, segregation coexists with integration. Low values are associated with systems that are either completely independent (segregated, disordered) or completely dependent (integrated, ordered). We will see that beneath this concept of intricacy there is another (new) basic notion of complexity, that we call average sample complexity. The definitions and study of intricacy and average sample complexity in dynamics were initiated in [153, 185].

One considers a model that consists of a family $X = \{X_i : i = 0, 1, \ldots, n - 1\}$ of random variables representing an isolated neural system with $n$ elementary components (maybe groups of neurons), each $X_i$ taking values in a discrete (finite or countable) set $E$. For each $n \in \mathbb{N}$ we define $n^* = \{0, 1, \ldots, n - 1\}$. The set $n^*$ represents the set of sites, and $E$ represents the set of states. It may seem that we are assuming that the brain is one-dimensional, but not so: the sites may be arranged in some geometrically important way, but at this stage we only number them and will take the geometry, distances, connections, etc. into account maybe at some later stage. The elements of the set $E$ (often $E = \{0, 1\}$) may encode (probably quantized) levels of excitation or something analogous. For $S \subset n^*$, $X_S = \{X_i : i \in S\}$, $S^c = n^* \setminus S$. Neural complexity measures the level of interdependence between action at the sites in $S$ and those in $S^c$, averaged over all subset $S$ of the set of sites, with some choice of weights.
The entropy of a random variable $X$ taking values in a discrete set $E$ is

$$H(X) = -\sum_{x \in E} \Pr\{X = x\} \log \Pr\{X = x\}.$$ 

The mutual information between two random variables $X$ and $Y$ over the same probability space $(\Omega, \mathcal{F}, P)$ is

$$MI(X, Y) = H(X) + H(Y) - H(X, Y).$$

$$= H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Here $(X, Y)$ is the random variable on $\Omega$ taking values in $E \times E$ defined by $(X, Y)(\omega) = (X(\omega), Y(\omega))$. $MI(X, Y)$ is a measure of how much $Y$ tells about $X$ (equivalently, how much $X$ tells about $Y$). $MI(X, Y) = 0$ if and only if $X$ and $Y$ are independent.

The neural complexity, $C_N$, of the family $X = \{X_i : i = 0, 1, \ldots, n - 1\}$ is defined to be the following average, over all subfamilies $X_S = \{X_i : i \in S\}$, of the mutual information between $X_S$ and $X_{S^c}$:

$$C_N(X) = \frac{1}{n+1} \sum_{S \subset n^*} \frac{1}{n} MI(X_S, X_{S^c}).$$

The weights are chosen to be uniform over all subsets of the same size, and then uniform over all sizes.

J. Buzzi and L. Zambotti [43] studied neural complexity in a general probabilistic setting, considering it as one of a family of functionals on processes that they called intricacies, allowing more general systems of weights for the averaging of mutual informations. They define a system of coefficients, $c^n_S$, to be a family of numbers satisfying for all $n \in \mathbb{N}$ and $S \subset n^*$

1. $c^n_S \geq 0$;
2. $\sum_{S \subset n^*} c^n_S = 1$;
3. $c^n_{S^c} = c^n_S$.

For a fixed $n \in \mathbb{N}$ let $X = \{X_i : i \in n^*\}$ be a collection of random variables all taking values in the same finite set. Given a system of coefficients, $c^n_S$, the corresponding mutual information functional, $\mathcal{I}^c(X)$, is defined by

$$\mathcal{I}^c(X) = \sum_{S \subset n^*} c^n_S MI(X_S, X_{S^c}).$$

**Definition 4.1.1.** An intricacy is a mutual information functional satisfying:

1. Exchangeability: invariance under permutations of $n^*$;
2. Weak additivity: $\mathcal{I}^c(X, Y) = \mathcal{I}^c(X) + \mathcal{I}^c(Y)$ for any two independent systems $X = \{X_i : i \in n^*\}$ and $Y = \{Y_j : j \in m^*\}$.

