

Effective symbolic dynamics, random points, statistical behavior, complexity and entropy[☆]

Stefano Galatolo^a, Mathieu Hoyrup^b, Cristóbal Rojas^{b,c}

^a*Dipartimento di Matematica Applicata, Università di Pisa*

^b*Département d'Informatique - Ecole Normale Supérieure, Paris*

^c*CREA, Ecole Polytechnique, Paris*

Abstract

We consider the dynamical behavior of Martin-Löf random points in dynamical systems over metric spaces with a computable dynamics and a computable invariant measure. We use computable partitions to define a sort of effective symbolic model for the dynamics. Through this construction we prove that such points have typical statistical behavior (the behavior which is typical in the Birkhoff ergodic theorem) and are recurrent. We introduce and compare some notions of complexity for orbits in dynamical systems and prove: (i) that the complexity of the orbits of random points equals the Kolmogorov-Sinai entropy of the system, (ii) that the supremum of the complexity of orbits equals the topological entropy.

Contents

1	Introduction	2
2	Preliminaries	5
2.1	Partial recursive functions on integers and numbered sets	5
2.2	Computability over the reals	5
2.3	Computable metric spaces	6
2.4	Computable probability spaces	8
2.5	Algorithmic randomness	9
2.6	Kolmogorov-Chaitin complexity	10
2.6.1	Simple estimates	11
3	Effective symbolic dynamics and statistics of random points	11
3.1	Symbolic dynamics of random points	12
3.2	Some statistical properties of random points	15
3.2.1	Recurrence	15

[☆]partly supported by ANR Grant 05 2452 260 ox

Email addresses: s.galatolo@docenti.ing.unipi.it (Stefano Galatolo), hoyrup@di.ens.fr (Mathieu Hoyrup), rojas@di.ens.fr (Cristóbal Rojas)

3.2.2	Typicalness	15
4	Entropies of dynamical systems	16
4.1	Measure-theoretic entropy	16
4.2	Topological entropy	18
4.2.1	Entropy as a capacity	18
4.2.2	Entropy as a dimension	18
5	Complexity of the orbits of a dynamical system	19
5.1	Symbolic orbit complexity	19
5.2	Shadowing orbit complexity	20
5.3	Equivalence of the two notions of orbit complexity for random points	21
6	Entropy vs orbit complexity	23
6.1	Measure-theoretic entropy	24
6.1.1	A simple proof of Brudno’s first theorem	24
6.1.2	Brudno’s theorems for random points	25
6.2	Topological entropy	26

1. Introduction

The ergodic theory of dynamical systems provides a framework to study the way randomness arises in deterministic systems. For instance, Birkhoff’s ergodic theorem establishes the typical statistical behavior of orbits in a given system, and entropies measure the randomness degree of a process.

On the other hand, computability offers an alternative way of understanding randomness as algorithmic unpredictability. A Martin-Löf random infinite binary sequence can be seen as a sequence having maximal Kolmogorov-Chaitin complexity. The set of such sequences has full measure, and many properties that hold with probability one actually hold for each single random sequence. As an example of statistical properties which hold for each random sequence we recall V’yugin result [V’y97] who proves the Birkhoff theorem for each random symbolic sequence under some computability assumptions on the system.

It is natural to study the relationship between these two different approaches.

The notion of Martin-Löf randomness was first defined for infinite strings and more recently generalized to effective topological spaces in [HW03] and to computable metric spaces in [Gác05, HR09b]. Computable metric spaces are separable metric spaces where the distance can be in some sense effectively computed (see Sect. 2.3). In those spaces, it is also possible to define “computable” functions, which are functions whose behavior is in some sense given by an algorithm, and “computable” measures (there is an algorithm to calculate the measure of nice sets). The space of infinite symbolic sequences, the real line or euclidean spaces, are examples of metric spaces which become computable in a very natural way.

Computable partitions and ergodic theorems for random points

In the classical ergodic theory, the powerful technique of symbolic dynamics allows to associate to an abstract system (X, μ, T) a shift on a space of infinite strings having similar statistical properties. In this paper we define computable measurable partitions (see Sect. 3) and construct an effective version of the above technique, defining the *effective* symbolic models of the dynamics, in which random points are associated to random infinite strings. This tool allows to easily generalize theorems which are proved in the symbolic setting to the more general setting of endomorphisms of computable probability spaces. For instance we use V'yugin's theorem to prove a version of Birkhoff's ergodic theorem for random points.

Theorem (3.2.2). *Let (X, μ) be a computable probability space and x a μ -random point. For any ergodic endomorphism T and any μ -continuity set A*

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} f_A \circ T^i(x) = \mu(A)$$

where f_A is the indicator function of A .

Here, the notion of endomorphism is in a measure-theoretic and computable sense, see Sect. 2.4 for precise definitions. On this line, we also prove a recurrence theorem for random points (Prop. 3.2.1).

In the remaining part of the paper, computable partitions are used to investigate relations between various definitions of orbit complexity, random points and entropy of the system.

Orbit complexity and entropy

In [Bru83], Brudno used Kolmogorov complexity to define a notion of complexity $\overline{\mathcal{K}}(x, T)$ for the orbits of a dynamical system on a compact space. It is a measure of the information rate which is necessary to describe the behavior of the orbit of x . Later, White [Whi93] also introduced a slightly different version $\underline{\mathcal{K}}(x, T)$. The following relations between entropy and orbit complexity were proved:

Theorem (Brudno, White). *Let X be a compact topological space and $T : X \rightarrow X$ a continuous map.*

1. *For any ergodic Borel probability measure μ the equality*

$$\underline{\mathcal{K}}(x, T) = \overline{\mathcal{K}}(x, T) = h_\mu(T)$$

holds for μ -almost all $x \in X$,

2. *For all $x \in X$, $\overline{\mathcal{K}}(x, T) \leq h(T)$.*

Here $h_\mu(T)$ is the Kolmogorov-Sinai entropy of (X, T) with respect to μ and $h(T)$ is the topological entropy of (X, T) . This result seems miraculous as no computability assumption is required on the space or on the transformation T . Actually, this miracle lies in the

compactness of the space, which makes it finite when observations are made with finite precision (open covers of the space can be reduced to *finite* open covers). Indeed, when the space is not compact, it is possible to construct systems for which the complexity $\overline{\mathcal{K}}(x, T)$ of orbits is correlated in no way to their dynamical complexity. In [Gal00], Brudno’s definition was generalized to non-compact computable metric spaces. This definition (see Sect. 5.2) coincides with Brudno’s one in the compact case. However, a relation with entropy was not stated in the non-compact case, or for non-continuous functions. This is in part because these definitions are topological. We propose an alternative notion of orbit complexity $\mathcal{K}_\mu(x, T)$ and prove its relation with entropy for non-compact spaces and for transformations which are not necessarily continuous. Our definition is “measure-theoretical” in the sense that it uses measurable (computable) partitions to encode orbits. With this tool we prove:

Theorem (6.1.2). *Let T be an ergodic endomorphism of the computable probability space (X, μ) ,*

$$\mathcal{K}_\mu(x, T) = h_\mu(T) \quad \text{for all } \mu\text{-random points } x.$$

We then prove that in the compact case our symbolic orbit complexity coincides with Brudno’s one at each random point:

Theorem (5.3.1). *Let T be an ergodic endomorphism of the computable probability space (X, μ) , where X is compact,*

$$\mathcal{K}_\mu(x, T) = \overline{\mathcal{K}}(x, T) \quad \text{for all } \mu\text{-random point } x.$$

The two above statements hence imply a pointwise version of the Brudno’s theorem for each random point.

In the topological context, we then consider $\overline{\mathcal{K}}(x, T)$ and strengthen the second part of Brudno’s theorem, showing:

Theorem (6.2.1). *Let T be a computable map on a compact computable metric space X ,*

$$\sup_{x \in X} \underline{\mathcal{K}}(x, T) = \sup_{x \in X} \overline{\mathcal{K}}(x, T) = h(T)$$

Observe that this was already implied by Brudno’s theorem, using the variational principle: $h(T) = \sup\{h_\mu(T) : \mu \text{ is } T\text{-invariant}\}$. Nevertheless, our proof uses purely topological and algorithmic arguments and no measure. In particular, it does not use the variational principle, and can be thought as an alternative proof of it.

Many of these statements require that the dynamics and the invariant measure be computable. The first assumption can easily be checked on concrete systems if the dynamics is given by a map that is effectively defined.

The second is more delicate: it is well-known that given a map on a metric space, there can be a continuous (even infinite dimensional) space of probability measures which are invariant for the map, and many of them will be non computable. An important part of the theory of dynamical systems is devoted to selecting measures which are particularly meaningful. From this point of view, an important class of these measures is the class of

SRB invariant measures, which are measures being in some sense the “physically meaningful ones” (for a survey on this topic see [You02]). It can be proved (see [GHR09a, GHR09b] and their references e.g.) that in several classes of dynamical systems where SRB measures are proved to exist, these measures are also computable. Hence this provides several classes of nontrivial concrete examples to which our results can be applied.

2. Preliminaries

2.1. Partial recursive functions on integers and numbered sets

In this section we recall some basic facts on recursion, mainly to fix a notation for what follows.

The notion of algorithm has been formalized independently by Turing, Church, Kleene among others. Each constructed model defines a set of partial (not defined everywhere) functions which can be computed by some *effective* mechanical or algorithmic procedure. Later, it has been proved that all this models of computation define the same class of functions, namely: the set of *partial recursive functions*. This fact supports a working hypothesis known as Church’s Thesis, which states that every (intuitively formalizable) algorithm is a partial recursive function. We will not give formal definitions, see for example, [Rog87]. There exists an effective procedure to enumerate the class of all partial recursive functions. More precisely, there is an enumeration $(\varphi_e)_{e \in \mathbb{N}}$ of all the partial recursive functions and a particular recursive function φ_u , called *universal*, such that $\varphi_u(\langle e, n \rangle) = \varphi_e(n)$ for all e, n , where $\langle \cdot, \cdot \rangle : \mathbb{N}^2 \rightarrow \mathbb{N}$ is some effective bijection. A number e such that $\varphi_e = \varphi$ is called a **Gödel number** of φ . Intuitively, it is the number of a program computing φ . A set of natural numbers is called **recursively enumerable** (*r.e.* for short) if it is the range of some partial recursive function, i.e. if there exists an algorithm listing the set. We denote by $E_e := \{\varphi_u(\langle e, n \rangle) : n \in \mathbb{N}\}$ the r.e. set associated to φ_e .

Strictly speaking, the above notions are defined on integers. However, when the objects of some class have been identified with integers, it makes sense to speak about algorithms acting on these objects.

Definition 2.1.1. A **numbered set** is a countable set \mathcal{O} together with a surjection $\nu_{\mathcal{O}} : \mathbb{N} \rightarrow \mathcal{O}$ called the **numbering**. We write o_n for $\nu_{\mathcal{O}}(n)$ and call n a **name** of o_n .

Some classical examples of numbered sets are \mathbb{N}^k , the set of partial recursive functions (with their Gödel numbers), the collection of all r.e. subsets of \mathbb{N} . The set of rational numbers has also a natural numbering $\mathbb{Q} = \{q_0, q_1, \dots\}$ that we fix once for all.

