MEASURING COMPLEXITY IN CANTOR DYNAMICS (SHORT VERSION)

KARL PETERSEN

Contents

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. The first 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 second 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 third 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 [3, 11, 29, 31, 47, 48, 53, 69].

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 [20, 54]). 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 below. 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 $\mu(X) = 1$. 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 ϕ is also one-to-one, then (X, T) and (Y, M) are *topologically conjugate* and ϕ 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 ϕ is one-to-one a.e., equivalently $\phi^{-1}\mathcal{C} = \mathcal{B}$ up to sets of measure 0, then ϕ is an *isomorphism*.

We focus in these notes especially on topological dynamical systems which are *shift dynamical systems*. Let $\mathcal{A}$ 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 $\mathcal{A}$-shift*, $\Sigma(\mathcal{A})$, is the collection of all bisequences of symbols from $\mathcal{A}$. If $\mathcal{A}$ has n elements,

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

The *one-sided full $\mathcal{A}$-shift* is the collection of all infinite sequences of symbols from $\mathcal{A}$ and is denoted

$$\Sigma^+(\mathcal{A}) = \Sigma_n^+ = \mathcal{A}^{\mathbb{N}} = \{x = (x_i)_{i\in\mathbb{N}} \ | \ x_i \in \mathcal{A} \text{ for all } i \in \mathbb{N}\}.$$

We also define the *shift transformation* $\sigma : \Sigma(\mathcal{A}) \to \Sigma(\mathcal{A})$ and $\Sigma^+(\mathcal{A}) \to \Sigma^+(\mathcal{A})$ by

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

The pair (Σ_n, σ) is called the *n-shift dynamical system.*

We give $\mathcal{A}$ the discrete topology and $\Sigma(\mathcal{A})$ and $\Sigma^+(\mathcal{A})$ the product topology. Furthermore, the topologies on $\Sigma(\mathcal{A})$ and $\Sigma^+(\mathcal{A})$ are compatible with the metric $d(x,y) = 1/2^n$, where $n = \inf\{|k| \ | \ x_k \neq y_k\}$ [47]. Thus two elements of $\Sigma(\mathcal{A})$ 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,σ) (or (X^+,σ)), 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 $\mathcal{A}$ 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 ϵ, on a fixed finite alphabet $\mathcal{A}$. The set of all blocks on $\mathcal{A}$, including the empty block ϵ, is denoted by $\mathcal{A}^*$. Given a subshift (X,σ) 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 $\mathcal{A}$ be a finite alphabet. A map $\theta : \mathcal{A} \to \mathcal{A}^*$ is called a *substitution.* A substitution θ is extended to $\mathcal{A}^*$ and $\mathcal{A}^{\mathbb{N}}$ by $\theta(b_1 b_2 \dots) = \theta(b_1)\theta(b_2)\dots$. A substitution θ is called *primitive* if there is m such that for all $a \in \mathcal{A}$ the block $\theta^m(a)$ contains every element of $\mathcal{A}$.

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 σ is the usual shift. Then (X_u^+, σ) is the dynamical system associated with u. For any sequence u, denote by $\mathcal{L}(u)$ the family of all subblocks of u.

Exercise 1.1. 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 $\mathcal{V}_n, n = 0, 1, 2, \dots$; $\mathcal{V}_0$ consists of a single vertex v_0, called the *root*; and the set $\mathcal{E}$ of edges is also the disjoint union of sets $\mathcal{E}_n, n = 1, 2, \dots$ 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 $x_{\max}$ and exactly one minimal path $x_{\min}$, then one may define $Tx_{\max} = x_{\min}$ and arrive at a minimal homeomorphism $T : X \to X$. See [14, 25] for surveys on Bratteli-Vershik 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 σ-algebra of X, T is a minimal homeomorphism (every orbit is dense), and μ is a unique T-invariant Borel probability measure on X.

2. The Krieger Embedding 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.

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 s 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." [43]

4. Lind and Thouvenot [46] proved that hyperbolic toral automorphisms (the matrix has no eigenvalue of modulus 1) are universal. This was extended by Quas and Soo [57] 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 [58].

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 [33, 34].

6. Every ergodic measure-preserving system is measure-theoretically isomorphic to a minimal Bratteli-Vershik system with a unique invariant Borel probability measure [66, 67].

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 [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 [17]: 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, ϵ-separated set. Then

$$h(T) = \lim_{\epsilon\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \log S(n, \epsilon).$$

Third definition [17]: For $n \in \mathbb{N}$ and $\epsilon > 0$, a subset $A \subset X$ is called n, ϵ-*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, ϵ-spanning set. Then

$$h(T) = \lim_{\epsilon\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \log R(n, \epsilon).$$

Exercise 2.1. 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 ("Variational Principle")**.** *$h(T) = \sup\{h_\mu(T) : \mu$ is an invariant (ergodic) Borel probability measure on $X\}$.*

2.2. **Ergodic-theoretic entropy.** A finite (or sometimes countable) measurable partition

$$\alpha = \{A_1, \ldots, A_r\}$$

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

$$H(\alpha) = \sum_{A\in\alpha} -\mu(A) \log \mu(A) \quad \text{(the logs can be base } e, 2, \text{ or } r);$$

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 α 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 α (or reading a symbol) at time 1, and $\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha$ represents the result of n repetitions of the experiment (or the reception of a string of n symbols). Then $H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha)/n$ is the average information gain per repetition (or per symbol received), and

$$h(\alpha, T) = \lim_{n\to\infty} \frac{1}{n} H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha)$$

is the long-term time average of the information gained per unit time. (This limit exists because of the *subadditivity* of H: $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.)

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 (Kolmogorov-Sinai). *If T has a finite* generator *α—a partition α such that the smallest σ-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 2.3. *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 $k \to \infty$.*

2.3. **Conditioning.** For a countable measurable partition α and sub-σ-algebra $\mathcal{F}$ of $\mathcal{B}$, we define the *conditional information function* of α 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 α (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 α 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 α. When $\mathcal{F}$ is the σ-algebra generated by a partition β, we often just write β in place of $\mathcal{F}$.

Proposition 2.4. *1. $H(\alpha \vee \beta|\mathcal{F}) = H(\alpha|\mathcal{F}) + H(\beta|\mathcal{B}(\alpha) \vee \mathcal{F})$.*
2. $H(\alpha|\mathcal{F})$ is increasing in its first variable and decreasing in its second.

Theorem 2.5. *For any finite (or countable finite-entropy) partition α,*

$$h(\alpha, T) = H(\alpha|\mathcal{B}(T^{-1}\alpha \vee T^{-2}\alpha \vee \ldots)).$$

2.4. **Examples.**

(1) Bernoulli shifts: $h = -\sum p_i \log p_i$. Consequently $\mathcal{B}(1/2, 1/2)$ is not isomorphic to $\mathcal{B}(1/3, 1/3, 1/3)$.
(2) Markov shifts: $h = -\sum p_i \sum P_{ij} \log P_{ij}$.
(3) Discrete spectrum: $h = 0$. (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 $h(\alpha, T) = H(\alpha|\alpha_1^\infty) = H(\alpha|\mathcal{B}) = 0$. It's especially easy to see for an irrational rotation of the circle, for if α is the partition into two disjoint arcs, then α_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 [52, p. 56].
(7) Skew products: $h(T \times \{S_x\}) = h(T) + h_T(S)$. Here the action is $(x, y) \to (Tx, S_x y)$, with each S_x a m.p.t. on Y, and the second term is the *fiber entropy*

$$h_T(S) = \sup\{\int_X H(\beta|S_x^{-1}\beta \vee S_x^{-1}S_{Tx}^{-1}\beta \vee \ldots) d\mu(x) : \beta \text{ is a finite partition of } Y\}.$$

(8) Automorphism of the torus: $h = \sum_{|\lambda_i|>1} \log|\lambda_i|$ (the λ_i are the eigenvalues of the integer matrix with determinant ± 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. □

Exercise 2.2. Prove the statements in 1–5 above.