**Theorem 4.1.2** (Buzzi-Zambotti). Let $c^n_S$ be a system of coefficients and $\mathcal{I}^c$ the associated mutual information functional. $\mathcal{I}^c$ is an intricacy if and only if there exists a symmetric probability measure $\lambda_c$ on $[0, 1]$ such that

$$c^n_S = \int_{[0,1]} x^{|S|} (1-x)^{n-|S|} \lambda_c(dx).$$

**Example 4.1.3.** 1. $c^n_S = \frac{1}{(n+1) \binom{n}{|S|}}$ (Edelman-Sporns-Tononi);

2. For $0 < p < 1$,

$$c^n_S = \frac{1}{2} \left( p^{|S|}(1-p)^{n-|S|} + (1-p)^{|S|} p^{n-|S|} \right) (p\text{-symmetric});$$
3. For \( p = 1/2 \), \( c_n^S = 2^{-n} \) (uniform).

**Exercise 4.1.1.** Prove that the neural (Edelman-Sporns-Tononi) weights correspond to \( \lambda \) being Lebesgue measure on \([0,1]\).

### 4.2 Topological intricacy and average sample complexity

Let \((X, T)\) be a topological dynamical system and \(\mathcal{U}\) an open cover of \(X\). Given \(n \in \mathbb{N}\) and a subset \(S \subset n^*\) define

\[\mathcal{U}_S = \bigvee_{i \in S} T^{-i} \mathcal{U}.\]

**Definition 4.2.1.** [153,185] Let \(c_n^S\) be a system of coefficients. Define the topological intricacy of \((X, T)\) with respect to the open cover \(\mathcal{U}\) to be

\[\text{Int}(X, \mathcal{U}, T) := \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_n^S \log \left( \frac{N(\mathcal{U}_S)N(\mathcal{U}_S^c)}{N(\mathcal{U}_n^c)} \right).\]

Applying the laws of logarithms and noting the symmetry of the sum with respect to sets \(S\) and their complements leads one to define the following quantity.

**Definition 4.2.2.** [153, 185] The topological average sample complexity of \(T\) with respect to the open cover \(\mathcal{U}\) is defined to be

\[\text{Asc}(X, \mathcal{U}, T) := \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_n^S \log N(\mathcal{U}_S).\]

**Proposition 4.2.3.** \(\text{Int}(X, \mathcal{U}, T) = 2 \text{Asc}(X, \mathcal{U}, T) - h_{\text{top}}(X, \mathcal{U}, T)\).

**Theorem 4.2.4.** The limits in the definitions of \(\text{Int}(X, \mathcal{U}, T)\) and \(\text{Asc}(X, \mathcal{U}, T)\) exist.

As usual this follows from subadditivity of the sequence

\[(4.2.1) \quad b_n := \sum_{S \subset n^*} c_n^S \log N(\mathcal{U}_S)\]

and Fekete’s Subadditive Lemma: For every subadditive sequence \(a_n\), \(\lim_{n \to \infty} a_n/n\) exists and is equal to \(\inf_n a_n/n\).

**Exercise 4.2.1.** Prove Fekete’s Lemma.

**Proposition 4.2.5.** For each open cover \(\mathcal{U}\), \(\text{Asc}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T)\), and hence \(\text{Int}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T)\).

In particular, a dynamical system with zero (or relatively low) topological entropy (one that is coherent or ordered) has zero (or relatively low) topological intricacy.

The intricacy of a subshift \((X, \sigma)\) with respect to the “time zero open cover” \(\mathcal{U}_0\) by cylinder sets defined by the first (or central) coordinate is determined by counting the numbers of different blocks that can be seen along specified sets of coordinates:

\[\text{Int}(X, \mathcal{U}_0, \sigma) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_n^S \log \left( \frac{|\mathcal{L}(X)| |\mathcal{L}(\mathcal{U}_S^c)(X)|}{|\mathcal{L}(\mathcal{U}_n^c)(X)|} \right)\]