It is straightforward to see how the notion of recursive function and algorithmic enumeration can be extended to numbered sets once a numbering is specified.

2.2. Computability over the reals

We now use the numbered set $\mathbb{Q} = \{q_0, q_1, \dots\}$ to define computability on the set \mathbb{R} of real numbers.

Definition 2.2.1. Let x be a real number. We say that:

- x is **lower semi-computable** if the set $E := \{i \in \mathbb{N} : q_i < x\}$ is r.e.,
- x is **upper semi-computable** if the set $E := \{i \in \mathbb{N} : q_i > x\}$ is r.e.,
- x is **computable** if it is lower and upper semi-computable.

Equivalently, a real number is computable if there exists an algorithmic enumeration of a sequence of rational numbers converging exponentially fast to x . That is:

Proposition 2.2.1. *A real number x is computable if and only if there exists an algorithm $\mathcal{A} : \mathbb{N} \rightarrow \mathbb{Q}$ such that $|\mathcal{A}(i) - x| < 2^{-i}$, for all i .*

Definition 2.2.2. Let $(x_n)_n$ be a sequence of real numbers. We say that x_n is computable **uniformly in n** if there exists an algorithm $\mathcal{A} : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q}$ such that $|\mathcal{A}(n, i) - x_n| < 2^{-i}$ for all n, i .

Sequences of uniformly lower (resp. upper) semi-computable reals are defined in the same way.

2.3. Computable metric spaces

We give a short and self-contained introduction to the concepts from computable analysis on metric spaces that we need in the sequel. More details on the subject can be found in [Wei00, BHW08].

Definition 2.3.1. A **computable metric space** is a triple $\mathcal{X} = (X, d, \mathcal{S})$, where

- (X, d) is a separable complete metric space,
- $\mathcal{S} = \{s_i\}_{i \in \mathbb{N}}$ is a numbered dense subset of X (called **ideal points**),
- the real numbers $d(s_i, s_j)$ are computable, uniformly in i, j .

$(\mathbb{R}^n, d_{\mathbb{R}^n}, \mathbb{Q}^n)$ with the euclidean metric and the standard numbering of \mathbb{Q}^n is an example of computable metric space. Another important example is the Cantor space $(\Sigma^{\mathbb{N}}, d, \mathcal{S})$ with Σ a finite alphabet and d the usual distance¹. In this case \mathcal{S} is the set of ultimately 0-stationary sequences. For further examples we refer to [Wei93].

Like in the case of the real numbers, let us say that a sequence of points $x_i \in X$ converges **fast** to x if $d(x_i, x) < 2^{-i}$ for all i .

Definition 2.3.2 (Computable points). A point $x \in X$ is said to be **computable** if there exists an algorithm $\mathcal{A} : \mathbb{N} \rightarrow \mathcal{S}$ that enumerates a sequence converging fast to x .

Let $(x_n)_n$ be a sequence of computable points. We say that x_n is computable **uniformly in n** if there exists an algorithm $\mathcal{A} : \mathbb{N} \times \mathbb{N} \rightarrow \mathcal{S}$ such that for all n , the sequence $(\mathcal{A}(n, i))_i$ converges fast to x_n .

¹ $d((s_i), (t_i)) = \sum_i |\Sigma|^{-i} \delta(s_i, t_i)$ where $\delta(a, b) = 1$ if $a = b$, 0 otherwise.

Let us consider the set of *ideal balls* $\mathcal{B} := \{B(s_i, q_j) : s_i \in \mathcal{S}, q_j \in \mathbb{Q}, q_j > 0\}$. The numberings of \mathcal{S} and \mathbb{Q} induce a canonical numbering $\mathcal{B} = \{B_0, B_1, \dots\}$ that is fixed once for all.

Definition 2.3.3 (Effective open set). We say that $U \subseteq X$ is an *effective open set* if there is some r.e. set $E \subseteq \mathbb{N}$ such that $U = \bigcup_{i \in E} B_i$.

Let $(U_n)_n$ be a sequence of open sets. We say that U_n is effectively open, *uniformly in n* if there exists a r.e. set $E \subseteq \mathbb{N} \times \mathbb{N}$ such that for all n it holds $U_n = \bigcup_{i:(n,i) \in E} B_i$.

Remarks 2.3.1.

- If U is an effective open set then the set of ideal points belonging to U is r.e.
- If $(U_n)_n$ is a sequence of uniformly effective open sets then the union $\bigcup_n U_n$ is an effective open set.
- The numbering $\{E_0, E_1, \dots\}$ of the r.e. subsets of \mathbb{N} induces a numbering $\{U_0, U_1, \dots\}$ of the collection \mathcal{U} of all effective open subsets of X , defining $U_n = \bigcup_{i \in E_n} B_i$.
- The numbered set \mathcal{U} is closed under finite unions and finite intersections. Furthermore, these operations are *effective* in the following sense: there exist recursive functions φ^\cup and φ^\cap such that for all $i, j \in \mathbb{N}$, $U_i \cup U_j = U_{\varphi^\cup(i,j)}$ and the same holds for φ^\cap . Equivalently, $U_i \cup U_j$ is effectively open, uniformly in i, j .

Definition 2.3.4 (Effective G_δ -set). An *effective G_δ -set* is the intersection $\bigcap_n U_n$ of a family of uniformly effective open sets U_n .

Let (X, d, \mathcal{S}) and (X', d', \mathcal{S}') be computable metric spaces with $\mathcal{B} = \{B_i\}_{i \in \mathbb{N}}$, $\mathcal{B}' = \{B'_i\}_{i \in \mathbb{N}}$ the corresponding numbered sets of ideal open balls.

Definition 2.3.5 (Computable function). A function $T : X \rightarrow X'$ is said to be *computable* if $T^{-1}(B'_n)$ is effectively open uniformly in n .

Remarks 2.3.2.

- The preimage of a sequence of uniformly effective open sets is again a sequence of uniformly effective open sets. This could be an alternative equivalent definition of computable function.
- If T is computable then the images of ideal points can be uniformly computed: $T(s_i)$ is a computable point, uniformly in i .
- The distance function $d : X \times X \rightarrow \mathbb{R}$ is a computable function.

By definition computable functions are continuous. Since we will work with functions that are not necessarily continuous everywhere, we shall consider functions that are computable on some subset of X . More precisely,

Definition 2.3.6. A function T is said to be *computable on $D \subseteq X$* if there are uniformly effective open sets $U_n \subseteq X$ such that $T^{-1}(B'_n) \cap D = U_n \cap D$ for all n . D is called the *domain of computability* of T .

2.4. Computable probability spaces

Now we turn our attention to computability on probability spaces. We will not consider general measurable spaces, but only complete separable metric spaces endowed with the Borel σ -field, as probability and ergodic theory take place mostly on such spaces. Strictly speaking a *computable probability space* should be the computable version of a probability space, given by a set, a σ -field and a probability measure, but for the sake of simplicity, we will use this name for any computable metric space endowed with a computable Borel probability measure (as defined below).

Let then X be a computable metric space. The set of Borel probability measures over X , denoted by $\mathcal{M}(X)$, can be endowed with a structure of computable metric space (this will be defined below, for more details, see [Gác05, HR09b]). A computable measure can then be defined as a computable point of $\mathcal{M}(X)$.

Let us first recall some prerequisites from measure theory. The weak topology on $\mathcal{M}(X)$ is defined by the notion of weak convergence of measures: we say that μ_n converge weakly to μ and write $\mu_n \rightarrow \mu$ if

$$\mu_n \rightarrow \mu \text{ iff } \int f d\mu_n \rightarrow \int f d\mu \text{ for all real bounded continuous } f. \quad (1)$$

Let us recall the Portmanteau theorem. We say that a Borel set A is a set of **μ -continuity** if $\mu(\partial A) = 0$, where $\partial A = \overline{A} \cap \overline{X \setminus A}$ is the boundary of A .

Theorem 2.4.1 (Portmanteau theorem). *Let μ_n, μ be Borel probability measures on a separable metric space (X, d) . The following are equivalent:*

1. μ_n converges weakly to μ ,
2. $\limsup_n \mu_n(F) \leq \mu(F)$ for all closed sets F ,
3. $\liminf_n \mu_n(G) \geq \mu(G)$ for all open sets G ,
4. $\lim_n \mu_n(A) = \mu(A)$ for all μ -continuity sets A .

This theorem easily implies the following: when (X, d) is a separable metric space, weak convergence can be proved using the following criterion:

Proposition 2.4.1. *Let \mathcal{A} be a countable basis of the topology which is closed under the formation of finite unions. If $\mu_n(A) \rightarrow \mu(A)$ for every $A \in \mathcal{A}$, then μ_n converge weakly to μ .*

Let us introduce on $\mathcal{M}(X)$ the structure of a computable metric space. As X is separable and complete, so is $\mathcal{M}(X)$. Let $D \subseteq \mathcal{M}(X)$ be the set of those probability measures that are concentrated in finitely many points of \mathcal{S} and assign rational values to them. It can be shown that this is a dense subset (see [Bil68]).

We consider the Prokhorov metric π on $\mathcal{M}(X)$ defined by:

$$\pi(\mu, \nu) := \inf\{\epsilon \in \mathbb{R}^+ : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for every Borel set } A\}.$$

where $A^\epsilon = \{x : d(x, A) < \epsilon\}$.

This metric induces the weak topology on $\mathcal{M}(X)$. Furthermore, it can be shown that the triple $(\mathcal{M}(X), \pi, D)$ is a computable metric space (see [Gác05]). By Def. 2.3.2 a measure μ is then computable if there is an algorithmic enumeration of a sequence of ideal measures $(\mu_n)_{n \in \mathbb{N}} \subseteq D$ converging fast to μ .

The following theorem gives a characterization for the computability of measures in terms of the computability of the measure of sets (for a proof see [HR09b]):

Theorem 2.4.2. *A measure $\mu \in \mathcal{M}(X)$ is computable if and only if the measures $\mu(B_{i_1} \cup \dots \cup B_{i_k})$ of finite unions of ideal open balls are lower-semi-computable, uniformly in i_1, \dots, i_k .*

Definition 2.4.1. A **computable probability space** is a pair (X, μ) where X is a computable metric space and μ is a computable Borel probability measure on X .

Definition 2.4.2 (Morphism). Let (X, μ) and (Y, ν) be two computable probability spaces. A **morphism** from (X, μ) to (Y, ν) is a measure-preserving function $F : X \rightarrow Y$ which is computable on an effective G_δ -set of μ -measure one.

We recall that F is measure-preserving if $\nu(A) = \mu(F^{-1}(A))$ for every Borel set A . Computable probability structures can be easily transferred:

Proposition 2.4.2. *Let (X, μ) be a computable probability space, Y a computable metric space and $F : X \rightarrow Y$ a function which is computable on an effective G_δ -set of μ -measure one. The induced measure μ_F on Y defined by $\mu_F(A) = \mu(F^{-1}(A))$ is computable and F is a morphism of computable probability space.*

2.5. Algorithmic randomness

The randomness of a particular outcome is always relative to some statistical test. The notion of algorithmically random infinite binary sequence, defined by Martin-Löf in 1966, is an attempt to have an “absolute” notion of randomness. This absoluteness is actually relative to all “effective” statistical tests, and lies on the hypothesis that this class of tests is sufficiently wide.

