Effective Kolmogorov-Sinai and Topological Entropy

Satyadev Nandakumar

Indian Institute of Technology Kanpur

September 9, 2014

Introduction

Kolmogorov's Programme:

"The application of probability theory can be put on a uniform basis. It is always a matter of hypotheses about the impossibility of reducing in one way or another the complexity of the description of objects in question."

Consider theorems in Probability theory which hold "almost everywhere". Can we show that if an object has maximum descriptional complexity, (i.e. is "random"), then it obeys the theorem?

Kolmogorov Theme

$$\begin{tikzcd}
\text{Computability Theory} \arrow[d, "\text{Compressibility}"] \\
\text{Information Theory} \arrow[d, "\text{Entropy}"] \\
\text{Dynamical Systems}
\end{tikzcd}$$

We have a theory of individual random sequences (e.g. Martin-Löf randomness), and their statistical properties

- e.g. in every random infinite binary sequence, under the uniform measure, 0 appears with an asymptotic frequency of $1/2$.

Symbolic Dynamics associates general dynamical systems with a shift on Σ^∞, $|\Sigma| < \infty$, with similar statistical properties.

Allows us to generalize results on infinite sequences on finite alphabets, to general metric spaces.

Kolmogorov-Sinai and Topological Entropies

Entropy in Symbolic Dynamics

There are two broad settings in which we study the notion on entropy in dynamical systems.

- ☐ The Topological Setting: (X, T) where X is a compact space, and $T : X \to X$ is a continuous transformation.
- ☐ The Measure-theoretic Setting: $(X, \mathcal{F}, \mu, T)$ where $(X, \mathcal{F}, \mu)$ is a probability space, and $T : X \to X$ is a *measure-preserving transformation*.

Associated with either, we define an analogue of Shannon entropy - the topological entropy and the Kolmogorov-Sinai entropy, respectively. (These are also related!)

Why entropy is important:

1. It gives us a way to quantify the "disorder" or "information" in a system.
2. In symbolic dynamics, it provides a way to "classify" systems. For example, "closely related" systems should have the "same entropy".

Definition 1. Let $(X, \mathcal{F}, P)$ be a probability space.

A measurable transformation $T : X \to X$ is called *measure-preserving* if for every $A \in \mathcal{F}$, $P(T^{-1}A) = P(A)$.

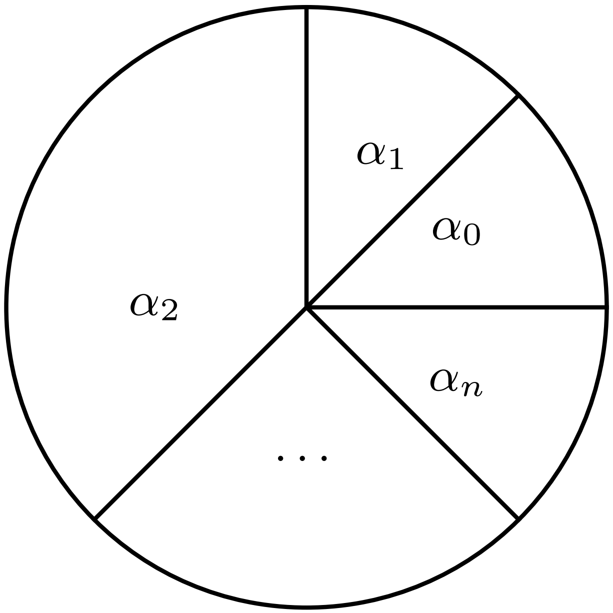
A measure-preserving map T is *ergodic* if for all $A \in \mathcal{F}$, $TA = A$ only when $P(A) \in \{0, 1\}$.

Example. If X is a finite set with the uniform distribution on it, then every permutation is a measure-preserving transformation.