**Example 4.2.6** (Computing \(|\mathcal{L}(X)|\) for the golden mean SFT). Let \(n = 3\), so that \(n^* = \{0,1,2\}\). The following figure shows how different numbers of blocks can appear along different sets of coordinates of the same cardinality: if \(S = \{0,1\}\) then \(N(S) = 3\), whereas if \(S = \{0,2\}\), \(N(S) = 4\).
Part 4. Balancing freedom and interdependence

$S = \{0, 1\}$  $S = \{0, 2\}$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$|L_S(X)| = 3$  $|L_S(X)| = 4$

When we average over all subsets $S \subset n^*$, we get an approximation (from above) to Int$(X, \emptyset_0, \sigma)$:

| $S$  | $S$ | $|L_S(X)|$ | $|L_{S^c}(X)|$ |
|------|------|----------|----------------|
| $\emptyset$  | $\{0, 1, 2\}$  | 1  | 5  |
| $\{0\}$  | $\{1, 2\}$  | 2  | 3  |
| $\{1\}$  | $\{0, 2\}$  | 2  | 4  |
| $\{2\}$  | $\{0, 1\}$  | 2  | 3  |
| $\{0, 1\}$  | $\{2\}$  | 3  | 2  |
| $\{0, 2\}$  | $\{1\}$  | 4  | 2  |
| $\{1, 2\}$  | $\{0\}$  | 3  | 2  |
| $\{0, 1, 2\}$  | $\emptyset$  | 5  | 1  |

**Example 4.2.7** (Computing $|L_S(X)|$ for the golden mean SFT).

$$\frac{1}{3 \cdot 2^3} \sum_{S \subset n^*} \log \left( \frac{|L_S(X)| |L_{S^c}(X)|}{|L_{n^*}(X)|} \right) = \frac{1}{24} \log \left( \frac{6^4 \cdot 8^2}{5^6} \right) \approx 0.070.$$  

Apparently one needs better formulas for Int and Asc than the definitions, which involve exponentially many calculations as $n$ grows. Here is a formula that applies to many SFT’s and permits accurate numerical estimates.

**Theorem 4.2.8.** Let $X$ be a shift of finite type with adjacency matrix $M$ such that $M^2 > 0$. Let $c^n_S = 2^{-n}$ for all $S$. Then

$$\text{Asc}(X, \emptyset_0, \sigma) = \frac{1}{4} \sum_{k=1}^{\infty} \log \frac{|L_k^*(X)|}{2^k}.$$  

This formula shows that, as expected, Asc is sensitive to word counts of all lengths and thus is a finer measurement than $h_{\text{top}}$, which just gives the asymptotic exponential growth rate. Below we will see examples of systems that can be distinguished by Asc and Int but not by their entropies, or even by their symbolic complexity functions.

The main ideas of the proof are:

1. Each $S \subset n^*$ decomposes into a union of disjoint intervals $I_j$ separated by gaps of length at least 1.
2. Because $M^2 > 0$, $\log N(S) = \sum_j \log N(I_j)$.
3. We may consider the $S \subset n^*$ that do not contain $n - 1$, then those that do, and use induction.

**Corollary 4.2.9.** For the full $r$-shift with $c^n_S = 2^{-n}$ for all $S$,

$$\text{Asc}(\Sigma_r, \emptyset_0, \sigma) = \frac{\log r}{2} \quad \text{and} \quad \text{Int}(\Sigma_r, \emptyset_0, \sigma) = 0.$$
In the following table we compare $h_{\text{top}}, \text{Int, and Asc}$ for the full 2-shift, the golden mean SFT, and the subshift consisting of a single periodic orbit of period two. The first is totally disordered, while the third is completely deterministic, so each of these has intricacy zero, while the golden mean SFT has some balance between freedom and discipline.

<table>
<thead>
<tr>
<th>Adjacency Graph</th>
<th>Entropy</th>
<th>Asc</th>
<th>Int</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disordered</td>
<td>0.693</td>
<td>0.347</td>
<td>0</td>
</tr>
<tr>
<td>Ordered</td>
<td>0.481</td>
<td>0.286</td>
<td>0.090</td>
</tr>
</tbody>
</table>

As with the definitions of topological and measure-theoretic entropies, one may seek to define an isomorphism invariant by taking the supremum over all open covers (or partitions). But this will lead to nothing new.