More recently the notion of Martin-Löf randomness was generalized to effective topological spaces in [HW03] and to computable metric spaces in [Gác05, HR09b]. In this section, (X, μ) is a computable probability space.

Definition 2.5.1. A **Martin-Löf μ -test** is a sequence $(U_n)_{n \in \mathbb{N}}$ of uniformly effective open sets which satisfy $\mu(U_n) < 2^{-n}$ for all n . Any subset of $\bigcap_n U_n$ is called an **effective μ -null set**.

A point $x \in X$ is called **μ -random** if x is contained in no effective μ -null set. The set of μ -random points is denoted by R_μ .

The set R_μ of μ -random points has full measure, so from a measure-theoretic point of view, we can work on R_μ instead of the whole space X . The advantage of this is that many results of the form

$P(x)$ holds for μ -almost every $x \in X$,

with P some predicate, can be converted into an “individual” result

$P(x)$ holds for every $x \in R_\mu$.

To put conversion into practice, we will need the following results (see [HW03], Thm 4.5 or [HR09b]).

Lemma 2.5.1. *Every μ -random point is in every effective open set of full measure.*

Proposition 2.5.1 (Morphisms preserve randomness). *Let F be a morphism of computable probability spaces (X, μ) and (Y, ν) . Then every μ -random point x is in the domain of computability of F and $F(x)$ is ν -random.*

2.6. Kolmogorov-Chaitin complexity

The idea is to define, for a finite object, the minimal amount of algorithmic information from which the object can be recovered. That is, the length of the shortest description (code) of the object. For a complete introduction we refer to standard texts [LV93, Gác].

Let Σ^* and $\Sigma^\mathbb{N}$ be the sets of finite and infinite words (over the finite alphabet Σ) respectively. A word $w \in \Sigma^*$ defines the **cylinder** $[w] \subseteq \Sigma^\mathbb{N}$ of all possible continuations of w . A set $D = \{w_1, w_2, \dots\} \subseteq \Sigma^*$ defines an open set $[D] = \bigcup_i [w_i] \subseteq \Sigma^\mathbb{N}$. D is called **prefix-free** if no word of D is prefix of another one, that is if the cylinders $[w_i]$ are pairwise disjoint.

Let X be Σ^* or \mathbb{N} or \mathbb{N}^* .

Definition 2.6.1. An **interpreter** is a partial recursive function $I : \{0, 1\}^* \rightarrow X$ with a prefix-free domain.

Definition 2.6.2. Let $I : \{0, 1\}^* \rightarrow X$ be an interpreter. The **Kolmogorov-Chaitin complexity** $K_I(x)$ of $x \in X$ relative to I is defined to be

$$K_I(x) = \begin{cases} |p| & \text{if } p \text{ is a shortest input such that } I(p) = x \\ \infty & \text{if there is no } p \text{ such that } I(p) = x. \end{cases}$$

It turns out that there exists an algorithmic enumeration of all the interpreters, which entails the existence of a universal interpreter U that is asymptotically optimal in the sense that the *invariance theorem* holds:

Theorem 2.6.1 (Invariance theorem). *For every interpreter I there exists $c_I \in \mathbb{N}$ such that for all $x \in X$ we have $K_U(x) \leq K_I(x) + c_I$.*

We fix a universal interpreter U and we let $K(x) := K_U(x)$ be the **Kolmogorov-Chaitin complexity** of x .

2.6.1. Simple estimates

Let us recall some simple estimates of the complexity that we will need later. Let f, g be real-valued functions. We say that g **additively dominates** f and write $f \stackrel{+}{\prec} g$ if there is a constant c such that $f \leq g + c$. As codes are always binary words, we use base-2 logarithms, which we denote by \log . We define $J(x) = x + 2 \log(x + 1)$ for $x \geq 0$. For $n \in \mathbb{N}$, $K(n) \stackrel{+}{\prec} J(\log n)$. For $n_1, \dots, n_k \in \mathbb{N}$, $K(n_1, \dots, n_k) \stackrel{+}{\prec} K(n_1) + \dots + K(n_k)$. The following property is a version of a result attributed to Kolmogorov, stated in terms of prefix complexity instead of plain complexity.

Proposition 2.6.1. *Let $E \subseteq \mathbb{N} \times X$ be a r.e. set such that $E_n = \{x : (n, x) \in E\}$ is finite for all n . Then for all $n \in \mathbb{N}$ and $s \in E_n$,*

$$K(s) \stackrel{+}{\prec} J(\log |E_n|) + K(n)$$

Proposition 2.6.2. *Let μ be a computable measure on $\Sigma^{\mathbb{N}}$. For all $w \in \Sigma^*$,*

$$K(w) \stackrel{+}{\prec} -\log \mu([w]) + K(|w|)$$

Theorem 2.6.2 (Coding theorem). *Let $P : X \rightarrow \overline{\mathbb{R}}^+$ be a lower semi-computable function such that $\sum_{x \in X} P(x) \leq 1$. Then $K(x) \stackrel{+}{\prec} -\log P(x)$, i.e. there is a constant c such that $K(x) \leq -\log P(x) + c$ for all $x \in X$.*

Moreover, the quantity $\sum_x 2^{-K(x)}$ is finite and smaller than 1 as it is the Lebesgue measure of the domain of the universal interpreter U . There is a relation between Kolmogorov-Chaitin complexity and randomness, initial segments of random infinite strings being maximally complex.

Theorem 2.6.3 (Schnorr). *Let μ be a computable measure over the finite alphabet Σ . Then $\omega \in \Sigma^{\mathbb{N}}$ is a μ -random sequence if and only if $\exists m \forall n K(\omega_{1:n}) \geq -\log \mu[\omega_{1:n}] - m$.*

The minimal such m , defined by $d_\mu(\omega) := \sup_n \{-\log \mu[\omega_{1:n}] - K(\omega_{1:n})\}$ and called the **randomness deficiency** of ω w.r.t μ , is not only finite almost everywhere: it has finite mean, that is $\int d_\mu(\omega) d\mu \leq 1$. For a proof see [LV93].

3. Effective symbolic dynamics and statistics of random points

Let us recall some basic facts about ergodic theory (see [Pet83, Wal82] for an introduction). Let X be a metric space, let $T : X \rightarrow X$ be a Borel measurable map. Let μ be a T -invariant Borel probability measure, i.e. a Borel probability measure on X such that $\mu(A) = \mu(T^{-1}(A))$ for each measurable set A . A measurable set A is called **T -invariant** if $T^{-1}(A) = A \pmod 0$ (the symmetric difference between the two sets has zero measure). The system (X, μ, T) is said to be **ergodic** if each T -invariant set has total or null measure. In such systems the famous Birkhoff ergodic theorem says that time averages computed along

μ -typical orbits coincides with space average with respect to μ . More precisely, for any $f \in L^1(X, \mu)$ it holds

$$\lim_{n \rightarrow \infty} \frac{S_n^f(x)}{n} = \int f \, d\mu, \quad (2)$$

for μ -almost each x , where $S_n^f = f + f \circ T + \dots + f \circ T^{n-1}$.

Now that we have algorithmic randomness at our disposal, it is then natural to ask if μ -random points satisfy Birkhoff's ergodic theorems, and for which transformations and observables. This problem was tackled by V'yugin [V'y97] who gave a positive answer on the Cantor space for *computable* transformations and *computable* observables. It was later observed in [Nan08] that V'yugin's theorem does not imply anything for discontinuous observables (which cannot be computable), and an extension of V'yugin's theorem for observables that are computable but on a countable set was then carried out. In this section, we develop the framework of *effective symbolic dynamics* to show that V'yugin's theorem does imply a more general result. Let us first state the problem in a slightly different way.

It is a classical result that the set of points x such that (2) holds for all continuous bounded f has measure one. Such points are called **μ -typical**. This notion can be reformulated. Given a point x , let us consider the measures $\nu_n = \frac{1}{n} \sum_{j < n} \delta_{T^j x}$, where δ_y is the Dirac probability measure concentrated on y . Let μ be an ergodic measure for T . A point x is μ -typical if and only if the associated measures ν_n converge weakly to μ .

Now the question is: are μ -random points μ -typical?

We now develop some more tools to give a positive answer to this question on any computable probability space and for any ergodic endomorphism (Def. 2.4.2). This result (Thm. 3.2.2 below) implies equality (2) for random points, bounded continuous (not necessarily computable) observables and indicators of sets of μ -continuity (without effectivity assumption). In a sequel paper [HR09a] we give a much more general answer, proving a version of Birkhoff's ergodic theorem for random points and *effectively μ -measurable* transformations and observables.

3.1. Symbolic dynamics of random points

Let T be an endomorphism of the probability space (X, μ) . In the classical construction of symbolic dynamics associated to a given system, one considers access to the system given by a finite measurable partition, that is a finite collection of pairwise disjoint Borel sets $\xi = \{C_1, \dots, C_k\}$ such that $\mu(\bigcup_i C_i) = 1$. To almost each point $x \in X$ corresponds an infinite sequence $\omega = (\omega_i)_{i \in \mathbb{N}} = \phi_\xi(x) \in \{1, \dots, k\}^{\mathbb{N}}$ defined by:

$$\phi_\xi(x) = \omega \iff \forall j \in \mathbb{N}, T^j(x) \in C_{\omega_j}$$

As ξ is a measurable partition, the map ϕ_ξ is measurable and then the measure μ induces the measure μ_ξ on $\{1, \dots, k\}^{\mathbb{N}}$ defined by $\mu_\xi(B) = \mu(\phi_\xi^{-1}(B))$ for all measurable sets $B \subseteq \{1, \dots, k\}^{\mathbb{N}}$. Let us define the *shift* endomorphism $\sigma : \{1, \dots, k\}^{\mathbb{N}} \rightarrow \{1, \dots, k\}^{\mathbb{N}}$ by $\sigma((\omega_i)_{i \in \mathbb{N}}) = (\omega_{i+1})_{i \in \mathbb{N}}$. The symbolic dynamical system $(\{1, \dots, k\}^{\mathbb{N}}, \mu_\xi, \sigma)$ is called the *symbolic model* of (X, μ, T) w.r.t. ξ .

The requirement of ϕ_ξ being measurable makes the symbolic model appropriate from the measure-theoretic view point, but is not enough to have a symbolic model compatible with the computational approach:

Definition 3.1.1. Let T be an endomorphism of the computable probability space (X, μ) and $\xi = \{C_1 \dots, C_k\}$ a finite measurable partition.

The associated symbolic model $(\{1, \dots, k\}^\mathbb{N}, \mu_\xi, \sigma)$ is said to be an **effective symbolic model** if the map $\phi_\xi : X \rightarrow \{1, \dots, k\}^\mathbb{N}$ is a morphism of computable probability space (here the space $\{1, \dots, k\}^\mathbb{N}$ is endowed with the standard computable structure).

The sets C_i are called the **atoms** of ξ and we denote by $\xi(x)$ the atom containing x (if there is one). Observe that ϕ_ξ is computable on its domain only if the atoms are effective open sets (in the domain). We hence define:

Definition 3.1.2 (Computable partitions). A measurable partition ξ is said to be a **computable partition** if its atoms are effective open sets.