2.5. **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) and open cover $\mathcal{U} = \{U_0, \ldots, U_{r-1}\}$ of X, for $x \in X$ and $n \geq 1$, we consider the set of codings of the initial n points in the orbit of x according to the partition $\mathcal{U}$: let $\mathcal{C}(x, n) =$ the set of n-blocks w on $\{0, \ldots, r-1\}$ such that $T^j x \in U_{w_j}$, $j = 1, \ldots, n$. Then we define the upper and lower *complexity of the orbit* of a point $x \in X$ to be

$$\sup K(x, T) = \sup_{\mathcal{U}} \limsup_{n\to\infty} \min\{\frac{K(w)}{n} : w \in \mathcal{C}(x, n)\}$$

and

$$\inf K(x, T) = \sup_{\mathcal{U}} \liminf_{n\to\infty} \min\{\frac{K(w)}{n} : w \in \mathcal{C}(x, n)\}$$

Theorem 2.6 (Brudno, White)**.** *If μ is an ergodic invariant measure on (X, T), then*

$$\sup K(x, T) = \inf K(x, T) = h_\mu(X, T) \text{ a.e. } d\mu(x).$$

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 $\mathcal{A}$. 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,

$$\limsup_{n\to\infty} \frac{\log p_x(n)}{n}, \tag{3.1}$$

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 [3, 11, 30, 31, 48]. Here we look at a few representative examples.

Exercise 3.1. [29] Show that the growth rate of the complexity function p_X of a subshift (X, σ) is an invariant of topological conjugacy by proving that if (X, σ) and (Y, σ) are topologically conjugate subshifts on finite alphabets, then there is a constant c such that

$$p_X(n-c) \le p_Y(n) \le p_X(n+c) \text{ for all } n > c. \tag{3.2}$$

3.1. **The complexity function in one-dimensional symbolic dynamics.** Let u be a one or two-sided infinite sequence on a finite alphabet $\mathcal{A}$, and let $p_u(n)$ denote the number of n-blocks in u. Since every block continues in at least one direction, $p_u(n+1) \ge p_u(n)$ for all n.

Exercise 3.2. Find the complexity functions of the bisequence $u = \dots 12121212321212121 \dots$ and the one-sided sequence $v = 321212121 \dots$.

Hedlund and Morse [49, 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 $p_u(n) \le n$.

Exercise 3.3. Show that for one-sided sequences u the following conditions are equivalent: (1) there is an n such that $p_u(n) \le n$;
(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 3.4. Show that for a two-sided sequence u, if there is an n such that $p_u(n) \le n$, then u is periodic.

3.2. **Sturmian sequences.** Hedlund and Morse [50] defined Sturmian sequences as those that have the smallest possible complexity among non-eventually-periodic sequences.

Definition 3.1. 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 3.5. The Fibonacci substitution is defined by:

$$\begin{aligned} \phi : 0 &\mapsto 01 \\ 1 &\mapsto 0. \end{aligned}$$

The fixed point of the Fibonacci substitution, $f = 0100101001001010010100100101...$, is called the Fibonacci sequence. Show that f is a Sturmian sequence.

Definition 3.2. A set S of blocks is *balanced* if for any pair of blocks u, v of the same length in S, $||u|_1 - |v|_1| \le 1$, where $|u|_1$ is the number of occurrences of 1 in u and $|v|_1$ is the number of occurences 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 $p_u(n) \ge n + 1$ for all n, and if u is balanced then $p_u(n) \le n + 1$ for all n. In fact, it can be proved that a sequence u is balanced and aperiodic if and only if it is Sturmian [48]. 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 3.3. Let α and β be real numbers with $0 \leq \alpha, \beta \leq 1$. We define two infinite sequences $x_{\alpha,\beta}$ and ${x'}_{\alpha,\beta}$ by

$$\begin{aligned} (x_{\alpha,\beta})_n &= \lfloor \alpha(n+1)+\beta \rfloor - \lfloor \alpha n+\beta \rfloor \\ ({x'}_{\alpha,\beta})_n &= \lceil \alpha(n+1)+\beta \rceil - \lceil \alpha n+\beta \rceil \end{aligned}$$

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 α and intercept β.

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 $F_n = (n, \lfloor \alpha n + \beta \rfloor)$. 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 $F'_n = (n, \lceil \alpha n + \beta \rceil)$. 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 [48].

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 [48] 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 3.4. 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:

$$\begin{aligned} (x_{\alpha,\beta})_0 =& \lfloor 1/\tau^2 \rfloor = 0 \\ (x_{\alpha,\beta})_1 =& \lfloor 2/\tau^2 \rfloor - \lfloor 1/\tau^2 \rfloor = 0 \\ (x_{\alpha,\beta})_2 =& \lfloor 3/\tau^2 \rfloor - \lfloor 2/\tau^2 \rfloor = 1 \\ (x_{\alpha,\beta})_3 =& \lfloor 4/\tau^2 \rfloor - \lfloor 3/\tau^2 \rfloor = 0 \\ (x_{\alpha,\beta})_4 =& \lfloor 5/\tau^2 \rfloor - \lfloor 4/\tau^2 \rfloor = 0 \\ (x_{\alpha,\beta})_5 =& \lfloor 6/\tau^2 \rfloor - \lfloor 5/\tau^2 \rfloor = 1 \\ & \vdots \end{aligned}$$

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 3.6. 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 $\mathcal{L}_n(u)$ 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 3.5. 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^+ [31, 48].

In a similar fashion, we define the right special blocks of $\mathcal{L}_n(u)$.

Definition 3.6. 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 $L_n(u)$ [31].

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 α associated to u. We use this α to determine the left special sequence of X_u^+.

Let $(d_1, d_2, ..., d_n, ...)$ 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 $(d_1, d_2, ..., d_n, ...)$ is its *directive sequence*. We can then determine the left special sequence of X_u^+ with the following proposition stated in [48].

Proposition 3.7. *Let $\alpha = [0, 1 + d_1, d_2, ...]$ be the continued fraction expansion of an irrational α 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 3.8. Let $\alpha = 1/\tau^2$, where $\tau = (1 + \sqrt{5})/2$ is the golden mean. The continued fraction expansion of $1/\tau^2$ is $[0, 2, \overline{1}]$. 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:

$$\begin{aligned} s_1 &= s_0^{d_1} s_{-1} = 01 \\ s_2 &= s_1^{d_2} s_0 = 010 \\ s_3 &= s_2^{d_3} s_1 = 01001 \\ s_4 &= s_3^{d_4} s_2 = 01001010 \\ &\vdots \end{aligned}$$

Continuing this process, the left special sequence of X_u^+, where u is a coding of a line with slope $1/\tau^2$, is

$$l = 010010100100101001... = f.$$

It follows that the left special sequence of X_f^+ is f.

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

$$\omega = 0110100110010110....$$

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 [31, Chapter 5]

3.4. **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 3.3) 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 [27]): For a d-dimensional configuration $x : \mathbb{Z}^d \to \mathcal{A}$ on a finite alphabet $\mathcal{A}$, define its *rectangular complexity function* to be the function $P_x(m_1, \dots, m_d)$ 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 vector $w \in \mathbb{Z}^2$ such that $x(v + w) = x(v)$ for all $v \in \mathbb{Z}^2$.

Cassaigne [21] characterized all two-dimensional configurations with complexity function $P_x(m_1, m_2) = m_1 m_2 + 1$.

Vuillon [68] considered tilings of the plane generated by a cut-and-project scheme. Recall (see [4]) 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 [13] showed that these two-dimensional configurations code the $\mathbb{Z}^2$ action of two translations on the 2-torus. 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 [12].