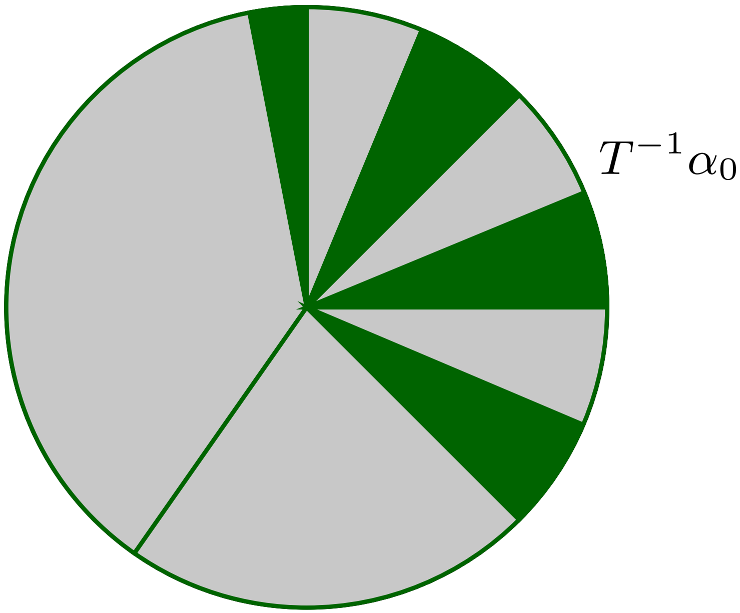
Any permutation consisting of a single cycle is an ergodic transformation.

Definition 2. A system $(X, \mathcal{F}, P, T)$ where $(X, \mathcal{F}, P)$ is a probability space and T is measure-preserving with respect to it, is called a *dynamical system*.

Partitions



Partitions



Kolmogorov-Sinai Entropy

The entropy of a partition $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$ of X is

$$H(\alpha) = \sum_{i=0}^{n-1} P(\alpha_i) \log \left(\frac{1}{P(\alpha_i)} \right).$$

The k-step entropy is

$$h_k(\alpha, T) = \frac{H\left(\alpha \vee \cdots \vee T^{-k+1}\alpha\right)}{k}.$$

The entropy of a transformation T wrt α is

$$h(\alpha, T) = \lim_{k \to \infty} h_k(\alpha, T).$$

The entropy of a transformation T is

$$h(T) = \sup\{h(\alpha, T) \mid \alpha \text{ is a finite partition of } X\}.$$

Definition 3. [AKM65] Let X be a compact Hausdorff space, and let C be a finite open cover.

Let $H(C)$ be the logarithm of the cardinality of the smallest subset of C which covers X.

For two covers C and D, let $C \vee D$ be their minimal common refinement.

For any continuous map $f : X \to X$, the following limit exists:

$$H(C, f) = \lim_{n\to\infty} \frac{1}{n} H(C \vee f^{-1}C \vee \cdots \vee f^{-n}(C).$$

The topological entropy of f, is

$$h(f) = \sup_C H(f, C).$$

If (X, f) is a topological dynamical system, and $(X, \mathcal{F}, \mu, f)$ is a measure-preserving dynamical system, the following holds.

Theorem 4. *[Din70], Goodman 1971*

$$h(f) = \sup\{h_\mu(f) \mid \mu \textit{ is an invariant Borel measure }\}.$$

Effective Symbolic Dynamics

We now explain some results in *effective* symbolic dynamics.

We focus on two things.

- ☐ Definitions of entropy using effective notions of complexities of orbits.
- ☐ Classification results.

We first need definitions. Idea: Define the complexity of the *orbits* of points in the space.

	Effective Topological Space	Effective Probability Space
Effective Definition	Brudno, White Galotolo, Hoyrup, Rojas	Galotolo, Hoyrup, Rojas
Computability	?	ess. follows from Kolmogorov-Sinai theorem

A few currently established results.