Theorem 3.1.1. *Let T be an endomorphism of the computable probability space (X, μ) and $\xi = \{C_1 \dots, C_k\}$ a finite computable partition. The associated symbolic model is effective.*

Proof. Let D be the domain of computability of T (it is a full-measure effective G_δ -set). Define the set

$$X_\xi = D \cap \bigcap_{n \in \mathbb{N}} T^{-n}(C_1 \cup \dots \cup C_k)$$

X_ξ is a full-measure effective G_δ -set: indeed, as $C_1 \cup \dots \cup C_k$ is effectively open and T is computable on D there are uniformly effective open sets U_n such that $D \cap T^{-n}(C_1 \cup \dots \cup C_k) = D \cap U_n$, so $X_\xi = D \cap \bigcap_n U_n$. As T is measure-preserving, all U_n have measure one.

Now, $X_\xi \cap \phi_\xi^{-1}[i_0, \dots, i_n] = X_\xi \cap C_{i_0} \cap T^{-1}C_{i_1} \cap \dots \cap T^{-n}C_{i_n}$. This proves that ϕ_ξ is computable over X_ξ . Proposition 2.4.2 allows to conclude. \square

After the definition an important question is: are there computable partitions? the answer depends on the existence of effective open sets with a zero-measure boundary.

Definition 3.1.3. A set A is said to be **almost decidable** if there are two effective open sets U and V such that:

$$U \subseteq A, \quad V \subseteq A^c, \quad \mu(U) + \mu(V) = 1$$

Remarks 3.1.1.

- a set is almost decidable if and only if its complement is almost decidable,
- an almost decidable set is always a continuity set,
- a μ -continuity ideal ball is always almost decidable,

Ignoring computability, the existence of open μ -continuity sets directly follows from the fact that the collection of open sets is uncountable and μ is finite. The problem in the computable setting is that there are only countable many effective open sets. Fortunately, there still always exists a basis of almost decidable balls. This will be used many times in the sequel, in particular it directly implies the existence of computable partitions. This result was independently obtained in [Bos08a, Bos08b, HR09b].

Theorem 3.1.2. *Let (X, μ) be a computable probability space. There is a sequence of uniformly computable reals $(r_j)_{j \in \mathbb{N}}$ that is dense in \mathbb{R}^+ and such that for every i, j , the ball $B(s_i, r_j)$ is almost decidable.*

We denote by B_n^μ the almost decidable ball $B(s_i, r_j)$ with $n = \langle i, j \rangle$. The family $\{B_n^\mu : n \in \mathbb{N}\}$ is a basis for the topology. It is even effectively equivalent to the basis of ideal balls : every ideal ball can be expressed as a r.e. union of almost decidable balls, and vice-versa. We finish presenting some results that will be needed in the next subsection.

Corollary 3.1.1. *On every computable probability space, there exists a family of uniformly computable partitions which generates the Borel σ -field.*

Proof. Take $\xi_{(i,j)} = \{B(s_i, r_j), X \setminus \overline{B}(s_i, r_j)\}$ where \overline{B} is the closed ball: as the almost decidable balls form a basis of the topology, the σ -field generated by the ξ_n is the Borel σ -field. \square

Proposition 3.1.1. *If A is almost decidable then $\mu(A)$ is a computable real number.*

Proof. Since U and V are effectively open, by Thm. 2.4.2 their measures are lower-semi-computable. As $\mu(U) + \mu(V) = 1$, their measures are also upper-semi-computable. \square

The following regards the computability of inducing a measure in a subset and will be used in the proof of prop. 3.2.1

Proposition 3.1.2. *Let μ be a computable measure and A be an almost decidable subset of X with $\mu(A) > 0$. Then the induced measure $\mu_A(\cdot) = \mu(\cdot|A)$ is computable. Furthermore, $R_{\mu_A} = R_\mu \cap A$.*

Proof. Let A be an almost decidable set, coming with U, V . Let $W = B_{n_1} \cup \dots \cup B_{n_k}$ be a finite union of ideal balls. As $A = U \bmod 0$, one has

$$\mu_A(W) = \mu(W \cap A) / \mu(A) = \mu(W \cap U) / \mu(A).$$

$W \cap U$ is an effective open set, so its measure is lower semi-computable. As $\mu(A)$ is computable, $\mu_A(W)$ is lower semi-computable. Note that everything is uniform in n_1, \dots, n_k . The result follows from Thm. 2.4.2.

Let U and V as in the definition of an almost decidable set. First note that $R_\mu \cap A = R_\mu \cap U$, as $R_\mu \subseteq U \cup V$ by Lem. 2.5.1. Again by Lem. 2.5.1, $R_{\mu_A} \subseteq U$, and as $\mu_A \leq \frac{1}{\mu(A)}\mu$, every μ -effective null set is also a μ_A -effective null set, so $R_{\mu_A} \subseteq R_\mu$. Hence, we have $R_{\mu_A} \subseteq R_\mu \cap U$.

Conversely, $R_{\mu_A}^c$ being a μ_A -effective null set, its intersection with U is a μ -effective null set, by definition of μ_A . So $R_{\mu_A}^c \cap U \subseteq R_\mu^c$, which is equivalent to $R_\mu \cap U \subseteq R_{\mu_A}$. \square

3.2. Some statistical properties of random points

Before coming back to typicalness of random points, let us study a weaker property, namely recurrence, for which the version for random points has a more simple proof.

3.2.1. Recurrence

We recall that the Poincaré recurrence theorem states that in a measure-preserving system, for each set E almost each orbit starting from E comes back to E infinitely often. On a metric space we can also consider:

Definition 3.2.1. Let X be a metric space. A point $x \in X$ is said to be **recurrent** for a transformation $T : X \rightarrow X$, if $\liminf_n d(x, T^n x) = 0$.

Poincaré recurrence theorem implies that in a measure-preserving transformation almost each point are recurrent. Under suitable computability assumptions the same holds for *all* random points.

Proposition 3.2.1 (Random points are recurrent). *Let (X, μ) be a computable probability space. If x is μ -random, then it is recurrent with respect to every endomorphism T of (X, μ) .*

Proof. Let x be μ -random and B an almost decidable open ball containing x . If B was a μ -null set, it would be an effective μ -null set and could not contain x , which is μ -random. Hence $\mu(B) > 0$, $\mu_B(\cdot) = \mu(\cdot|B)$ is well-defined and x is μ_B -random by Prop. 3.1.2. Let D be the domain of computability of T . There is an effective open set U such that:

$$\bigcup_{n \geq 1} T^{-n}B \cap D = U \cap D.$$

The Poincaré recurrence theorem states that μ -almost every point in B comes back to B , so $\mu_B(U) = 1$. As x is μ_B -random, $x \in U$ by Lem. 2.5.1. \square

3.2.2. Typicalness

To prove that μ -random points are μ -typical, we will use the following particular case of V'yugin's main theorem.

Theorem 3.2.1 (V'yugin). *Let μ be a computable shift-invariant ergodic measure on the Cantor space $\{0, 1\}^{\mathbb{N}}$. For each μ -random sequence ω :*

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} \omega_i = \mu([1]). \quad (3)$$

We are now able to prove:

Theorem 3.2.2. *Let (X, μ) be a computable probability space. Then each μ -random point x is μ -typical for every ergodic endomorphism T .*

Proof. Let f_A be the characteristic function of the set A . First, let us show that if A is an almost decidable set then for all μ -random point x :

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} f_A \circ T^i(x) = \mu(A) \quad (4)$$

Indeed, consider the *computable partition* defined by $\xi := \{U, V\}$ with U and V as in Def. 3.1.3 and the associated symbolic model $(\{1, \dots, k\}^{\mathbb{N}}, \mu_\xi, \sigma)$. By Theorem 3.1.1 and Proposition 2.5.1 $\phi_\xi(x)$ is a well-defined μ_ξ -random infinite sequence, so Thm. 3.2.1 applies and gives (4). As explained at the beginning of Sect. 3, this can be reformulated as the convergence of $\nu_n(A)$ to $\mu(A)$, where $\nu_n = \frac{1}{n} \sum_{j < n} \delta_{T^j x}$. Now, the collection of almost decidable sets satisfies Prop. 2.4.1, so ν_n converges weakly to μ : x is μ -typical. \square

Observe that the version of Birkhoff's theorem for random holds all ergodic endomorphisms (of computable probability space, i.e., in both a measure-theoretic and computable sense, see Def. 2.4.2) and for all bounded continuous (not necessarily computable) observables. Moreover, by Prop. 2.4.1 it holds for all indicators of μ -continuity sets (again, without any computability assumption). This result hence improves [Nan08].

4. Entropies of dynamical systems

4.1. Measure-theoretic entropy

Suppose that symbols from a finite alphabet are produced by some source at each integer time. The tendency of the source toward producing such object more than such other can be modeled by a probability distribution. The Shannon entropy of the source measures the degree of uncertainty about future symbols.

Any ergodic dynamical system (X, μ, T) can be seen as a source of outputs. Kolmogorov and Sinai adapted Shannon's theory to dynamical systems in order to measure the degree of unpredictability or chaoticity of an ergodic system. The first step consists in discretizing the space X using finite partitions. Let $\xi = \{C_1, \dots, C_n\}$ be a finite measurable partition of X . Then let $T^{-k}\xi$ be the partition whose atoms are the pre-images $T^{-k}C_i$. Then let

$$\xi_n = \xi \vee T^{-1}\xi \vee T^{-2}\xi \vee \dots \vee T^{-(n-1)}\xi$$

be the partition given by the sets of the form

$$C_{i_0} \cap T^{-1}C_{i_1} \cap \dots \cap T^{-(n-1)}C_{i_{n-1}},$$

varying C_{i_j} among all the atoms of ξ . Knowing which atom ξ_n a point x belongs to comes to knowing which atoms of the partition ξ the orbit of x visits up to time $n - 1$.

The measure-theoretical entropy of the system w.r.t. the partition ξ can then be thought as the rate (per time unit) of gained information (or removed uncertainty) when observations of the type " $T^n(x) \in C_i$ " are performed. This is of great importance when classifying dynamical systems: it is a measure-theoretical invariant, which enables one to distinguish non-isomorphic systems.

We briefly recall the definition. For more details, we refer the reader to [Bil65, Wal82, Pet83, HK95].

Given a partition ξ and a point x , $\xi(x)$ denotes the atom of the partition x belongs to. Let us consider the **Shannon information function** relative to the partition ξ_n (the information which is gained by observing that $x \in \xi_n(x)$),

$$I_\mu(x|\xi_n) := -\log \mu(\xi_n(x))$$

and its mean, the entropy of the partition ξ_n ,

$$H_\mu(\xi_n) := \int I_\mu(\cdot|\xi_n) d\mu = \sum_{C \in \xi_n} -\mu(C) \log \mu(C)$$

The **measure-theoretical** or **Kolmogorov-Sinai entropy** of T relative to the partition ξ is defined as:

$$h_\mu(T, \xi) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(\xi_n).$$

(which exists and is an infimum, since the sequence $H_\mu(\xi_n)_n$ is sub-additive). With the Shannon information function, it is possible to define a kind of point-wise notion of entropy with respect to a partition ξ :

$$\limsup_n \frac{1}{n} I_\mu(x|\xi_n).$$

This point-wise entropy is related to the global entropy of the system by the celebrated Shannon-McMillan-Breiman theorem:

Theorem (Shannon-McMillan-Breiman). *Let T be an ergodic measure preserving transformation of (X, \mathcal{B}, μ) and ξ a finite measurable partition. Then for μ -almost every x ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} I_\mu(x|\xi_n) = h_\mu(T, \xi). \quad (5)$$

The convergence also holds in $L^1(X, \mathcal{B}, \mu)$.