Sander and Tijdeman [61–63] 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 $A = \{a_1, \dots, a_n\}$, each $a_i \in \mathbb{Z}^d$, be a fixed non-empty *sampling window*, and define

$$P_x(A) = \operatorname{card}\{(x(v + a_1), \dots, x(v + a_n)) : v \in \mathbb{Z}^d\} \tag{3.3}$$

to be the number of distinct A-*patterns* in x (written here as ordered $|A|$-tuples). A natural extension of (3.3) 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 $A \subset \mathbb{Z}$ 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 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.

The last statement above was recently improved by Cyr and Kra [23]: 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 [42] (see also [41]) represented configurations in $\mathbb{Z}^d$ as formal power series in d variables with coefficients from $\mathcal{A}$ 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 [27] 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 [59] by combinatorial and geometrical arguments to replace $1/144$ by $1/16$, and by Cyr and Kra [24] 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 [32, 60, 64] 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 3.1 for subshifts, Frank and Sadun [56] and A. Julien [39] (see also [40]) 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 [5, 55]) and even on trees [6–8].

Analogues of the Nivat Conjecture for general Delaunay sets in $\mathbb{R}^d$ were proved by Lagarias and Pleasants [44, 45]. Huck and Richard [37] 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 [26] 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.

3.5. **Topological complexity.** Let (X,T) be a topological dynamical system. If (X,T) is a subshift and $\mathcal{U}$ 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 $\mathcal{U}_0^{n-1} = \mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-n+1}\mathcal{U}$; i.e., in this case $p_X(n)$ equals the $N(\mathcal{U}_0^{n-1})$ of the definition of topological entropy (see Sections 3.1 and 2.1). Blanchard, Host, and Maass [15] took this as the definition of the *topological complexity function*: $p_{\mathcal{U}}(n) = N(\mathcal{U}_0^{n-1})$ = the minimum possible cardinality of a subcover of $\mathcal{U}_0^{n-1}$.

Theorem 3.9. [15] *A topological dynamical system is equicontinuous if and only if every finite open cover has bounded complexity function. (Cf. Sections 3.1 and* **??***.)*

Exercise 3.7. 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 3.10. 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 [15].

(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 $\mathcal{U}$ of X has complexity function satisfying $p_{\mathcal{U}}(n) \geq n+2$ 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 $\mathcal{U}$ of X with $p_{\mathcal{U}}(n) \leq n+1$ 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 $X \times Y$.)

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 [16] 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 [65] 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 [54, 70].

One considers a model that consists of a family $X = \{X_i : i = 0, 1, \dots, n-1\}$ 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 $n^* = \{0, 1, \dots, n-1\}$. 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

$$\begin{aligned} MI(X,Y) &= H(X) + H(Y) - H(X,Y). \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned}$$

Here (X,Y) is the random variable on Ω 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 $X = \{X_i : i = 0, 1, \dots, n-1\}$ 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}:

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

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 [18] 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_S^n, to be a family of numbers satisfying for all $n \in \mathbb{N}$ and $S \subset n^*$

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

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_S^n, the corresponding *mutual information functional*, $\mathcal{I}^c(X)$, is defined by

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

Definition 4.1. 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 4.2 (Buzzi-Zambotti)**.** *Let c_S^n 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 λ_c on $[0,1]$ such that*

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

Example 4.3. 1. $c_S^n = \dfrac{1}{(n+1)} \dfrac{1}{\binom{n}{|S|}}$ (Edelman-Sporns-Tononi);

2. For $0 < p < 1$,

$$c_S^n = \frac{1}{2}(p^{|S|}(1-p)^{n-|S|} + (1-p)^{|S|}p^{n-|S|}) \text{ (p-symmetric);}$$

3. For $p = 1/2$, $c_S^n = 2^{-n}$ (uniform).

Exercise 4.1. Prove that the neural (Edelman-Sporns-Tononi) weights correspond to λ being Lebesgue measure on $[0,1]$.

4.2. **Topological intricacy and average sample complexity.** Let (X,T) be a topological dynamical system and $\mathscr{U}$ an open cover of X. Given $n \in \mathbb{N}$ and a subset $S \subset n^*$ define

$$\mathscr{U}_S = \bigvee_{i \in S} T^{-i} \mathscr{U}.$$

Definition 4.4. [54, 70] Let c_S^n be a system of coefficients. Define the *topological intricacy of* (X,T) *with respect to the open cover* $\mathscr{U}$ to be

$$\operatorname{Int}(X,\mathscr{U},T) := \lim_{n\to\infty} \frac{1}{n} \sum_{S\subset n^*} c_S^n \log\left(\frac{N(\mathscr{U}_S)N(\mathscr{U}_{S^c})}{N(\mathscr{U}_{n^*})}\right).$$

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 4.5. [54, 70] The *topological average sample complexity of* T *with respect to the open cover* $\mathscr{U}$ is defined to be

$$\operatorname{Asc}(X,\mathscr{U},T) := \lim_{n\to\infty} \frac{1}{n} \sum_{S\subset n^*} c_S^n \log N(\mathscr{U}_S).$$

Proposition 4.6. $\operatorname{Int}(X,\mathscr{U},T) = 2\operatorname{Asc}(X,\mathscr{U},T) - h_{\text{top}}(X,\mathscr{U},T)$.

Theorem 4.7. *The limits in the definitions of* $\operatorname{Int}(X,\mathscr{U},T)$ *and* $\operatorname{Asc}(X,\mathscr{U},T)$ *exist .*

As usual this follows from subadditivity of the sequence

$$b_n := \sum_{S\subset n^*} c_S^n \log N(\mathscr{U}_S)$$

and Fekete's Subadditive Lemma: For every subadditive sequence a_n, the limit $\lim_{n\to\infty} a_n/n$ exists and is equal to $\inf_n a_n/n$.

Exercise 4.2. Prove Fekete's Lemma.

Proposition 4.8. *For each open cover* $\mathscr{U}$*,* $\operatorname{Asc}(X,\mathscr{U},T) \le h_{\text{top}}(X,\mathscr{U},T) \le h_{\text{top}}(X,T)$*, and hence* $\operatorname{Int}(X,\mathscr{U},T) \le h_{\text{top}}(X,\mathscr{U},T) \le h_{\text{top}}(X,T)$.

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,σ) with respect to the "time zero open cover" $\mathscr{U}_0$ 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:

$$\operatorname{Int}(X,\mathscr{U}_0,\sigma) = \lim_{n\to\infty} \frac{1}{n} \sum_{S\subset n^*} c_S^n \log\left(\frac{|\mathscr{L}_S(X)||\mathscr{L}_{S^c}(X)|}{|\mathscr{L}_{n^*}(X)|}\right)$$

Example 4.9 (Computing $|\mathscr{L}_S(X)|$ 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 $S = \{0,2\}$, $N(S) = 4$.

$S = \{0,1\}$

0	0	
0	1	
1	0	

$|\mathscr{L}_S(X)| = 3$

$S = \{0,2\}$

0		0
0		1
1		0
1		1

$|\mathscr{L}_S(X)| = 4$

When we average over all subsets $S \subset n^*$, we get an approximation (from above) to $\operatorname{Int}(X,\mathscr{U}_0,\sigma)$:

S	S	$\lvert\mathscr{L}_S(X)\rvert$	$\lvert\mathscr{L}_{S^c}(X)\rvert$
$\emptyset$	$\{0,1,2\}$	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 4.10 (Computing $|\mathscr{L}_S(X)|$ for the golden mean SFT)**.**

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

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

$$\operatorname{Asc}(X,\mathscr{U}_0,\sigma)=\frac{1}{4}\sum_{k=1}^{\infty}\frac{\log|\mathscr{L}_{k^*}(X)|}{2^k}.$$

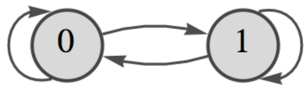
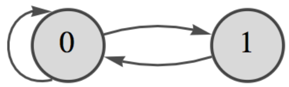
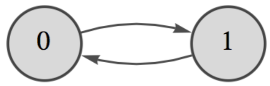
This formula shows that, as expected, Asc is sensitive to word counts of all lengths and thus is a finer measurement than h_{top}, which just gives the asymptotic exponential growth rate. Below we will see examples of systems that can be distinguished by Asc and Int but not by their entropies, or even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

The proof is too long to be sketched here, but a main idea is that most subsets $S\subset n^*$ are also subsets of $(n-1)^*$.