**Theorem 4.2.10.** Let $(X, T)$ be a topological dynamical system and fix the system of coefficients to be $c^n_S = 2^{-n}$. Then

$$\sup_{\mathcal{U}} \text{Asc}(X, \mathcal{U}, T) = h_{\text{top}}(X, T).$$

**Proof.** The proof depends on the structure of average subsets of $n^* = \{0, 1, \ldots, n - 1\}$: most $S \subset n^*$ have size about $n/2$, so are not too sparse.

When computing the ordinary topological entropy of a subshift, to get at the supremum over open covers it is enough to start with the time-0 partition (or open cover) $\alpha$, then iterate and refine, replacing $\alpha$ by $\alpha_k = \alpha_0^{k-1}$. Then for fixed $k$, when we count numbers of blocks (configurations), we are looking at $\alpha_{(n+k)^*}$ instead of $\alpha_{n^*}$; and when $k$ is fixed, as $n$ grows the result is the same.

When computing $\text{Asc}$ and $\text{Int}$, start with the time-0 partition, and code by $k$-blocks. Then $S \subset n^*$ is replaced by $S + k^*$, and the effect on $\alpha_{S+k^*}$ as compared to $\alpha_S$ is similar, since it acts similarly on each of the long subintervals comprising $S$.

Here is a still sketchy but slightly more detailed indication of the idea. Fix a $k$ for coding by $k$-blocks (or looking at $N((\mathcal{U}_k)_S)$ or $H((\alpha_k)_S)$). Cut $n^*$ into consecutive blocks of length $k/2$. When $s \in S$ is in one of these intervals of length $k/2$, then $s + k^*$ covers the next interval of length $k/2$. So if $S$ hits many of the intervals of length $k/2$, then $S + k^*$ starts to look like a union of long intervals, say each with $|E_j| > k$. By shaving a little off each of these relatively long intervals, we can assume that also the gaps have length at least $k$.

Given $\epsilon > 0$, we may assume that $k$ is large enough that for every interval $I \subset \mathbb{N}$ with $|I| \geq k/2$,

$$0 \leq \frac{\log N(I)}{\text{card}(I)} - h_{\text{top}}(X, \sigma) < \epsilon.$$
We let $\mathcal{B}$ denote the set of $S \subset n^*$ which miss at least $2n\epsilon/k$ of the intervals of length $k/2$ and show that
\[ \lim_{n \to \infty} \frac{\text{card}(\mathcal{B})}{2^n} = 0. \]
If $S \notin \mathcal{B}$, then $S$ hits many of the intervals of length $k/2$, and hence $S + k^*$ is the union of intervals of length at least $k$, and we can arrange that the gaps are also long enough to satisfy the estimate in 4.2.2 comparing (average of log of) number of blocks to $h_{\text{top}}(X, \sigma)$.

**Exercise 4.2.2.** Prove that $\sup_{\mathcal{A}} \text{Int}(X, \mathcal{A}, T) = h_{\text{top}}(X, T)$ (for the system of coefficients $c^0_S = 2^{-n}$).

### 4.3 Ergodic-theoretic intricacy and average sample complexity

We turn now to the formulation and study of the measure-theoretic versions of intricacy and average sample complexity. For a partition $\alpha$ of $X$ and a subset $S \subset n^*$ define
\[ \alpha_S = \bigvee_{i \in S} T^{-i}\alpha. \]

**Definition 4.3.1.** [153, 185] Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, $\alpha = \{A_1, \ldots, A_n\}$ a finite measurable partition of $X$, and $c^\beta_S$ a system of coefficients. The *measure-theoretic intricacy of $T$ with respect to the partition $\alpha$* is
\[ \text{Int}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^\beta_S \left[ H_\mu(\alpha_S) + H_\mu(\alpha_{S^c}) - H_\mu(\alpha_{n^*}) \right]. \]

The *measure-theoretic average sample complexity of $T$ with respect to the partition $\alpha$* is
\[ \text{Asc}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^\beta_S H_\mu(\alpha_S). \]

**Theorem 4.3.2.** The limits in the definitions of measure-theoretic intricacy and measure-theoretic average sample complexity exist.