The partition-dependency is suppress taking the supremum over finite measurable partitions: the **Kolmogorov-Sinai entropy** of (X, μ, T) is

$$h_\mu(T) := \sup\{h_\mu(T, \xi) : \xi \text{ finite measurable partition}\}.$$

We recall the following two results that we will need later. The first proposition follows directly from the definitions.

Proposition 4.1.1. *If $(\Sigma^{\mathbb{N}}, \mu_\xi, \sigma)$ is the symbolic model associated to (X, μ, T) w.r.t. ξ then $h_\mu(T, \xi) = h_{\mu_\xi}(\sigma)$.*

The next proposition is taken from [Pet83]:

Proposition 4.1.2. *If $(\xi_i)_{i \in \mathbb{N}}$ is a family of finite measurable partitions which generates the Borel σ -field up to sets of measure 0, then $h_\mu(T) = \sup_i h_\mu(T, \xi_0 \vee \dots \vee \xi_i)$.*

4.2. Topological entropy

In this section, X is a metric space and $T : X \rightarrow X$ a continuous map.

Bowen's definition of topological entropy is reminiscent of the capacity (or box counting dimension) of a totally bounded subset of a metric space. We first recall this definition, and then present another characterization given by Pesin, expressing the topological entropy as a kind of Hausdorff dimension. We will use it in the sequel.

4.2.1. Entropy as a capacity

We recall the definition: for $n \geq 0$, let us define the distance $d_n(x, y) = \max\{d(T^i x, T^i y) : 0 \leq i < n\}$ and the Bowen ball $B_n(x, \epsilon) = \{y : d_n(x, y) < \epsilon\}$, which is open by continuity of T . Given a totally bounded set $Y \subseteq X$ and numbers $n \geq 0, \epsilon > 0$, let $N(Y, n, \epsilon)$ be the minimal cardinality of a cover of Y by Bowen balls $B_n(x, \epsilon)$. A set of points E such that $\{B_n(x, \epsilon) : x \in E\}$ is a cover of Y is also called an (n, ϵ) -spanning set of Y . One then defines:

$$h_1(T, Y, \epsilon) = \limsup_{n \rightarrow \infty} \frac{\log N(Y, n, \epsilon)}{n}$$

which is non-decreasing as $\epsilon \rightarrow 0$, so the following limit exists:

$$h_1(T, Y) = \lim_{\epsilon \rightarrow 0} h_1(Y, T, \epsilon).$$

When X is compact, the **topological entropy** of T is $h(T) = h_1(T, X)$. It measures the exponential growth-rate of the number of distinguishable orbits of the system.

Remark 4.2.1. The topological entropy can be defined using separated sets instead of open covers: a subset A of X is (n, ϵ) -separated if for any distinct points $x, y \in A$, $d_n(x, y) > \epsilon$. Let us define $M(Y, n, \epsilon)$ as the maximal cardinality of an (n, ϵ) -separated subset of Y . It is easy to see that $M(Y, n, 2\epsilon) \leq N(Y, n, \epsilon) \leq M(Y, n, \epsilon)$, and hence $h_1(T, Y)$ can be alternatively defined using $M(Y, n, \epsilon)$ in place of $N(Y, n, \epsilon)$.

4.2.2. Entropy as a dimension

It is possible to define a topological entropy which is an analog of Hausdorff dimension. This definition coincides with the classical one in the compact case. Hausdorff dimension has stronger stability properties than box dimension, which has important consequences, as we will see in what follows. We refer the reader to [Pes98, HK02] for more details.

Let X be a metric space and $T : X \rightarrow X$ a continuous map. The ϵ -size of $E \subseteq X$ is 2^{-s} where

$$s = \sup\{n \geq 0 : \text{diam}(T^i E) \leq \epsilon \text{ for } 0 \leq i < n\}.$$

It measures how long the orbits starting from E are ϵ -close. As ϵ decreases, the ϵ -size of E is non-decreasing. The 2ϵ -size of a Bowen ball $B_n(x, \epsilon)$ is at most 2^{-n} .

In a way that is reminiscent from the definition of Hausdorff measure, let us define

$$m_\delta^s(Y, \epsilon) = \inf_{\mathcal{G}} \left\{ \sum_{U \in \mathcal{G}} (\epsilon\text{-size}(U))^s \right\}$$

where the infimum is taken over all countable covers \mathcal{G} of Y by open sets of ϵ -size $< \delta$. This quantity is monotonically increasing as δ tends to 0, so the limit $m^s(Y, \epsilon) := \lim_{\delta \rightarrow 0^+} m_\delta^s(Y, \epsilon)$ exists and is a supremum. There is a critical value s_0 such that $m^s(Y, \epsilon) = \infty$ for $s < s_0$ and $m^s(Y, \epsilon) = 0$ for $s > s_0$. Let us define $h_2(T, Y, \epsilon)$ as this critical value:

$$h_2(T, Y, \epsilon) := \inf \{s : m^s(Y, \epsilon) = 0\} = \sup \{s : m^s(Y, \epsilon) = \infty\}.$$

As less and less covers are allowed when $\epsilon \rightarrow 0$ (the ϵ -size of sets does not decrease), the following limit exists

$$h_2(T, Y) := \lim_{\epsilon \rightarrow 0^+} h_2(T, Y, \epsilon)$$

and is a supremum. In [Pes98], it is proved that:

Theorem 4.2.1. *When Y is a T -invariant compact set, $h_1(T, Y) = h_2(T, Y)$.*

In particular, if the space X is compact, then $h(T) = h_1(T, X) = h_2(T, X)$.

5. Complexity of the orbits of a dynamical system

5.1. Symbolic orbit complexity

In this section, T is an endomorphism of the computable probability space (X, μ) and $\xi = \{C_1, \dots, C_k\}$ is a computable partition. Let $(\Sigma^{\mathbb{N}}, \mu_\xi, \sigma)$ be the effective symbolic model of (X, μ, T, ξ) where $\Sigma = \{1, \dots, k\}$ (see Sect. 3.1).

Kolmogorov-Chaitin complexity (see Sect. 2.6) was introduced as a quantity of information, on the same level as Shannon information. When the measure, the transformation and the partition are computable, it makes sense to define the algorithmic equivalents of the notions defined above. It turns out that the two points of view are strongly related.

An atom C of the partition ξ_n can then be seen as a word of length n on the alphabet Σ , which allows one to consider its Kolmogorov-Chaitin complexity $K(C)$. For those points whose all iterates are covered by ξ (they form a dense effective G_δ -set of full measure), we define the **Kolmogorov-Chaitin information function** relative to the partition ξ_n :

$$\mathcal{I}(x|\xi_n) := K(\xi_n(x))$$

which is independent of μ . We can then define **algorithmic entropy** of the partition ξ_n as the mean of \mathcal{I} :

$$\mathcal{H}_\mu(\xi_n) := \int \mathcal{I}(\cdot|\xi_n) d\mu = \sum_{C \in \xi_n} \mu(C)K(C).$$

We also define a point-wise notion of algorithmic entropy, which we call symbolic orbit complexity:

Definition 5.1.1 (Symbolic orbit complexity). Let T be an endomorphism of the computable probability space (X, μ) . For any finite computable partition ξ , we define

$$\begin{aligned} \mathcal{K}_\mu(x, T|\xi) &:= \limsup_n \frac{1}{n} \mathcal{I}(x|\xi_n), \\ \mathcal{K}_\mu(x, T) &:= \sup\{\mathcal{K}_\mu(x, T|\xi) : \xi \text{ computable partition}\}. \end{aligned}$$

The quantity $\mathcal{K}_\mu(x, T|\xi)$ was introduced by Brudno in [Bru83] without any computability restriction on the space, the measure nor the transformation. He could not suppress the dependency on ξ by taking the supremum over all finite partitions, as he remarked that this supremum is infinite as soon as the orbit of x is not eventually periodic. Here we restrict the class of admissible partitions to some class of regular but meaningful partitions (see also [BBG⁺04] Sect. 4 or [Ken08]). We will see through Thm. 6.1.2 that this restricted supremum makes sense.

Without the notion of computable partition, Brudno did not go further with this approach and proposed a topological definition using open covers instead of partitions, that we present now, in the more general version proposed in [Gal00].

5.2. Shadowing orbit complexity

In this section, (X, d, \mathcal{S}) is a computable metric space and $T : X \rightarrow X$ a transformation (for the moment, no continuity or computability assumption is put on T). We will consider a notion of orbit complexity which quantifies the algorithmic information needed to describe the orbit of x with finite but arbitrarily accurate precision. This definition was introduced by one of the authors in [Gal00] who proved that it coincides on compact spaces and for continuous maps with Brudno's original definition (using open covers).

Given $\epsilon > 0$ and $n \in \mathbb{N}$, the algorithmic information that is needed to list a sequence of ideal points which follows the orbit of x for n steps at a distance less than ϵ is:

$$\mathcal{K}_n(x, T, \epsilon) := \min\{K(i_0, \dots, i_{n-1}) : d(s_{i_j}, T^j x) < \epsilon \text{ for } j = 0, \dots, n-1\}$$

where K is the Kolmogorov-Chaitin complexity.

We then define the maximal and minimal growth-rates of this quantity:

$$\begin{aligned} \overline{\mathcal{K}}(x, T, \epsilon) &:= \limsup_{n \rightarrow \infty} \frac{1}{n} \mathcal{K}_n(x, T, \epsilon) \\ \underline{\mathcal{K}}(x, T, \epsilon) &:= \liminf_{n \rightarrow \infty} \frac{1}{n} \mathcal{K}_n(x, T, \epsilon). \end{aligned}$$

As ϵ tends to 0, these quantities increase (or at least do not decrease), hence they have limits (which can be infinite).

Definition 5.2.1. The *upper* and *lower shadowing orbit complexities* of x under T are respectively defined by:

$$\begin{aligned} \overline{\mathcal{K}}(x, T) &:= \lim_{\epsilon \rightarrow 0^+} \overline{\mathcal{K}}(x, T, \epsilon) \\ \underline{\mathcal{K}}(x, T) &:= \lim_{\epsilon \rightarrow 0^+} \underline{\mathcal{K}}(x, T, \epsilon). \end{aligned}$$

Remark 5.2.1. If T is computable, and assuming that ϵ takes only rational values, the n first iterates of x could be ϵ -shadowed by the orbit of a single ideal point instead of a pseudo-orbit

of n ideal points. Actually it is easy to see that it gives the same quantities $\overline{\mathcal{K}}(x, T, \epsilon)$ and $\underline{\mathcal{K}}(x, T, \epsilon)$: let $\mathcal{K}'_n(x, T, \epsilon) = \min\{K(i) : d(T^j s_i, T^j x) < \epsilon \text{ for } j < n\}$, one has:

$$\begin{aligned}\mathcal{K}'_n(x, T, 2\epsilon) &\stackrel{\pm}{\leq} \mathcal{K}_n(x, T, \epsilon) + K(\epsilon) \\ \mathcal{K}_n(x, T, \epsilon) &\stackrel{\pm}{\leq} \mathcal{K}'_n(x, T, \epsilon/2) + K(n, \epsilon)\end{aligned}$$

Indeed, from ϵ and i_0, \dots, i_{n-1} some ideal point can be algorithmically found in the effective open set $B(s_{i_0}, \epsilon) \cap \dots \cap T^{-(n-1)}B(s_{i_{n-1}}, \epsilon)$, uniformly in i_0, \dots, i_{n-1} . Its n first iterates 2ϵ -shadow the orbit of x , which proves the first inequality. For the second inequality, some i_0, \dots, i_{n-1} can be algorithmically found from n, ϵ , and a point s_i whose n first iterates $\epsilon/2$ -shadow the orbit of x , taking any $s_{i_j} \in B(T^j s_i, \epsilon/2)$.