Corollary 4.12. *For the full r-shift with $c_S^n=2^{-n}$ for all S,*

$$\operatorname{Asc}(\Sigma_r,\mathscr{U}_0,\sigma)=\frac{\log r}{2}\quad \textit{and}\quad \operatorname{Int}(\Sigma_r,\mathscr{U}_0,\sigma)=0.$$

In the following table we compare h_{top}, 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.

	Adjacency Graph	Entropy	Asc	Int
Disordered		0.693	0.347	0
		0.481	0.286	0.090
Ordered		0	0	0

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.

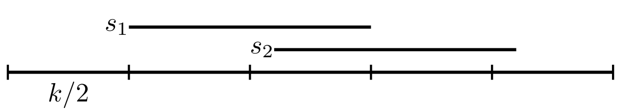
Theorem 4.13. *Let (X,T) be a topological dynamical system and fix the system of coefficients to be $c_S^n = 2^{-n}$. Then*

$$\sup_{\mathscr{U}} \operatorname{Asc}(X,\mathscr{U},T) = h_{\mathrm{top}}(X,T).$$

Proof. The proof depends on the structure of average subsets of $n^* = \{0,1,\dots,n-1\}$: 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) α, then iterate and refine, replacing α 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 α_{n^*}; and when k is fixed, as n grows the result is the same.

When computing Asc and 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 α_{S+k^*} as compared to α_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 $N((\mathscr{U}_k)_S)$ 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 $|I| \geq k/2$,

$$0 \leq \frac{\log N(I)}{\operatorname{card}(I)} - h_{\mathrm{top}}(X,\sigma) < \epsilon. \tag{4.1}$$

We let $\mathfrak{B}$ denote the set of $S \subset n^*$ which miss at least $2n\epsilon/k$ of the intervals of length $k/2$ and show that

$$\lim_{n\to\infty} \frac{\operatorname{card}(\mathfrak{B})}{2^n} = 0.$$

If $S \notin \mathfrak{B}$, 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 4.1 comparing (average of log of) number of blocks to $h_{\text{top}}(X, \sigma)$. □

Exercise 4.3. Prove that $\sup_{\mathscr{U}} \text{Int}(X, \mathscr{U}, T) = h_{\text{top}}(X, T)$ (for the system of coefficients $c_S^n = 2^{-n}$).

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 α of X and a subset $S \subset n^*$ define

$$\alpha_S = \bigvee_{i \in S} T^{-i}\alpha.$$

Definition 4.14 (P-W)**.** Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system, $\alpha = \{A_1, \dots, A_n\}$ a finite measurable partition of X, and c_S^n a system of coefficients. The *measure-theoretic intricacy of T with respect to the partition α* is

$$\text{Int}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \left[H_\mu(\alpha_S) + H_\mu(\alpha_{S^c}) - H_\mu(\alpha_{n^*}) \right].$$

The *measure-theoretic average sample complexity of T with respect to the partition α* is

$$\text{Asc}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n H_\mu(\alpha_S).$$

Theorem 4.15. *The limits in the definitions of measure-theoretic intricacy and measure-theoretic average sample complexity exist.*

Theorem 4.16. *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system and fix the system of coefficients $c_S^n = 2^{-n}$. 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 4.17 (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 the processes considered are finite-state ergodic stochastic processes $X = (x_1, x_2, \dots)$, and a "finitely observable functional" is the a.s. limit $F(X)$ of a sequence of functions $f_n(x_1, x_2, \dots, x_n)$ 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 $\mathscr{U}$, fix a k and find the topological average sample complexity $Asc(X, \mathscr{U}_k, \sigma) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \log N((\mathscr{U}_k)_S)$. Or, do not take the limit on n, and study the quantity as a function of n, analogously to the symbolic or topological complexity functions. Similarly for the measure-theoretic version: fix a partition α and study the limit, or the function of n.

$$\text{Asc}_\mu(X, T, \alpha) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n H_\mu(\alpha_S).$$

So we begin study of the Asc of a fixed open cover as a function of n,

$$\text{Asc}(X, \sigma, \mathscr{U}_k, n) = \frac{1}{n} \sum_{S \subset n^*} c_S^n \log N(S),$$

especially for SFT's and $\mathscr{U} = \mathscr{U}_0$, the natural time-0 cover (and partition).

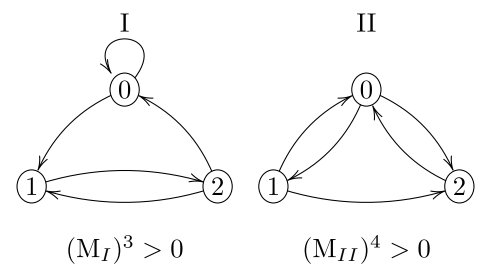


FIGURE 1. Graphs of two subshifts with the same complexity function but different average sample complexity functions.

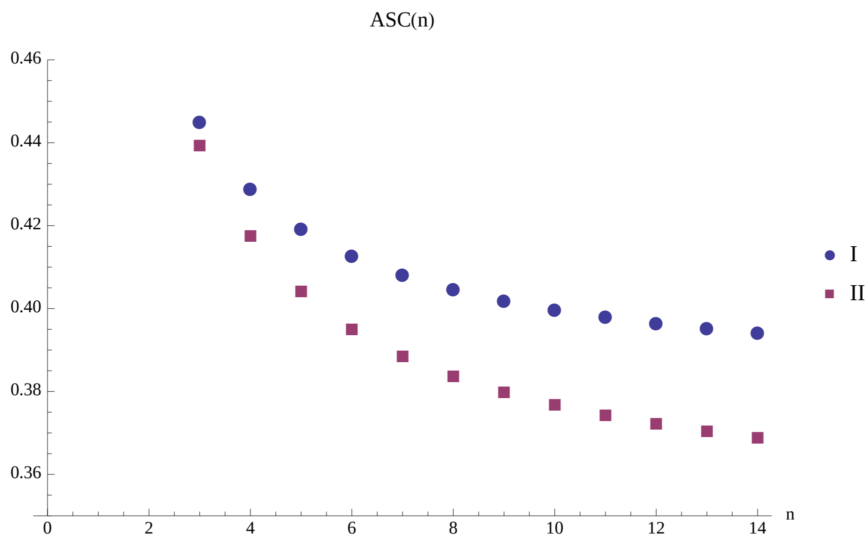
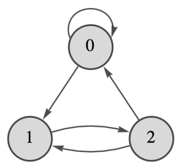
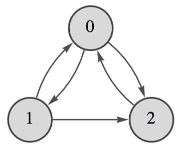


Figure 4.18 shows two SFT's that have the same number of n-blocks for every n but different Asc functions.

Example 4.18.

$$\mathrm{Asc}(n) = \frac{1}{n}\frac{1}{2^n}\sum_{S \subset n^*} \log N(S)$$

Numerical evidence (up to $n = 10$) indicates that these two SFT's have different values of Asc and Int, although they have identical complxity functions and hence the same topological entropy..