**Theorem 4.3.3.** Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and fix the system of coefficients $c^\beta_S = 2^{-n}$. Then
\[ \sup_\alpha \text{Asc}_\mu(X, \alpha, T) = \sup_\alpha \text{Int}_\mu(X, \alpha, T) = h_\mu(X, T). \]

The proofs are similar to those for the corresponding theorems in topological setting. These observations indicate that there may be a topological analogue of the following result.

**Theorem 4.3.4** (Ornstein-Weiss, 2007). If $J$ is a finitely observable functional defined for ergodic finite-valued processes that is an isomorphism invariant, then $J$ is a continuous function of the measure-theoretic entropy.

Here $X = (x_1, x_2, \ldots)$ is a finite-state ergodic stochastic process, and a “finitely observable functional” is the a.s. limit $F(X)$ of a sequence of functions $f_n(x_1, x_2, \ldots, x_n)$ taking values in some metric space which for every such process converges almost surely. The integral of $x_1$ and the entropy of the process are examples of finitely observable functionals.

### 4.4 The average sample complexity function

The observations in the preceding situation suggest that one should examine these Asc and Int *locally*. For example, for a fixed open cover $\mathcal{U}$, fix a $k$ and find the topological average sample complexity
\[ \text{Asc}(X, \mathcal{U}, k, \sigma) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^\beta_S \log N((\mathcal{U}_k)_S). \]
4.4. The average sample complexity function

Figure 4.1: Graphs of two subshifts with the same complexity function but different average sample complexity functions.

The average sample complexity function

\[ (M_I)^3 > 0 \quad (M_{II})^4 > 0 \]

Figure 4.4.1 shows two SFT’s that have the same number of \( n \)-blocks for every \( n \) but different Asc functions.

Example 4.4.1.

\[ \text{Asc}(n) = \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq \mathcal{N}^n} \log N(S) \]
Numerical evidence (up to \( n = 10 \)) indicates that these two SFT's have different values of Asc and Int, although they have identical complexity functions and hence the same topological entropy.

<table>
<thead>
<tr>
<th>Adjacency Graph</th>
<th>( h_{\text{top}} )</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.481</td>
<td>0.399</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>0.481</td>
<td>0.377</td>
<td>0.208</td>
</tr>
</tbody>
</table>

### 4.5 Computing measure-theoretic average sample complexity

For a fixed partition \( \alpha \), we develop a relationship between \( \text{Asc}_\mu(X,\alpha,T) \) and a series summed over \( i \) involving the conditional entropies \( H_\mu(\alpha_i|\alpha_i^\infty) \). The series can serve as a computational tool analogous to the series in Theorem 4.2.8.

The idea is to view a subset \( S \subset n^* \) as corresponding to a random binary string of length \( n \) generated by the Bernoulli measure \( B(1/2, 1/2) \) on the full 2-shift. For example \( \{0, 2, 3\} \subset 5^* \leftrightarrow 10110 \). The average entropy, \( H_\mu(\alpha_\infty) \), over all \( S \subset n^* \), is then an integral and can be interpreted in terms of the entropy of a first-return map to the cylinder \( A = [1] \) in a cross product of our system \( X \) and the full 2-shift, \( \Sigma_2 \).

**Theorem 4.5.1.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system and \( \alpha \) a finite measurable partition of \( X \). Let \( A = [1] = \{\xi \in \Sigma_2^+: \xi_0 = 1\} \) and \( \beta = \alpha \times A \) the related finite partition of \( X \times A \). Denote by \( T_{X \times A} \) the first-return map on \( X \times A \) and let \( P_A = P/P[1] \) denote the measure \( P \) restricted to \( A \) and normalized. Let \( c_3^2 = 2^{-n} \) for all \( S \subset n^* \). Then

\[
\text{Asc}_\mu(X,\alpha,T) = \frac{1}{2} h_{\mu \times P_A}(X \times A, \beta, T_{X \times A}).
\]

Applying the definition of the entropy of a transformation with respect to a fixed partition as the integral of the corresponding information function and breaking up the integral into a sum of integrals over sets where the first-return time to \( X \times A \) takes a fixed value produces the following result.