Remark 5.2.2. Under the same assumptions, one could define $K(B_n(s_i, \epsilon))$ to be $K(i, n, \epsilon)$, and replace $K(i)$ by $K(B_n(s_i, \epsilon))$ in the definition of $\mathcal{K}'_n(x, T, \epsilon)$, without changing the quantities $\overline{\mathcal{K}}(x, T, \epsilon)$ and $\underline{\mathcal{K}}(x, T, \epsilon)$. Indeed,

$$K(i) \stackrel{\pm}{\leq} K(B_n(s_i, \epsilon)) \stackrel{\pm}{\leq} K(i) + K(n) + K(\epsilon)$$

5.3. Equivalence of the two notions of orbit complexity for random points

We now prove:

Theorem 5.3.1. *Let T be an ergodic endomorphism of the computable probability space (X, μ) , where X is compact. Then for every μ -random point x ,*

$$\overline{\mathcal{K}}(x, T) = \mathcal{K}_\mu(x, T).$$

Proof of $\overline{\mathcal{K}}(x, T) \leq \mathcal{K}_\mu(x, T)$. Let $\epsilon > 0$. Choose a computable partition ξ of diameter $< \epsilon$ (this is why we require X to be compact). To every cell of ξ , associate an ideal point which is inside (as ξ is computable, this can be done in a computable way, but we actually do not need that). The translation of symbolic sequences in sequences of ideal points through this finite dictionary is effective, and transforms the symbolic orbit of a point x into a sequence of ideal points which is ϵ -close to the orbit of x . So $\overline{\mathcal{K}}(x, T, \epsilon) \leq \mathcal{K}_\mu(x, T|\xi)$. The inequality follows letting ϵ tend to 0. \square

To prove the other inequality, we recall some technical stuff. The Kolmogorov-Chaitin complexity of natural numbers $k \geq 1$ satisfies

$$K(k) \stackrel{\pm}{\leq} f(k)$$

where $f(x) = \log x + 1 + 2 \log(\log x + 1)$ for all $x \in \mathbb{R}, x \geq 1$. f is a concave increasing function and $x \mapsto xf(1/x)$ is an increasing function on $(0, 1/2]$ which tends to 0 as $x \rightarrow 0$.

We recall that for finite sequences of natural numbers (k_1, \dots, k_n) , one has

$$K(k_1, \dots, k_n) \stackrel{\pm}{\leq} K(k_1) + \dots + K(k_n)$$

as the shortest descriptions for k_1, \dots, k_n can be extracted from their concatenation (this is one reason to use the self-delimiting Kolmogorov-Chaitin complexity instead of the plain Kolmogorov complexity).

Lemma 5.3.1. *Let Σ be a finite alphabet and $n \in \mathbb{N}$. Let $u, v \in \Sigma^n$ and $0 < \alpha < 1/2$ such that the density of the set of positions where u and v differ is less than α , that is:*

$$\frac{1}{n} |\{i \leq n : u_i \neq v_i\}| < \alpha < 1/2$$

Then $|\frac{1}{n}K(u) - \frac{1}{n}K(v)| < \alpha f(1/\alpha) + \alpha f(|\Sigma|) + \frac{c}{n}$ where c is a constant independent of u, v and n .

Proof. Let (i_1, \dots, i_p) be the ordered sequence of indices where u and v differ. By hypothesis, $p/n < \alpha$. Put $k_1 = i_1$ and $k_j = i_j - i_{j-1}$ for $2 \leq j \leq p$.

We now show that u can be recovered from v and roughly $\alpha(f(1/\alpha) + f(|\Sigma|))n$ bits more. Indeed u can be computed from $(v, k_1, \dots, k_p, u_{i_1}, \dots, u_{i_p})$, constructing the string which coincides with v everywhere but at positions $k_1, k_1 + k_2, \dots, k_1 + \dots + k_p$, where the symbols u_{i_1}, \dots, u_{i_p} are used instead. Hence $K(u) \stackrel{+}{\leq} K(v) + K(k_1) + \dots + K(k_p) + K(u_{i_1}) + \dots + K(u_{i_p}) \stackrel{+}{\leq} K(v) + f(k_1) + \dots + f(k_p) + pf(|\Sigma|)$ as each symbol of Σ can be identified with a natural number between 1 and $|\Sigma|$.

Now, as f is a concave increasing function, one has:

$$\frac{1}{p} \sum_{j \leq p} f(k_j) \leq f\left(\frac{1}{p} \sum_{j \leq p} k_j\right) = f\left(\frac{i_p}{p}\right) \leq f\left(\frac{n}{p}\right)$$

As a result,

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \frac{p}{n}f\left(\frac{n}{p}\right) + \frac{p}{n}f(|\Sigma|) + \frac{c}{n}$$

where c is some constant independent of u, v, n, p . As $p/n < \alpha < 1/2$ and $x \mapsto xf(1/x)$ is increasing for $x \leq 1/2$, one has:

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \alpha f(1/\alpha) + \alpha f(|\Sigma|) + \frac{c}{n}$$

Switching u and v gives the result (c might be changed). □

We are now able to prove the other inequality.

Proof of $\mathcal{K}_\mu(x, T) \leq \overline{\mathcal{K}}(x, T)$. Fix some computable partition ξ . We show that for any $\beta > 0$ there is some $\epsilon > 0$ such that for every μ -random point x , $\mathcal{K}_\mu(x, T|\xi) \leq \overline{\mathcal{K}}(x, T, \epsilon) + \beta$. As $\overline{\mathcal{K}}(x, T, \epsilon)$ increases as $\epsilon \rightarrow 0^+$ and β is arbitrary, the inequality follows.

First take $0 < \alpha < 1/2$ small enough such that $\alpha f(1/\alpha) + \alpha f(|\xi|) < \beta$, and remark that

$$\lim_{\epsilon \rightarrow 0^+} \mu\left(\overline{(\partial\xi)^\epsilon}\right) = \mu(\partial\xi) = 0$$

Hence there is some ϵ such that $\mu\left(\overline{(\partial\xi)^{2\epsilon}}\right) < \alpha$. From a sequence of ideal points we will reconstruct the symbolic orbit of a random point with a density of errors less than α . Lemma 5.3.1 will then allow to conclude.

We define an algorithm $\mathcal{A}(\epsilon, i_0, \dots, i_{n-1})$ with $\epsilon \in \mathbb{Q}_{>0}$ and $i_0, \dots, i_{n-1} \in \mathbb{N}$ which outputs a word $a_0 \dots a_{n-1}$ on the alphabet ξ . To compute a_j , \mathcal{A} semi-decides in a dovetail picture:

- $s_{i_j} \in C$ for every $C \in \xi$,
- $s \in C$ for every $s \in B(s_{i_j}, \epsilon)$ and every $C \in \xi$.

The first test which stops provides some $C \in \xi$: put $a_j = C$.

Let x be a random point whose iterates are covered by ξ , and $s_{i_0}, \dots, s_{i_{n-1}}$ be ideal points which ϵ -shadow the first n iterates of x . We claim that \mathcal{A} will halt on $(\epsilon, i_0, \dots, i_{n-1})$. Indeed, as $T^j x$ belongs to some $C \in \xi$, $C \cap B(s_{i_j}, \epsilon)$ is a non-empty open set and then contains at least one ideal point s , which will be eventually dealt with.

We now compare the symbolic orbit of x with the symbolic sequence computed by \mathcal{A} . A discrepancy at rank j can appear only if $T^j x \in (\partial\xi)^{2\epsilon}$. Indeed, if $T^j x \notin (\partial\xi)^{2\epsilon}$ then $B(T^j x, 2\epsilon) \subseteq C$ where C is the cell $T^j x$ belongs to. As $d(s_{i_j}, T^j x) < \epsilon$, $B(s_{i_j}, \epsilon) \subseteq B(x, 2\epsilon) \subseteq C$, so the algorithm gives the right cell.

Now, as x is μ -typical by Thm. 3.2.2,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} |\{j < n : T^j x \in (\partial\xi)^{2\epsilon}\}| \leq \mu \left(\overline{(\partial\xi)^{2\epsilon}} \right) < \alpha$$

so there is some n_0 such that for all $n \geq n_0$, $\frac{1}{n} |\{j < n : T^j x \in (\partial\xi)^{2\epsilon}\}| < \alpha$. This implies that for all $n \geq n_0$ and ideal points $s_{i_0}, \dots, s_{i_{n-1}}$ which ϵ -shadow the first n iterates of x and with minimal complexity, the algorithm $\mathcal{A}(\epsilon, i_0, \dots, i_{n-1})$ produces a symbolic string u which differs from the symbolic orbit v of x of length n with a density of errors $< \alpha$. As $K(u) \stackrel{+}{\leq} K(\epsilon) + \mathcal{K}_n(x, T, \epsilon)$ and $\alpha f(1/\alpha) + \alpha f(|\xi|) < \beta$, applying Lem. 5.3.1 gives:

$$\begin{aligned} \frac{1}{n} K(\xi_n(x)) = \frac{1}{n} K(v) &\leq \frac{1}{n} K(u) + \alpha f(1/\alpha) + \alpha f(|\xi|) + \frac{c}{n} \\ &\leq \frac{1}{n} (\mathcal{K}_n(x, T, \epsilon) + K(\epsilon) + c') + \beta + \frac{c}{n} \end{aligned}$$

where c' is independent of n . Taking the lim sup as $n \rightarrow \infty$ gives:

$$\mathcal{K}_\mu(x, T|\xi) \leq \overline{\mathcal{K}}(x, T, \epsilon) + \beta$$

□

6. Entropy vs orbit complexity

In [Bru83] Brudno proved:

Theorem 6.0.2 (Brudno's first theorem). $\mathcal{K}_\mu(x, T|\xi) = h_\mu(T, \xi)$ for μ -almost every point.

Theorem 6.0.3 (Brudno's second theorem). Let X be a compact topological space and $T : X \rightarrow X$ a continuous map.

1. For any ergodic Borel probability measure μ the equality

$$\overline{\mathcal{K}}(x, T) = h_\mu(T)$$

holds for μ -almost all $x \in X$,

2. For all $x \in X$, $\overline{\mathcal{K}}(x, T) \leq h(T)$.

Observe that Brudno did not consider the quantity $\underline{\mathcal{K}}(x, T)$, which was later introduced by White [Whi93], who improved Brudno's second theorem showing that $\underline{\mathcal{K}}(x, T) = h_\mu(T)$ holds for μ -almost all $x \in X$.