Adjacency Graph	h_{top}	Asc(10)	Int(10)
	0.481	0.399	0.254
	0.481	0.377	0.208

4.5. **Computing measure-theoretic average sample complexity.** For a fixed partition α, we develop a relationship between $\operatorname{Asc}_\mu(X, \alpha, T)$ and a series summed over i involving the conditional entropies $H_\mu(\alpha \mid \alpha_i^\infty)$. The series can serve as a computational tool analogous to the series in Theorem 4.11.

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 $\mathscr{B}(1/2, 1/2)$ on the full 2-shift. For example $\{0, 2, 3\} \subset 5^* \leftrightarrow 10110$. The average entropy, $H_\mu(\alpha_S)$, 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, Σ_2.

Theorem 4.19. *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system and α 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 $c_S^n = 2^{-n}$ for all $S \subset n^*$. Then*

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

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 4.20. *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system and α a finite measurable partition of X. Let $c_S^n = 2^{-n}$ for all $S \subset n^*$. Then*

$$\operatorname{Asc}_\mu(X, \alpha, T) \geq \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} H_\mu\left(\alpha \mid \alpha_i^\infty\right).$$

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 4.20 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 Int, there is the added problem of the minus sign in

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

Maybe some modern work on local or relative variational principles, almost subadditive potentials, equilibrium states for shifts with infinite alphabets, etc. could apply? (See [9,10,19,22,28,35,36,38,51,71] etc.)

But the above theorem does give up some information immediately:

Proposition 4.21. *When $T : X \to X$ is an expansive homeomorphism on a compact metric space (e.g., (X,T) is a subshift on finite alphabet), $\mathrm{Asc}_\mu(X,T,\alpha)$ is an affine upper semicontinuous (in the weak* topology) function of μ, so the set of maximal measures for $\mathrm{Asc}_\mu(X,T,\alpha)$ is nonempty, compact, and convex and contains some ergodic measures (see [69, p. 198 ff.]).*

We try now to find measures of maximal Asc or 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 $\mu_{\max}$. Given a potential function ϕ that is a function of just two coordinates, again there is a unique measure that maximizes

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

See [52].

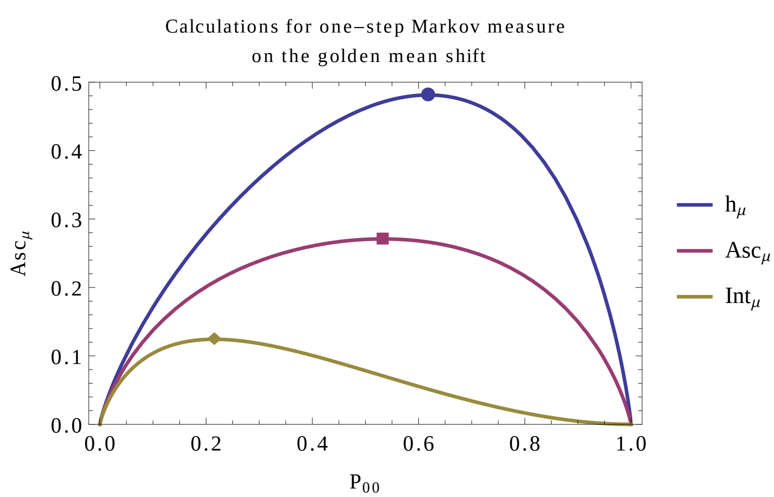
A 1-step Markov measure on the full shift space (Σ_n, σ) is given by s stochastic matrix $P = (P_{ij})$ and fixed probability vector $p = \begin{pmatrix} p_0 & p_1 & \dots & p_{n-1} \end{pmatrix}$, i.e. $\sum p_i = 1$ and $pP = p$. The measure $\mu_{P,p}$ is defined as usual on cylinder sets by

$$\mu_{p,P}[i_0 i_1 \dots i_k] = p_{i_0} P_{i_0 i_1} \cdots P_{i_{k-1} i_k}.$$

Example 4.22 (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

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

Using the series formula in Theorem 4.20 and known equations for conditional entropy, we can approximate Asc_μ and Int_μ for Markov measures on SFTs. Let's look first at 1-step Markov measures.

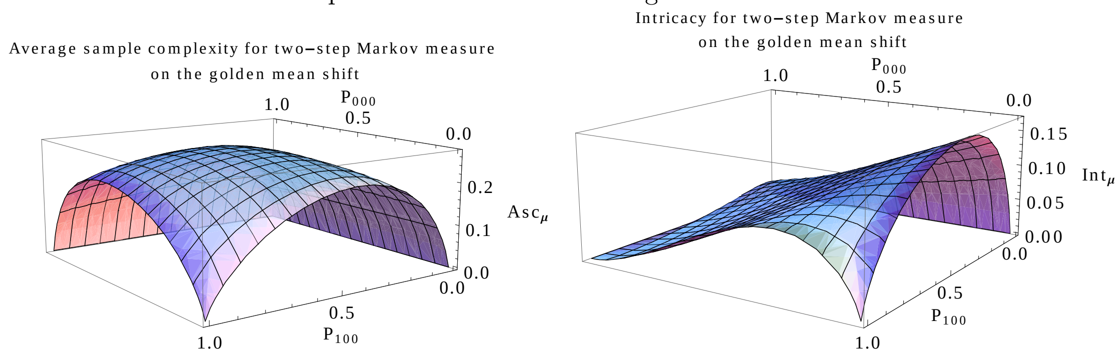


P_{00}	h_μ	Asc_μ	Int_μ
0.618	0.481	0.266	0.051
0.533	0.471	0.271	0.071
0.216	0.292	0.208	0.124

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 Asc_μ and Int_μ; and the maxima for Asc_μ, Int_μ, and h_μ are achieved by different measures.

Now let's calculate Asc and Int for various 2-step Markov measures on the golden mean SFT.

2-step Markov measures on the golden mean shift

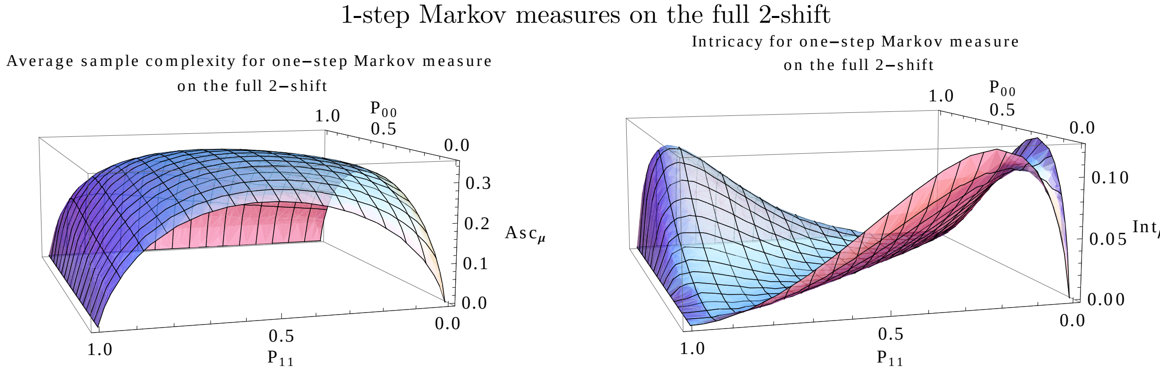


P_{000}	P_{100}	h_μ	Asc_μ	Int_μ
0.618	0.618	0.481	0.266	0.051
0.483	0.569	0.466	0.272	0.078
0	0.275	0.344	0.221	0.167

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

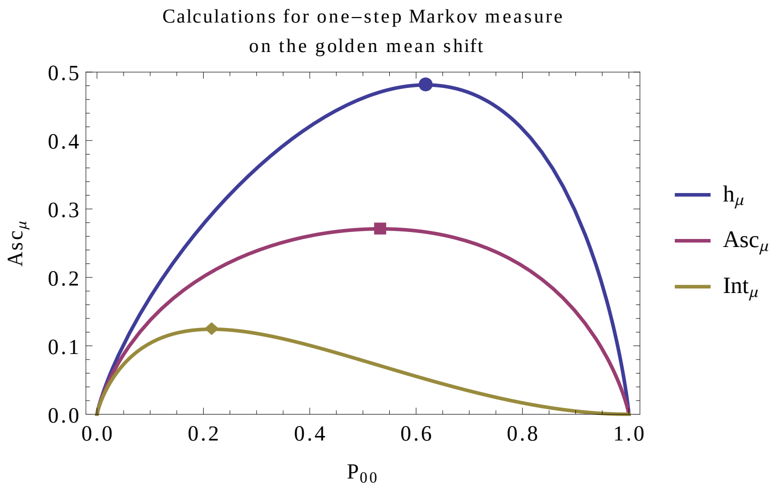
Let's move from the golden mean SFT to the full 2-shift.