**Theorem 4.5.2.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system and \( \alpha \) a finite measurable partition of \( X \). Let \( c_3^2 = 2^{-n} \) for all \( S \subset n^* \). Then

\[
\text{Asc}_\mu(X,\alpha,T) \geq \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} H_\mu(\alpha_i|\alpha_i^\infty).
\]

Equality holds in certain cases (in particular, for Markov shifts)

### 4.6 The search for maximizing measures on subshifts

Given a topological dynamical system \((X,T)\), one would like to find the measures that maximize Asc and Int, since the nature of these measures might tell us a lot about the balance between freedom and determinism within the system. For ordinary topological entropy and topological pressure with respect to a given potential, maximizing measures (measures of maximal entropy, equilibrium states) are of great
importance and are regarded as natural measures on the system. We hope that Theorem 4.5.2 might be helpful in the identification of these extremal measures. In the topological case the first-return map $T_{X \times A}$ is not continuous nor expansive nor even defined on all of $X \times A$ in general, so known results about measures of maximal entropy and equilibrium states do not apply. To maximize $\text{Int}$, there is the added problem of the minus sign in

$$\text{Int}(X, \mathcal{U}, T) = 2 \text{Asc}(X, \mathcal{U}, T) - h_{\text{top}}(X, \mathcal{U}, T).$$

Maybe some modern work on local or relative variational principles, almost subadditive potentials, equilibrium states for shifts with infinite alphabets, etc. could apply? (See [17, 18, 44, 50, 73, 100, 103, 106, 145, 187] etc.)

But the above theorem does give up some information immediately:

**Proposition 4.6.1.** When $T : X \to X$ is an expansive homeomorphism on a compact metric space (e.g., $(X, T)$ is a subshift on finite alphabet), $\text{Asc}_\mu(X, T, \alpha)$ is an affine upper semicontinuous (in the weak* topology) function of $\mu$, so the set of maximal measures for $\text{Asc}_\mu(X, T, \alpha)$ is nonempty, compact, and convex and contains some ergodic measures (see [181, p. 198 ff.]).

We try now to find measures of maximal $\text{Asc}$ or $\text{Int}$ on SFT’s, or at least maximal measures among all Markov measures of a fixed memory. Recall that a measure of maximal entropy on an SFT is unique and is a Markov measure, called the Shannon-Parry measure, denoted here by $\mu_{\text{max}}$. Further, given a potential function $\phi$ that is a function of just two coordinates, again there is a unique equilibrium measure that maximizes

$$P_\mu(\phi) = h_\mu(\sigma) + \int_X \phi d\mu.$$

See [149].

A 1-step Markov measure on the full shift space $(\Sigma_n, \sigma)$ is given by a stochastic matrix $P = (P_{ij})$ and fixed probability vector $p = \left( p_0, p_1, \ldots, p_{n-1} \right)$, i.e. $\sum p_i = 1$ and $pP = p$. The measure $\mu_{P,p}$ is defined as usual on cylinder sets by

$$\mu_{P,p}[i_0i_1\ldots i_k] = p_{i_0}P_{i_0i_1}\cdots P_{i_{k-1}i_k}.$$

**Example 4.6.2** (1-step Markov measure on the golden mean shift). Denote by $P_{00} \in [0,1]$ the probability of going from 0 to 0 in a sequence of $X_{(11)} \subset \Sigma_2$. Then

$$P = \begin{pmatrix} P_{00} & 1 - P_{00} \\ 1 - P_{00} & 0 \end{pmatrix}, \quad p = \begin{pmatrix} \frac{1}{2 - P_{00}} & \frac{1 - P_{00}}{2 - P_{00}} \end{pmatrix}.$$