First, we show how the algorithmic theory of randomness and information on the space of symbolic sequences provides powerful results that enable one to obtain Brudno's first theorem in an easier way. Then we will strengthen Brudno's two theorems, proving versions for μ -random points. Finally, we will study in more details the relation between the topological quantities $\underline{\mathcal{K}}(x, T)$, $\overline{\mathcal{K}}(x, T)$ and $h(T)$.

6.1. Measure-theoretic entropy

6.1.1. A simple proof of Brudno's first theorem

Thm. 2.6.3 and Prop. 2.6.2 enable one to give tight relations between the algorithmic entropies \mathcal{I}_μ and \mathcal{H}_μ and the Shannon entropies I_μ and H_μ . First let us gather these two inequalities: if $\Sigma^\mathbb{N}$ is endowed with a computable probability measure ν , then for all $\omega \in \Sigma^\mathbb{N}$,

$$-\log \nu[\omega_{0..n-1}] - d_\nu(\omega) \leq K(\omega_{0..n-1}) \stackrel{\pm}{\leq} -\log \nu[\omega_{0..n-1}] + K(n) \quad (6)$$

where d_ν is the deficiency of randomness, which satisfies $\int_{\Sigma^\mathbb{N}} d_\nu d\nu < 1$ and is finite exactly on ν -random sequences (the constant in $\stackrel{\pm}{\leq}$ does not depend on ω and n , see Sect. 2.6.1).

Now we show how to obtain Brudno's theorem from (6). Applying it to $\nu = \mu_\xi$ directly gives:

$$I_\mu(\cdot | \xi_n) - d_\mu \circ \phi_\xi \leq \mathcal{I}(\cdot | \xi_n) \stackrel{\pm}{\leq} I_\mu(\cdot | \xi_n) + K(n) \quad (7)$$

where it is defined (almost everywhere, at least on random points). Every μ -random point x is mapped by ϕ_ξ to a μ_ξ -random sequence (see Prop. 2.5.1), whose randomness deficiency is finite. It then follows that the point-wise entropies using Shannon information I_μ and Kolmogorov-Chaitin information \mathcal{I}_μ coincide on μ -random points:

Proposition 6.1.1. *For every μ -random point x ,*

$$\mathcal{K}_\mu(x, T | \xi) := \limsup_n \frac{1}{n} \mathcal{I}_\mu(x | \xi_n) = \limsup_n \frac{1}{n} I_\mu(x | \xi_n). \quad (8)$$

This equality together with the Shannon-McMillan-Breiman theorem (5) gives directly Brudno's theorem (Thm. 6.0.2). Hence, as the collection of computable partitions is generating (see Cor. 3.1.1 and Prop. 4.1.2) and countable, taking the supremum over computable partitions gives $\mathcal{K}_\mu(x, T) = h_\mu(T)$ for μ -almost every x . We will strengthen this in the next section, proving that it holds for all μ -random points.

Remark 6.1.1. The Kolmogorov-Sinai entropy, originally expressed using Shannon entropy, can be expressed using algorithmic entropy. Indeed, taking the mean in (7), one obtains:

$$H_\mu(\xi_n) - 1 \leq \mathcal{H}_\mu(\xi_n) \stackrel{\pm}{\leq} H_\mu(\xi_n) + K(n),$$

so

$$h_\mu(T, \xi) = \lim_n \frac{H_\mu(\xi_n)}{n} = \lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n}.$$

Again, as the collection of computable partitions is generating the Kolmogorov-Sinai entropy of (X, μ, T) can be characterized by:

$$h_\mu(T) = \sup \left\{ \lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n} : \xi \text{ finite computable partition} \right\}.$$

6.1.2. Brudno's theorems for random points

On the Cantor space, V'yugin [V'y98] and later Nakamura [Nak05] proved a slightly weaker version of the Shannon-McMillan-Breiman for Martin-Löf random sequences. In particular, we will use:

Theorem 6.1.1 (V'yugin). *Let μ be a computable shift-invariant ergodic measure on $\Sigma^\mathbb{N}$. Then, for any μ -random sequence ω ,*

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mu([\omega_{0..n-1}]) = h_\mu(\sigma).$$

Note that it is not known yet if the limit exists for all random sequences. Using effective symbolic models, this can be easily extended to any computable probability space.

Corollary 6.1.1 (Shannon-McMillan-Breiman for random points). *Let T be an ergodic endomorphism of the computable probability space (X, μ) , and ξ a computable partition. For every μ -random point x ,*

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mu(\xi_n(x)) = h_\mu(T, \xi).$$

Proof. Since ξ is computable, the symbolic model $(\{1, \dots, k\}^\mathbb{N}, \mu_\xi, \sigma)$ is effective. Every μ -random point x is mapped to a μ_ξ -random sequence ω , for which the preceding theorem holds. Using the facts that $\mu(\xi_n(x)) = \mu_\xi([\omega_{0..n-1}])$ and $h_\mu(T, \xi) = h_{\mu_\xi}(\sigma)$ allows to conclude. \square

Finally, this implies our first announced result:

Theorem 6.1.2. *Let T be an ergodic endomorphism of the computable probability space (X, μ) , and ξ be a computable partition. For every μ -random point x :*

$$\begin{aligned} \mathcal{K}_\mu(x, T|\xi) &= h_\mu(T, \xi), \\ \mathcal{K}_\mu(x, T) &= h_\mu(T). \end{aligned}$$

Proof. We combine equality (8) and Cor. 6.1.1: for every random point x , $\mathcal{K}_\mu(x, T|\xi) = \limsup_n \frac{1}{n} I_\mu(x|\xi_n) = h_\mu(T, \xi)$. Since the collection of all computable partitions generates the Borel σ -field (Cor. 3.1.1), $\mathcal{K}_\mu(x, T) = \sup\{h_\mu(T, \xi) : \xi \text{ computable partition}\} = h_\mu(T)$ (Prop. 4.1.2). \square

Combining Thms. 5.3.1 and 6.1.2, we obtain a version of Brudno's second theorem (Thm. 6.0.3) for Martin-Löf random points.

Corollary 6.1.2. *Let T be an ergodic endomorphism of the computable probability space (X, μ) , where X is compact. Then for every μ -random point x :*

$$\overline{\mathcal{K}}(x, T) = h_\mu(T)$$

6.2. Topological entropy

Now we prove:

Theorem 6.2.1 (Topological entropy vs orbit complexity). *Let X be a compact computable metric space, and $T : X \rightarrow X$ a computable map. Then*

$$h(T) = \sup_{x \in X} \underline{\mathcal{K}}(x, T) = \sup_{x \in X} \overline{\mathcal{K}}(x, T).$$

In order to prove this theorem, we define an effective version of the topological entropy, which is strongly related to the complexity of orbits. To do this, let us give first a simple characterization of topological entropy which will accommodate to effectivisation.

Definition 6.2.1. A *null s -cover* of $Y \subseteq X$ is a set $E \subseteq \mathbb{N}^3$ such that:

1. $\sum_{(i,n,p) \in E} 2^{-sn} < \infty$,
2. for each $k, p \in \mathbb{N}$, the set $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E, n \geq k\}$ is a cover of Y .

The idea is simple: every null s -cover induces open covers of arbitrary small size and arbitrary small weight. Remark that any null s -cover of Y is also a null s' -cover for all $s' > s$.

Lemma 6.2.1. $h_2(T, Y) = \inf\{s : Y \text{ has a null } s\text{-cover}\}$.

Proof. Suppose $s > h_2(T, Y)$. We fix $p, k \in \mathbb{N}$ and put $\epsilon = 2^{-p}$ and $\delta = 2^{-k}$. As $m_\delta^s(Y, \epsilon) = 0$, there is a cover $(U_{j,k,p})_j$ of Y by open sets of ϵ -size $\delta_{j,k,p} < \delta$ with $\sum_j \delta_{j,k,p}^s < 2^{-(k+p)}$. Let s_i be any ideal point in $U_{j,k,p}$. If $\delta_{j,k,p} > 0$, then $\delta_{j,k,p} = 2^{-n}$ for some $n > k$. If $\delta_{j,k,p} = 0$, take any $n \geq \max\{k, (j+k+p)/s\}$. In both cases, $U_{j,k,p}$ is included in the Bowen ball $B_n(s_i, \epsilon)$. We define $E_{k,p}$ as the set of (i, n, p) obtained this way, and $E = \bigcup_{k,p} E_{k,p}$. By construction, for each k, p , $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E, n \geq k\}$ is a cover of Y as it contains $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E_{k,p}\}$. Moreover, $\sum_{(i,n,p) \in E_{k,p}} 2^{-sn} \leq \sum_j \delta_{j,k,p}^s + \sum_j 2^{-(j+k+p)} \leq 2^{-(k+p)+2}$, so $\sum_{(i,n,p) \in E} 2^{-sn} < \infty$.

Conversely, if Y has a null s -cover E , take $\epsilon, \delta > 0$ and p, k such that $\epsilon > 2^{-p+1}$ and $\delta > 2^{-k}$. For all $k' \geq k$, the family $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E, n \geq k'\}$ is a cover of Y by open sets of ϵ -size at most $2^{-n} \leq \delta$. Moreover, $\sum_{(i,n,p) \in E, n \geq k'} 2^{-sn}$ tends to 0 as k' grows, so $m_\delta^s(Y, \epsilon) = 0$. It follows that $s \geq h_2(T, Y)$. \square

By an *effective* null s -cover, we mean a null s -cover E which is a r.e. subset of \mathbb{N}^3 .

Definition 6.2.2. The *effective topological entropy* of T on Y is defined by

$$h_2^{\text{eff}}(T, Y) = \inf\{s : Y \text{ has an effective null } s\text{-cover}\}$$

As less null s -covers are allowed in the effective version, $h_2(T, Y) \leq h_2^{\text{eff}}(T, Y)$. Of course, if $Y \subseteq Y'$ then $h_2^{\text{eff}}(T, Y) \leq h_2^{\text{eff}}(T, Y')$. We now prove:

Theorem 6.2.2 (Effective topological entropy vs lower orbit complexity). *Let X be a computable metric space and $T : X \rightarrow X$ a continuous map. For all $Y \subseteq X$,*

$$h_2^{\text{eff}}(T, Y) = \sup_{x \in Y} \underline{\mathcal{K}}(x, T)$$

which implies in particular that $h_2^{\text{eff}}(T, \{x\}) = \underline{\mathcal{K}}(x, T)$: the restriction of the system to a single orbit may have *positive* effective topological entropy.

This kind of result has already been obtained for the Hausdorff dimension of subsets of the Cantor space, proving that the effective dimension of a set A is the supremum of the lower growth-rate of Kolmogorov-Chaitin complexity of sequences in A (which corresponds to Thm. 6.2.2 for sub-shifts). This remarkable property is a counterpart of the countable stability property of Hausdorff dimension ($\dim Y = \sup_i \dim Y_i$ when $\bigcup_i Y_i = Y$) (see [CH94, May01, Lut03, Rei04, Sta05]).

Theorem 6.2.2 is a direct consequence of the two following lemmas.