1-step Markov measures on the full 2-shift



P_{00}	P_{11}	h_μ	Asc_μ	Int_μ
0.5	0.5	0.693	0.347	0
0.216	0	0.292	0.208	0.124
0	0.216	0.292	0.208	0.124
0.905	0.905	0.315	0.209	0.104

Asc_μ appears to be strictly convex, so it would have a unique maximum among 1-step Markov measures. Int_μ 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



P_{00}	h_μ	Asc_μ	Int_μ
0.618	0.481	0.266	0.051
0.533	0.471	0.271	0.071
0.216	0.292	0.208	0.124

TABLE 1. Calculations for one-step Markov measures on the golden mean shift. Numbers in bold are maxima for the given categories.

P_{000}	P_{100}	h_μ	Asc_μ	Int_μ
0.618	0.618	0.481	0.266	0.051
0.483	0.569	0.466	0.272	0.078
0	0.275	0.344	0.221	0.167

TABLE 2. Calculations for two-step Markov measures on the golden mean shift.

Int_μ among all 1-step Markov measures. The maxima for Asc_μ, Int_μ, and h_μ are achieved by different measures.

We summarize some of the questions generated above.

Conjecture 4.23. On the golden mean SFT, for each r there is a unique r-step Markov measure μ_r that maximizes $\mathrm{Asc}_\mu(X, \sigma, \alpha)$ among all r-step Markov measures.

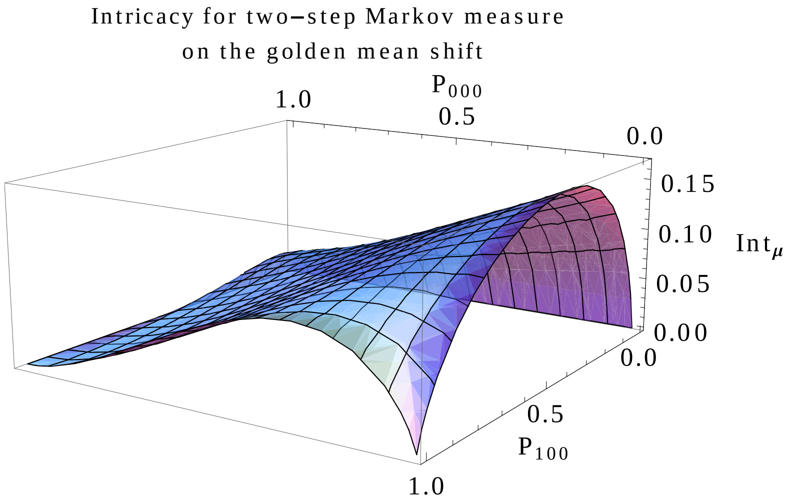
Conjecture 4.24. $\mu_2 \neq \mu_1$

Conjecture 4.25. On the golden mean SFT there is a unique measure that maximizes $\mathrm{Asc}_\mu(X, T, \alpha)$. It is not Markov of any order (and of course is not the same as $\mu_{\max}$).

Conjecture 4.26. On the golden mean SFT for each r there is a unique r-step Markov measure that maximizes $\mathrm{Int}_\mu(X, T, \alpha)$ among all r-step Markov measures.

Conjecture 4.27. On the 2-shift there are *two* 1-step Markov measures that maximize $\mathrm{Int}_\mu(X, T, \alpha)$ among all 1-step Markov measures. They are supported on the golden mean SFT and its image under the dualizing map $0 \leftrightarrow 1$.

Conjecture 4.28. On the 2-shift there is a 1-step Markov measure that is *fully supported* and is a local maximum point for $\mathrm{Int}_\mu(X, T, \alpha)$ among all 1-step Markov measures.



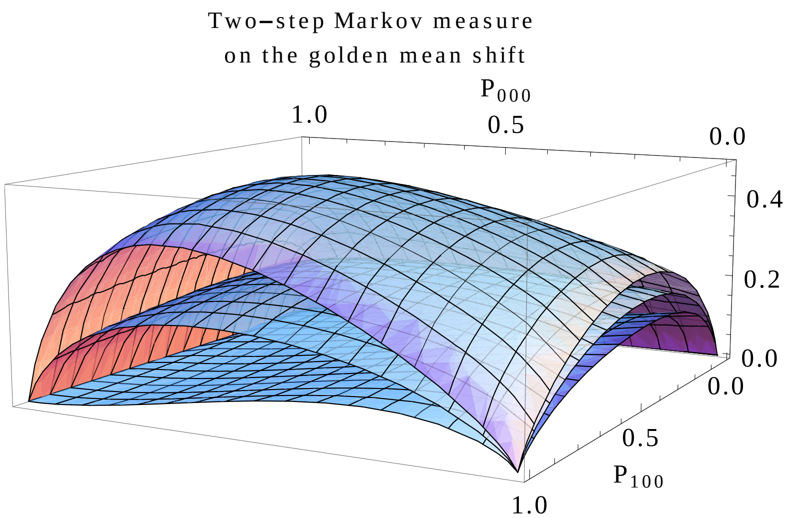
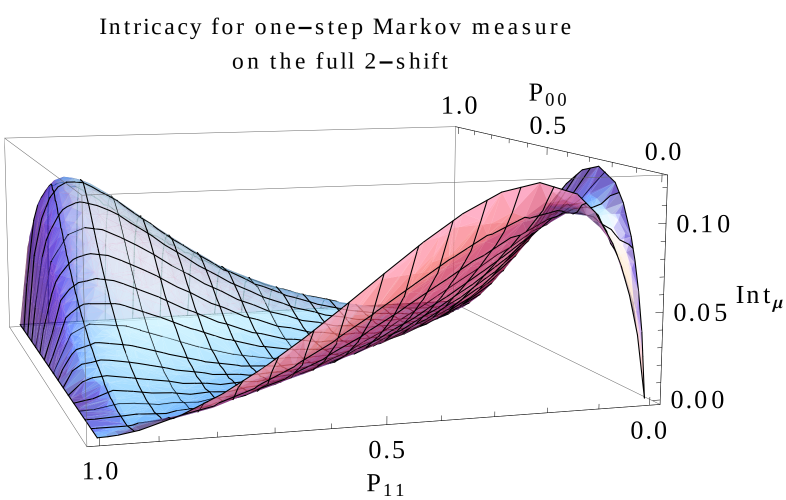
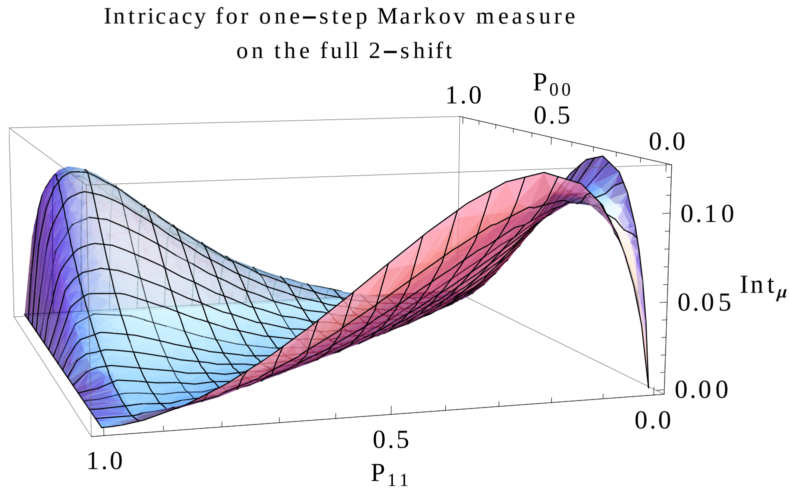


FIGURE 2. Combination of the plots of h_μ, Asc_μ, and Int_μ for two-step Markov measures on the golden mean shift.