- Effective Ergodic Theorem for effective symbolic dynamical systems. (Galotolo, Hoyrup, Rojas 2010)
- An effective variational principle. (GHR10, Simpson 2011)
- A Kolmogorov-Complexity proof of the Kolmogorov-Sinai Theorem (Day 2014).
- Effective Ornstein Isomorphism (Ghosh, Nandakumar, Pal 2014.)

Effective entropies via Orbit Complexities

Computable Probability Spaces

Definition 5. A *computable metric space* is a triple (X, d, S) where (X, d) is a complete separable metric space, $S = \langle s_i \rangle_{i \in \mathbb{N}}$ is a recursively enumerable dense set of points in X (called *ideal points*) and $d(s_i, s_j)$ are computable, uniformly in i, j.

Elements of the set $\{B(s_i, q_j) \mid s_i \in S, q_j \in \mathbb{Q}\}$ are called *ideal balls*.

Let (X, d_X, S_X) and (Y, d_Y, S_Y) be computable metric spaces. A function $T : X \to Y$ is called *computable* if $T^{-1}(B_n)$ is recursively enumerably open, uniformly in n.

Definition 6. A measure μ on X is computable if the measure of finite unions of ideal open balls is lower semicomputable.

Definition 7. A *computable probability space* is a pair $(\mathcal{X}, \mu)$ where $\mathcal{X}$ is a computable metric space and μ is a computable probability measure.

Effective Symbolic Dynamics

Let (X, μ, T) be a dynamical system, and $\alpha = \{a_1, \ldots, a_k\}$ be a finite measurable partition.

The associated model $(X_\alpha, \mu_\alpha, \sigma)$ is said to be an *effective symbolic model* if the map $\phi_\alpha : X \to \{1, \ldots, k\}^{\mathbb{N}}$ is a measure-preserving function defined on a constructive G_δ set of measure 1.

These can also be defined in terms of *computable partitions* - where every atom p_i has two open sets U and V such that $U \subseteq A$, $V \subseteq p_i^c$, and $\mu(U) + \mu(V) = 1$.

Orbit Complexity - Algorithmic Entropy Viewpoint: "Label" complexity

An atom of the partition $\alpha^{(n)} = \alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-n+1}\alpha$ can be seen as an n-length string on the alphabet α.
For $x \in X$, its *Kolmogorov information* relative to the partition $\alpha^{(n)}$ is

$$\mathcal{K}\left(\alpha^{(n)}(x)\right).$$

Note that this is independent of μ.

The *algorithmic entropy* of $\alpha^{(n)}$ is defined as

$$h_\mu\left(\alpha^{(n)}\right) = \sum_{w \in \alpha^{(n)}} \mu(w)\mathcal{K}(w).$$

Define

$$\mathcal{K}_\mu(x, T \mid \alpha) = \limsup_n \frac{-\log \mu\left(\alpha^{(n)}(x)\right)}{n},$$

and

$$\mathcal{K}_\mu(x, T) = \sup\left\{\mathcal{K}_\mu(x, T \mid \alpha) \mid \alpha \text{ is a computable partition}\right\},$$

as the *symbolic orbit complexity* of x under T. (This is a "label complexity".)

Correspondence between the views

Theorem 8. *[GHR10]*

$$K_\mu(x, T \mid \alpha) = h_\mu(T, \alpha)$$

for every μ-random x.

The result follows from the fact that

$$-\log \mu(\omega[0 \ldots n-1]) - d_\mu(\omega) \leq K(\omega[0 \ldots n-1]) \leq -\log \mu(\omega[0 \ldots n-1]) + K(n).$$

(Brudno had an a.e. version of the above theorem.)