Lemma 6.2.2. *Let $\alpha \geq 0$ and $Y_\alpha = \{x : \underline{\mathcal{K}}(x, T) \leq \alpha\}$. One has $h_2^{\text{eff}}(T, Y_\alpha) \leq \alpha$.*

Proof. Let $\beta > \alpha$ be a rational number. We define the r.e. set $E = \{(i, n, p) : K(i, n, p) < \beta n\}$. Let $p \in \mathbb{N}$ and $\epsilon = 2^{-p}$. If $x \in Y_\alpha$ then $\underline{\mathcal{K}}(x, T, \epsilon) \leq \alpha < \beta$ so for infinitely many n , there is some s_i such that $x \in B_n(s_i, \epsilon)$ and $K(i, n, p) < \beta n$. So for all k , $\{B_n(s_i, 2^{-p}) : (i, n, p) \in E, n \geq k\}$ covers Y_α . Moreover, $\sum_{(i, n, p) \in E} 2^{-\beta n} \leq \sum_{(i, n, p) \in E} 2^{-K(i, n, p)} \leq 1$.

E is then an effective null β -cover of Y_α , so $h_2^{\text{eff}}(T, Y_\alpha) \leq \beta$. And this is true for every rational $\beta > \alpha$. \square

Lemma 6.2.3. *Let $Y \subseteq X$. For all $x \in Y$, $\underline{\mathcal{K}}(x, T) \leq h_2^{\text{eff}}(T, Y)$.*

Proof. Let $s > h_2^{\text{eff}}(T, Y)$: Y has an effective null s -cover E . As $\sum_{(i, n, p) \in E} 2^{-sn} < \infty$, by the coding theorem $K(i, n, p) \leq sn + c$ for some constant c , which does not depend on i, n, p . If $x \in Y$, then for each p, k , x is in a ball $B_n(s_i, 2^{-p})$ for some $n \geq k$ with $(i, n, p) \in E$. Then $\mathcal{K}_n(x, T, 2^{-p}) \leq sn + c$ for infinitely many n , so $\underline{\mathcal{K}}(x, T, 2^{-p}) \leq s$. As this is true for all p , $\underline{\mathcal{K}}(x, T) \leq s$. As this is true for all $s > h_2^{\text{eff}}(T, Y)$, we can conclude. \square

Proof of Thm. 6.2.2. By Lem. 6.2.3, $\alpha := \sup_{x \in Y} \underline{\mathcal{K}}(x, T) \leq h_2^{\text{eff}}(T, Y)$. Now, as $Y \subseteq Y_\alpha$, $h_2^{\text{eff}}(T, Y) \leq h_2^{\text{eff}}(T, Y_\alpha) \leq \alpha$ by Lem. 6.2.2. \square

The definition of an effective null α -cover involves a summable computable sequence. The universality of the sequence $2^{-K(i)}$ among summable lower semi-computable sequences is at the core of the proof of the preceding theorem, which states that there is a universal

effective null α -cover, for every $\alpha \geq 0$. In other words, there is a maximal set of effective topological entropy $\leq \alpha$, and this set is $Y_\alpha = \{x \in X : \underline{\mathcal{K}}(x, T) \leq \alpha\}$.

The definition of the topological entropy as a capacity could be also made effective, restricting to effective covers. Classical capacity does not share with Hausdorff dimension the countable stability. For the same reason, its effective version is not related with the orbit complexity as strongly as the effective topological entropy is. Nevertheless, a weaker relation holds, which is sufficient for our purpose: the upper complexity of orbits is bounded by the effective capacity. We do not develop this and only state the needed property (which implicitly uses the fact that the effective capacity coincides with the classical capacity for a compact computable metric space):

Lemma 6.2.4. *Let X be a compact computable metric space, and $T : X \rightarrow X$ a computable map. For all $x \in X$, $\overline{\mathcal{K}}(x, T) \leq h_1(T, X)$.*

Proof. We first construct a r.e. set $E \subseteq \mathbb{N}^3$ such that for each n, p , $\{s_i : (i, n, p) \in E\}$ is a $(n, 2^{-p})$ -spanning set and a $(n, 2^{-p-2})$ -separated set. Let us fix n and p and enumerate $E_{n,p} = \{i : (i, n, p) \in E\}$, in a uniform way. The algorithm starts with $S = \emptyset$ and $i = 0$. At step i it analyzes s_i and decides to add it to S or not, and goes to step $i + 1$. $E_{n,p}$ is the set of points which are eventually added to S .

Step i for each ideal point $s \in S$, test in parallel $d_n(s_i, s) < 2^{-p-1}$ and $d_n(s_i, s) > 2^{-p-2}$: at least one of them must stop. If the first one stops first, reject s_i and go to Step $i + 1$. If the second one stops first, go on with the other points $s \in S$: if all S has been considered, then add s_i to S and go to Step $i + 1$.

By construction, the set of selected ideal points forms a $(n, 2^{-p-2})$ -separated set. If there is $x \in X$ which is at distance at least 2^{-p} from every selected point, then let s_i be an ideal point s_i with $d_n(x, s_i) < 2^{-p-1}$: s_i is at distance at least 2^{-p-1} from every selected point, so at step i it must have been selected, as the first test could not stop. This is a contradiction: the selected points form a $(n, 2^{-p})$ -spanning set.

From the properties of $E_{n,p}$ it follows that $N(X, n, 2^{-p}) \leq |E_{n,p}| \leq M(X, n, 2^{-p-2})$, and then

$$\sup_p \left(\limsup \frac{1}{n} \log |E_{n,p}| \right) = h_1(T, X)$$

If $\beta > h_1(T, X)$ is a rational number, then for each p , there is $k \in \mathbb{N}$ such that $\log |E_{n,p}| < \beta n$ for all $n \geq k$.

Now, for $s_i \in E_{n,p}$, $K(i) \stackrel{\pm}{\leq} \log |E_{n,p}| + 2 \log \log |E_{n,p}| + K(n, p)$ by Prop. 2.6.1. Take $x \in X$: x is in some $B_n(s_i, 2^{-p})$ for each n , so $\overline{\mathcal{K}}(x, T, 2^{-p}) \leq \limsup_n \frac{1}{n} \log |E_{n,p}| \leq \beta$ as $\log |E_{n,p}| < \beta n$ for all $n \geq k$. As this is true for all p and all $\beta > h_1(T, X)$, $\overline{\mathcal{K}}(x, T) \leq h_1(T, X)$ and this for all $x \in X$. \square

We are now able to prove Thm. 6.2.1. Combining the several results established above:

$$h_1(T, X) = h_2(T, X) \leq h_2^{\text{eff}}(T, X) = \sup_{x \in X} \underline{\mathcal{K}}(x, T) \leq \sup_{x \in X} \overline{\mathcal{K}}(x, T) \leq h_1(T, X)$$

(Thm. 4.2.1) (Thm. 6.2.2) (Lem. 6.2.4)

and the statement is proved.

References

- [BBG⁺04] V. Benci, C. Bonanno, S. Galatolo, G. Menconi, and M. Virgilio. Dynamical systems and computable information. *Disc. Cont. Dyn. Syst.-B*, 4:935–960, 2004.
- [BHW08] Vasco Brattka, Peter Hertling, and Klaus Weihrauch. A tutorial on computable analysis. In *Changing Conceptions of What is Computable*, pages 425 – 492, Orlando, FL, USA, 2008. Springer-Verlag.
- [Bil65] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley, New York, 1965.
- [Bil68] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1968.
- [Bos08a] Volker Bosserhoff. Notions of probabilistic computability on represented spaces. In *CCA 2007, R. Dillhage et al (eds.,* pages 45 – 78. Informatik Berichte 338, Fernuniversitaet Hagen, 2008.
- [Bos08b] Volker Bosserhoff. Notions of probabilistic computability on represented spaces. *J.UCS*, 14(6):965–995, 2008.
- [Bru83] A.A. Brudno. Entropy and the complexity of the trajectories of a dynamical system. *Transactions of the Moscow Mathematical Society*, 2:127–151, 1983.
- [CH94] Jin-Yi Cai and Juris Hartmanis. On Hausdorff and topological dimensions of the Kolmogorov complexity of the real line. In *Proceedings of the 30th IEEE symposium on Foundations of computer science*, pages 605–619, Orlando, FL, USA, 1994. Academic Press, Inc.
- [Gác] Peter Gács. Lecture notes on descriptive complexity and randomness. <http://www.cs.bu.edu/faculty/gacs/papers/ait-notes.pdf>.
- [Gác05] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005.
- [Gal00] Stefano Galatolo. Orbit complexity by computable structures. *Nonlinearity*, 13:1531–1546, 2000.
- [GHR09a] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Dynamics and abstract computability: computing invariant measures, 2009. arXiv:0903.2385.
- [GHR09b] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. An effective Borel-Cantelli lemma. Constructing orbits with required statistical properties. *Theoretical Computer Science*, 2009. In press, arXiv:0711.1478.

- [HK95] Boris Hasselblatt and Anatole Katok. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1995.
- [HK02] Boris Hasselblatt and Anatole Katok. *Handbook of Dynamical Systems*, volume 1A, chapter Principle Structures. Elsevier, 2002.
- [HR09a] Mathieu Hoyrup and Cristóbal Rojas. Applications of effective probability theory to martin-löf randomness. In *Proceedings of ICALP*, 2009.
- [HR09b] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 2009. In press, arXiv:0709.0907.
- [HW03] Peter Hertling and Klaus Weihrauch. Random elements in effective topological spaces with measure. *Information and Computation*, 181(1):32–56, 2003.
- [Ken08] Robert Kenny. *Orbit complexity and computable Markov partitions*. PhD Thesis. University of Western Australia, 2008.
- [Lut03] Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.
- [LV93] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993.
- [May01] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Electronic Colloquium on Computational Complexity (ECCC)*, 8(059), 2001.
- [Nak05] Masahiro Nakamura. Ergodic theorems for algorithmically random sequences. *Proceedings of the Symposium on Information Theory and Its Applications*, 2005.
- [Nan08] Satyadev Nandakumar. An effective ergodic theorem and some applications. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 39–44, New York, NY, USA, 2008. ACM.
- [Pes98] Yakov B. Pesin. *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*. Chicago Lectures in Mathematics. The University of Chicago Press, 1998.
- [Pet83] Karl Petersen. *Ergodic Theory*. Cambridge Univ. Press, 1983.
- [Rei04] Jan Reimann. *Computability and fractal dimension*. PhD thesis, Ruprecht-Karls Universität Heidelberg, 2004.
- [Rog87] Hartley Jr. Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, Cambridge, MA, USA, 1987.

- [Sta05] Ludwig Staiger. Constructive dimension equals Kolmogorov complexity. *Information Processing Letters*, 93(3):149–153, 2005.
- [V’y97] Vladimir V. V’yugin. Effective convergence in probability and an ergodic theorem for individual random sequences. *SIAM Theory of Probability and Its Applications*, 42(1):39–50, 1997.
- [V’y98] Vladimir V. V’yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(4):343–361, 1998.
- [Wal82] Peter Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.
- [Wei93] Klaus Weihrauch. Computability on computable metric spaces. *Theoretical Computer Science*, 113(2):191–210, 1993.
- [Wei00] Klaus Weihrauch. *Computable Analysis*. Springer, Berlin, 2000.
- [Whi93] Homer S. White. Algorithmic complexity of points in dynamical systems. *Ergodic Theory and Dynamical Systems*, 13:807–830, 1993.
- [You02] Lai-Sang Young. What are srb measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.