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 [70]:

1. We do not know whether a variational principle $\sup_\mu \mathrm{Asc}_\mu(X,T,\alpha) = \mathrm{Asc}_{\mathrm{top}}(X,\alpha,T)$ 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. 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 Asc and Int for more subshifts and systems.
6. Find the subsets or patterns S that maximize $\log N(\mathscr{U}_S)$ or $\log[N(\mathscr{U}_S)N(\mathscr{U}_{S^c})]/N(\mathscr{U}_{n^*})$, 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 Alt and 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!

References

1. Daniel M. Abrams and Mark J. Panaggion, *A model balancing cooperation and competition explains our right-handed world and the dominance of left-handed athletes*, arXiv:1109.6061v1.
2. R. L. Adler, A. G. Konheim, and M. H. McAndrew, *Topological entropy*, Trans. Amer. Math. Soc. **114** (1965), 309–319. MR 0175106 (30 #5291)
3. Jean-Paul Allouche, *Sur la complexité des suites infinies*, Bull. Belg. Math. Soc. Simon Stevin **1** (1994), no. 2, 133–143, Journées Montoises (Mons, 1992). MR 1318964 (96d:68167)
4. Pierre Arnoux, Valérie Berthé, Hiromi Ei, and Shunji Ito, *Tilings, quasicrystals, discrete planes, generalized substitutions, and multidimensional continued fractions*, Discrete Models: Ccombinatorics, Computation, and Geometry (Paris, 2001), Discrete Math. Theor. Comput. Sci. Proc., AA, Maison Inform. Math. Discrèt. (MIMD), Paris, 2001, pp. 059–078 (electronic). MR 1888763 (2002k:37025)
5. Nathalie Aubrun, *Sofic subshifts on finitely generated groups*, Lectures at the Workshop on Symbolic Dynamics on Finitely Generated Groups.

6. Nathalie Aubrun and Marie-Pierre Béal, *Tree-shifts of finite type*, Theoret. Comput. Sci. **459** (2012), 16–25. MR 2974235
7. ______, *Sofic tree-shifts*, Theory Comput. Syst. **53** (2013), no. 4, 621–644. MR 3084356
8. Jung-Chao Ban and Chih-Hung Chang, *Tree-shifts: Irreducibility, mixing, and the chaos of tree-shifts*, arXiv:1509.01355v1.
9. Luis Barreira, *Almost additive thermodynamic formalism: Some recent developments*, Rev. Math. Phys. **22** (2010), no. 10, 1147–1179. MR 2740708 (2012a:37059)
10. Luis M. Barreira, *A non-additive thermodynamic formalism and applications to dimension theory of hyperbolic dynamical systems*, Ergodic Theory Dynam. Systems **16** (1996), no. 5, 871–927. MR 1417767 (98a:58124)
11. Valérie Berthé and Michel Rigo (eds.), *Combinatorics, Automata and Number Theory*, Encyclopedia of Mathematics and its Applications, vol. 135, Cambridge University Press, Cambridge, 2010. MR 2742574 (2011k:68006)
12. Valérie Berthé and Laurent Vuillon, *Suites doubles de basse complexité*, J. Théor. Nombres Bordeaux **12** (2000), no. 1, 179–208. MR 1827847 (2002d:68085)
13. ______, *Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences*, Discrete Math. **223** (2000), no. 1-3, 27–53. MR 1782038 (2001i:68139)
14. S. Bezuglyi and O. Karpel, *Bratteli diagrams: structure, measures, dynamics*, arXiv:1503.03360v1.
15. F. Blanchard, B. Host, and A. Maass, *Topological complexity*, Ergodic Theory Dynam. Systems **20** (2000), no. 3, 641–662. MR 1764920 (2002b:37019)
16. Ron Blei, *Measurements of interdependence*, Lith. Math. J. **51** (2011), no. 2, 141–154. MR 2805733 (2012g:46002)
17. Rufus Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Trans. Amer. Math. Soc. **153** (1971), 401–414. MR 0274707 (43 #469)
18. J. Buzzi and L. Zambotti, *Approximate maximizers of intricacy functionals*, Probab. Theory Related Fields **153** (2012), no. 3-4, 421–440. MR 2948682
19. Yong-Luo Cao, De-Jun Feng, and Wen Huang, *The thermodynamic formalism for sub-additive potentials*, Discrete Contin. Dyn. Syst. **20** (2008), no. 3, 639–657. MR 2373208 (2008k:37072)
20. Kathleen Carroll and Karl Petersen, *Markov diagrams for some non-Markovian systems*, arXiv:1504.03988v1.
21. Julien Cassaigne, *Subword complexity and periodicity in two or more dimensions*, Developments in language theory (Aachen, 1999), World Sci. Publ., River Edge, NJ, 2000, pp. 14–21. MR 1880477 (2002m:68092)
22. Wen-Chiao Cheng, Yun Zhao, and Yongluo Cao, *Pressures for asymptotically sub-additive potentials under a mistake function*, Discrete Contin. Dyn. Syst. **32** (2012), no. 2, 487–497. MR 2837070 (2012j:37046)
23. Van Cyr and Bryna Kra, *Complexity of short rectangles and periodicity*, arXiv:1307.0098v1.
24. ______, *Nonexpansive $\mathbb{Z}^2$-subdynamics and Nivat's conjecture*, Trans. Amer. Math. Soc. **367** (2015), no. 9, 6487–6537. MR 3356945
25. Fabien Durand, *Combinatorics on Bratteli diagrams and dynamical systems*, Combinatorics, Automata and Number Theory, Encyclopedia Math. Appl., vol. 135, Cambridge Univ. Press, Cambridge, 2010, pp. 324–372. MR 2759109 (2012h:37023)
26. Fabien Durand and Michel Rigo, *Multidimensional extension of the Morse-Hedlund theorem*, European J. Combin. **34** (2013), no. 2, 391–409. MR 2994406
27. Chiara Epifanio, Michel Koskas, and Filippo Mignosi, *On a conjecture on bidimensional words*, Theoret. Comput. Sci. **299** (2003), no. 1-3, 123–150. MR 1973149 (2004c:68116)
28. K. J. Falconer, *A subadditive thermodynamic formalism for mixing repellers*, J. Phys. A **21** (1988), no. 14, L737–L742. MR 968557 (89k:58154)
29. Sébastien Ferenczi, *Complexity of sequences and dynamical systems*, Discrete Math. **206** (1999), no. 1-3, 145–154, Combinatorics and number theory (Tiruchirappalli, 1996). MR 1665394 (2000f:68093)
30. ______, *Complexity of sequences and dynamical systems*, Discrete Math. **206** (1999), no. 1-3, 145–154, Combinatorics and Number Theory (Tiruchirappalli, 1996). MR 1665394 (2000f:68093)
31. N. P. Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, vol. 1794, Springer Science & Business Media, 2002, Edited by V. Berthé, S. Ferenczi, C. Mauduit, and A. Siegel.
32. Natalie Priebe Frank, *A primer of substitution tilings of the Euclidean plane*, Expo. Math. **26** (2008), no. 4, 295–326. MR 2462439 (2010d:52047)
33. Thierry Giordano, Ian F. Putnam, and Christian F. Skau, *Topological orbit equivalence and C^*-crossed products*, J. Reine Angew. Math. **469** (1995), 51–111. MR 1363826 (97g:46085)
34. Richard H. Herman, Ian F. Putnam, and Christian F. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Internat. J. Math. **3** (1992), no. 6, 827–864. MR 1194074 (94f:46096)
35. W. Huang, A. Maass, P. P. Romagnoli, and X. Ye, *Entropy pairs and a local Abramov formula for a measure theoretical entropy of open covers*, Ergodic Theory Dynam. Systems **24** (2004), no. 4, 1127–1153. MR 2085906 (2005e:37027)
36. Wen Huang, Xiangdong Ye, and Guohua Zhang, *A local variational principle for conditional entropy*, Ergodic Theory Dynam. Systems **26** (2006), no. 1, 219–245. MR 2201946 (2006j:37015)
37. Christian Huck and Christoph Richard, *On pattern entropy of weak model sets*, arXiv:1412.6307v3.
38. Godofredo Iommi and Yuki Yayama, *Almost-additive thermodynamic formalism for countable Markov shifts*, Nonlinearity **25** (2012), no. 1, 165–191. MR 2864381
39. Antoine Julien, *Complexity as a homeomorphism invariant for tiling spaces*, arXiv:1212.1320v1.
40. ______, *Complexity and cohomology for cut-and-projection tilings*, Ergodic Theory Dynam. Systems **30** (2010), no. 2, 489–523. MR 2599890 (2011i:52045)
41. Jarkko Kari and Michal Szabados, *An algebraic geometric approach to multidimensional words.*