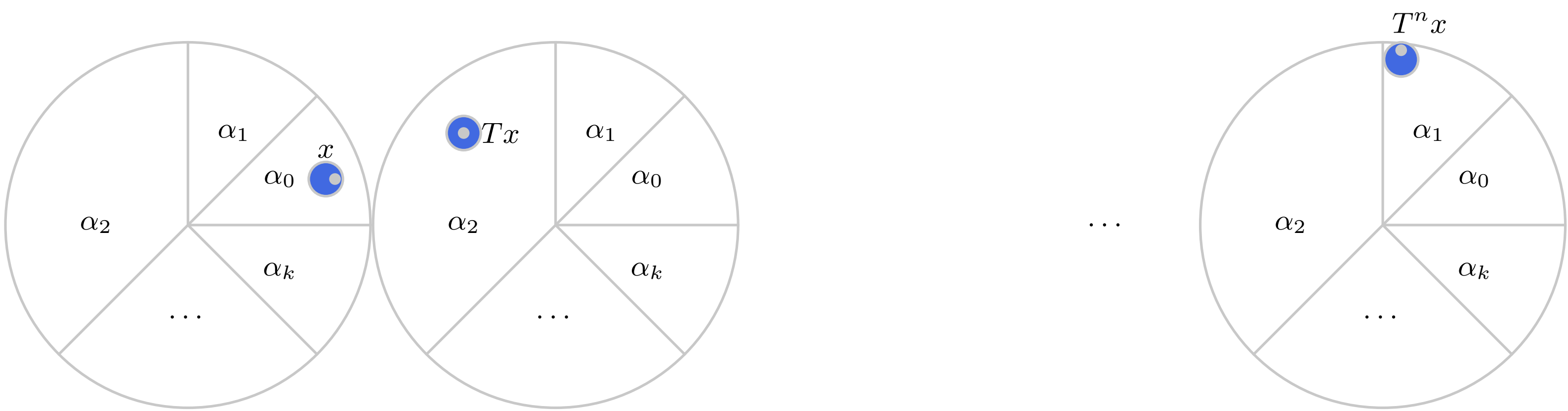
Moreover, we also have :

Theorem 9. *[GHR10] Let (X, μ) be a computable probability space and $T : X \to X$ be a computable ergodic transformation. For every μ-random point x, we have*

$$K_\mu(x, T) = h_\mu(T).$$

The result follows from the effective Shannon-McMillan-Breiman Theorem [Hoc09], [Hoy12], and the fact that the collection of all computable partitions generates the Borel σ-field.

Orbit Complexity for Computable Probability Spaces



"Shadows" for $\mathsf{OC}_n(x, T, \epsilon)$

Given $\epsilon > 0$, the algorithmic information of a sequence of ideal points which shadow the orbit of x for at least n steps is:

$$\mathsf{OC}_n(x, T, \epsilon) = \min\{K(i_0, \ldots, i_{n-1}) \mid d(s_{i_j}, T^j(x) < \epsilon\}.$$

Then its maximal growth rate is

$$\overline{\mathsf{OC}}(x, T, \epsilon) = \limsup_{n\to\infty} \frac{1}{n}\mathsf{OC}_n(x, T, \epsilon).$$

As ϵ decreases to zero, the maximal growth rate does not decrease. Hence we define

$$\overline{\mathsf{OC}}(K, T) = \lim_{\epsilon\to 0^+} \overline{\mathsf{OC}}(K, T, \epsilon).$$

Theorem 10. *Let (X, μ) be a computable* **compact** *probability space, and $T : X \to X$ be a computable ergodic transformation. Then for every μ-random x,*

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

(follows from an argument on reconstruction of typical orbits from ideal “shadows”.)