Using the series formula in Theorem 4.5.2 and known equations for conditional entropy, we can approximate $\text{Asc}_\mu$ and $\text{Int}_\mu$ for Markov measures on SFTs. Let’s look first at 1-step Markov measures.
Note that the maximum value of \( h_\mu = h_{\text{top}} = \log \phi \) occurs when \( P_{00} = 1/\phi \); there are unique maxima among 1-step Markov measures for \( \text{Asc}_\mu \) and \( \text{Int}_\mu \); and the maxima for \( \text{Asc}_\mu \), \( \text{Int}_\mu \), and \( h_\mu \) are achieved by different measures.

Now let’s calculate \( \text{Asc} \) and \( \text{Int} \) for various 2-step Markov measures on the golden mean SFT.

**2-step Markov measures on the golden mean shift**

\[
\begin{array}{cccc}
P_{00} & h_\mu & \text{Asc}_\mu & \text{Int}_\mu \\
0.618 & 0.481 & 0.266 & 0.051 \\
0.533 & 0.471 & 0.271 & 0.071 \\
0.216 & 0.292 & 0.208 & 0.124 \\
\end{array}
\]

\( \text{Asc}_\mu \) appears to be strictly convex, so it would have a unique maximum among 2-step Markov measures. \( \text{Int}_\mu \) appears to have a unique maximum among 2-step Markov measures on a proper subshift (\( P_{000} = 0 \)). The maxima for \( \text{Asc}_\mu \), \( \text{Int}_\mu \), and \( h_\mu \) are achieved by different measures, and are different from the measures that are maximal among 1-step Markov measures.

Let’s move from the golden mean SFT to the full 2-shift.
4.6. The search for maximizing measures on subshifts

Calculations for one-step Markov measure on the golden mean shift

1-step Markov measures on the full 2-shift

<table>
<thead>
<tr>
<th>$P_{00}$</th>
<th>$P_{11}$</th>
<th>$h_\mu$</th>
<th>$\text{Asc}_\mu$</th>
<th>$\text{Int}_\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.693</td>
<td>0.347</td>
<td>0</td>
</tr>
<tr>
<td>0.216</td>
<td>0.0</td>
<td>0.292</td>
<td>0.208</td>
<td>0.124</td>
</tr>
<tr>
<td>0.0</td>
<td>0.216</td>
<td>0.292</td>
<td>0.208</td>
<td>0.124</td>
</tr>
<tr>
<td>0.905</td>
<td>0.905</td>
<td>0.315</td>
<td>0.209</td>
<td>0.104</td>
</tr>
</tbody>
</table>

$\text{Asc}_\mu$ appears to be strictly convex, so it would have a unique maximum among 1-step Markov measures. $\text{Int}_\mu$ appears to have two maxima among 1-step Markov measures on proper subshifts ($P_{00} = 0$ and $P_{11} = 0$). There seems to be a 1-step Markov measure that is fully supported and is a local maximum for $\text{Int}_\mu$ among all 1-step Markov measures. The maxima for $\text{Asc}_\mu$, $\text{Int}_\mu$, and $h_\mu$ are achieved by different measures.

We summarize some of the questions generated above.

**Conjecture 4.6.3.** On the golden mean SFT, for each $r$ there is a unique $r$-step Markov measure $\mu_r$ that maximizes $\text{Asc}_\mu(X,\sigma,\alpha)$ among all $r$-step Markov measures.

**Conjecture 4.6.4.** $\mu_2 \neq \mu_1$

**Conjecture 4.6.5.** On the golden mean SFT there is a unique measure that maximizes $\text{Asc}_\mu(X,T,\alpha)$. It is not Markov of any order (and of course is not the same as $\mu_{\text{max}}$).

**Conjecture 4.6.6.** On the golden mean SFT for each $r$ there is a unique $r$-step Markov measure that maximizes $\text{Int}_\mu(X,T,\alpha)$ among all $r$-step Markov measures.
Table 4.1: Calculations for one-step Markov measures on the golden mean shift. Numbers in bold are maxima for the given categories.