42. ———, *An algebraic geometric approach to nivats conjecture*, Automata, Languages, and Programming (Magns M. Halldrsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, eds.), Lecture Notes in Computer Science, vol. 9135, Springer Berlin Heidelberg, 2015, pp. 273–285 (English).
43. Wolfgang Krieger, *On the subsystems of topological Markov chains*, Ergodic Theory Dynam. Systems **2** (1982), no. 2, 195–202 (1983). MR 693975 (85b:28020)
44. Jeffrey C. Lagarias and Peter A. B. Pleasants, *Local complexity of Delone sets and crystallinity*, Canad. Math. Bull. **45** (2002), no. 4, 634–652, Dedicated to Robert V. Moody. MR 1941231 (2003i:52023)
45. ———, *Repetitive Delone sets and quasicrystals*, Ergodic Theory Dynam. Systems **23** (2003), no. 3, 831–867. MR 1992666 (2005a:52018)
46. D. A. Lind and J.-P. Thouvenot, *Measure-preserving homeomorphisms of the torus represent all finite entropy ergodic transformations*, Math. Systems Theory **11** (1977/78), no. 3, 275–282. MR 0584588 (58 #28434)
47. Douglas Lind and Brian Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995. MR 1369092 (97a:58050)
48. M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002, A collective work by Jean Berstel, Dominique Perrin, Patrice Seebold, Julien Cassaigne, Aldo De Luca, Steffano Varricchio, Alain Lascoux, Bernard Leclerc, Jean-Yves Thibon, Veronique Bruyere, Christiane Frougny, Filippo Mignosi, Antonio Restivo, Christophe Reutenauer, Dominique Foata, Guo-Niu Han, Jacques Desarmenien, Volker Diekert, Tero Harju, Juhani Karhumaki and Wojciech Plandowski, With a preface by Berstel and Perrin. MR 1905123 (2003i:68115)
49. Marston Morse and Gustav A. Hedlund, *Symbolic Dynamics*, Amer. J. Math. **60** (1938), no. 4, 815–866. MR 1507944
50. ———, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42. MR 0000745 (1,123d)
51. Anna Mummert, *The thermodynamic formalism for almost-additive sequences*, Discrete Contin. Dyn. Syst. **16** (2006), no. 2, 435–454. MR 2226491 (2007f:37045)
52. William Parry and Selim Tuncel, *Classification Problems in Ergodic Theory*, London Mathematical Society Lecture Note Series, vol. 67, Cambridge University Press, Cambridge-New York, 1982, Statistics: Textbooks and Monographs, 41. MR 666871 (84g:28024)
53. Karl Petersen, *Ergodic Theory*, Cambridge University Press, Cambridge New York, 1989.
54. Karl Petersen and Benjamin Wilson, *Measuring interactions within a dynamical system*, in preparation.
55. Steven T. Piantadosi, *Symbolic dynamics on free groups*, Discrete Contin. Dyn. Syst. **20** (2008), no. 3, 725–738. MR 2373212 (2008k:37026)
56. Natalie Priebe Frank and Lorenzo Sadun, *Fusion tilings with infinite local complexity*, Topology Proc. **43** (2014), 235–276. MR 3104922
57. Anthony Quas and Terry Soo, *Ergodic universality of some topological dynamical systems*, arXiv:1208.3501v3.
58. ———, *Weak mixing suspension flows over shifts of finite type are universal*, J. Mod. Dyn. **6** (2012), no. 4, 427–449. MR 3008405
59. Anthony Quas and Luca Zamboni, *Periodicity and local complexity*, Theoret. Comput. Sci. **319** (2004), no. 1-3, 229–240. MR 2074955 (2005e:68204)
60. E. Arthur Robinson, Jr., *Symbolic dynamics and tilings of* $\mathbb{R}^d$, Symbolic dynamics and its applications, Proc. Sympos. Appl. Math., vol. 60, Amer. Math. Soc., Providence, RI, 2004, pp. 81–119. MR 2078847 (2005h:37036)
61. J. W. Sander and R. Tijdeman, *The complexity of functions on lattices*, Theoret. Comput. Sci. **246** (2000), no. 1-2, 195–225. MR 1780238 (2001g:68085)
62. ———, *Low complexity functions and convex sets in* $\mathbf{Z}^k$, Math. Z. **233** (2000), no. 2, 205–218. MR 1743434 (2000m:37019)
63. ———, *The rectangle complexity of functions on two-dimensional lattices*, Theoret. Comput. Sci. **270** (2002), no. 1-2, 857–863. MR 1871099 (2002j:68080)
64. Boris Solomyak, *Dynamics of self-similar tilings*, Ergodic Theory Dynam. Systems **17** (1997), no. 3, 695–738. MR 1452190 (98f:52030)
65. Giulio Tononi, Olaf Sporns, and Gerald M Edelman, *A measure for brain complexity: relating functional segregation and integration in the nervous system*, Proceedings of the National Academy of Sciences **91** (1994), no. 11, 5033–5037.
66. A. M. Vershik, *A theorem on Markov periodic approximation in ergodic theory*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **115** (1982), 72–82, 306, Boundary value problems of mathematical physics and related questions in the theory of functions, 14. MR 660072 (84m:28030)
67. ———, *A theorem on the Markov periodical approximation in ergodic theory*, J. Sov. Math. **28** (1985), 667–674.
68. Laurent Vuillon, *Combinatoire des motifs d'une suite sturmienne bidimensionnelle*, Theoret. Comput. Sci. **209** (1998), no. 1-2, 261–285. MR 1647534 (99j:68112)
69. Peter Walters, *An Introduction to Ergodic Theory (Graduate Texts in Mathematics)*, Springer, 2000.
70. Benjamin Wilson, *Measuring complexity in dynamical systems*, Ph.D. Dissertation, University of North Carolina at Chapel Hill.
71. Yuki Yayama, *Existence of a measurable saturated compensation function between subshifts and its applications*, Ergodic Theory Dynam. Systems **31** (2011), no. 5, 1563–1589. MR 2832258 (2012j:37052)

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