<table>
<thead>
<tr>
<th>$P_{00}$</th>
<th>$h_\mu$</th>
<th>Asc$_\mu$</th>
<th>Int$_\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.618</td>
<td>0.481</td>
<td>0.266</td>
<td>0.051</td>
</tr>
<tr>
<td>0.533</td>
<td>0.471</td>
<td>0.271</td>
<td>0.071</td>
</tr>
<tr>
<td>0.216</td>
<td>0.292</td>
<td>0.208</td>
<td>0.124</td>
</tr>
</tbody>
</table>

Table 4.2: Calculations for two-step Markov measures on the golden mean shift.

<table>
<thead>
<tr>
<th>$P_{000}$</th>
<th>$P_{100}$</th>
<th>$h_\mu$</th>
<th>Asc$_\mu$</th>
<th>Int$_\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.618</td>
<td>0.618</td>
<td>0.481</td>
<td>0.266</td>
<td>0.051</td>
</tr>
<tr>
<td>0.483</td>
<td>0.569</td>
<td>0.466</td>
<td>0.272</td>
<td>0.078</td>
</tr>
<tr>
<td>0</td>
<td>0.275</td>
<td>0.344</td>
<td>0.221</td>
<td>0.167</td>
</tr>
</tbody>
</table>

Conjecture 4.6.7. On the 2-shift there are two 1-step Markov measures that maximize Int$_\mu(X, T, \alpha)$ among all 1-step Markov measures. They are supported on the golden mean SFT and its image under the dualizing map 0 ↔ 1.

Conjecture 4.6.8. On the 2-shift there is a 1-step Markov measure that is fully supported and is a local maximum point for Int$_\mu(X, T, \alpha)$ among all 1-step Markov measures.

The conjectures extend to arbitrary shifts of finite type and other dynamical systems. Many other natural questions suggested by the definitions and properties established so far of intricacy and average sample complexity can be found in the dissertation of Ben Wilson [185]:

1. We do not know whether a variational principle $\sup_{\mu} \text{Asc}_\mu(X, T, \alpha) = \text{Asc}_{\text{top}}(X, \alpha, T)$ holds.

2. Analogous definitions, results, and conjectures exist when entropy is generalized to pressure, by including a potential function which measures the energy or cost associated with each configuration.
4.6. The search for maximizing measures on subshifts

Two-step Markov measure on the golden mean shift

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.2.png}
\caption{Combination of the plots of $h_\mu$, $\text{Asc}_\mu$, and $\text{Int}_\mu$ for two-step Markov measures on the golden mean shift.}
\end{figure}

First one can consider a function of just a single coordinate that gives the value of each symbol. Maximum intricacy may be useful for finding areas of high information activity, such as working regions in a brain (Edelman-Sporns-Tononi) or coding regions in genetic material (Koslicki-Thompson).

3. Higher-dimensional versions, where subsets $S$ of coordinates are replaced by patterns, are naturally defined and waiting to be studied.

4. One can define and then study average sample complexity of individual finite blocks.

5. We need formulas for $\text{Asc}$ and $\text{Int}$ for more subshifts and systems.

6. Find the subsets or patterns $S$ that maximize $\log N(\mathcal{U}_S)$ or $\log[N(\mathcal{U}_S)N(\mathcal{U}_{S^c})]/N(\mathcal{U}_{n^*})$, and similarly for the measure-preserving case.

7. In the topological case, what are the natures of the quantities that arise if one changes the definitions of $\text{Alt}$ and $\text{Int}$ by omitting the logarithms?

8. Consider not just subsets $S$ and their complements, but partitions of $n^*$ into a finite number of subsets. For the measure-preserving case, there is a natural definition of the mutual information among a finite family of random variables on which one could base the definition of intricacy.

We welcome help in resolving such questions and exploring further the ideas of intricacy, average sample complexity, and complexity in general!


144. _____, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42. MR 0000745 (1,123d)


Department of Mathematics
CB 3250 Phillips Hall
University of North Carolina
Chapel Hill, NC 27599 USA
petersen@math.unc.edu