Theorem 11. *[Effective Poincare Recurrence] Let* (X, μ) *be a computable probability space and* T *be a computable measure-preserving transformation. Then every* μ*-random* x *is recurrent - that is,*

$$\liminf_{n} d(x, T^n x) = 0.$$

Theorem 12. *Let* $(\mathcal{X}, \mu)$ *be a computable probability space, and* $T : X \to X$ *be an ergodic transformation. Then for every continuous bounded function* f *and every* μ*-random* x*,*

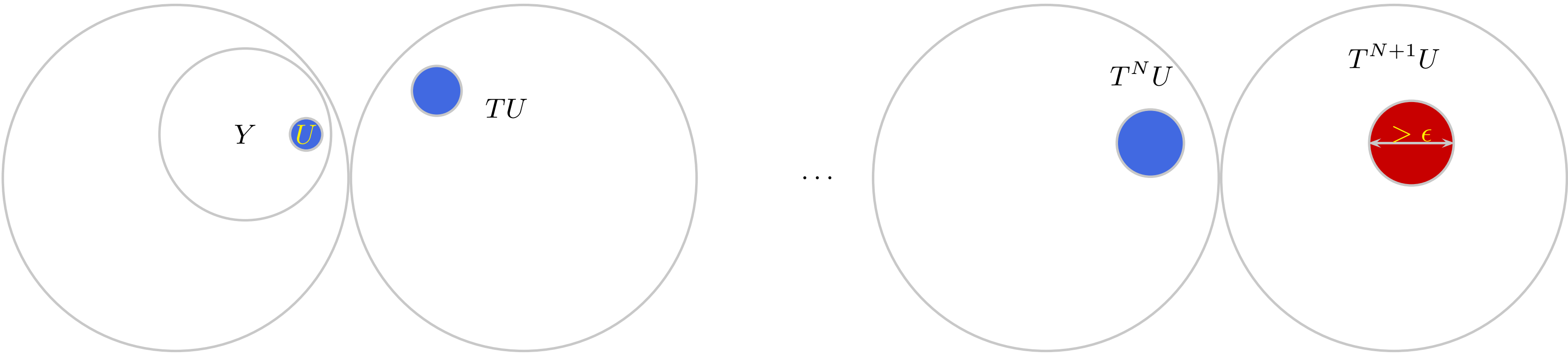
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int f d\mu.$$

Effective Topological Entropy

Topological Entropy - Bowen Balls

To effectivize topological entropy, [GHR10] deals with three equivalent definitions. An alternate characterization of topological entropy that we use to establish a characterization which can be interpreted as the orbit complexity of a point.

Topological Orbit Complexity



The ϵ-size of U in a cover of Y is N.

Topological Entropy - Dimension viewpoint

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

$$N = \sup \left\{ n \geq 0 \mid \text{diameter} \left(T^i E \right) < \epsilon, \text{ for all } 0 \leq i < n \right\}.$$

Define, for $s, \delta, \epsilon \in \mathbb{R}^+$,

$$m^s_\delta(Y, \epsilon) = \inf \left\{ \sum_{U \in \mathcal{G}} (\epsilon - size(U))^s \mid \mathcal{G} \text{ is a countable cover of } Y \text{ by open sets of } \epsilon - \text{size } < \delta \right\}.$$

As δ decreases to 0, the above quantity increases, hence define

$$m^s(Y, \epsilon) = \lim_{\delta \to 0^+} m^s_\delta(Y, \epsilon).$$

There is a critical value for s, below which $m^s_\delta(Y, \epsilon)$ is infinite and above which it is zero.
Define

$$h_{\text{top}}(T, Y, \epsilon) = \inf \left\{ s \mid m^s(Y, \epsilon) = 0 \right\}.$$

As fewer covers are admissible when ϵ decreases to 0, the following limit exists, and is equal to the topological entropy when Y is compact.

$$h_{\text{top}}(T, Y) = \lim_{\epsilon \to 0^+} h_{\text{top}}(T, Y)(T, Y, \epsilon).$$

Theorem 13. *Let X be a computable* **compact** *metric space, and $T : X \to X$ be a computable map. Then,*

$$h_{top}(T, X) = \sup_{x \in X} \overline{OC}(x, T).$$

A third, equivalent, definition of topological entropy will be useful in effectivizing the notion.

Definition 14. A *null s-cover* of $Y \subseteq X$ is a set of triples of natural numbers E such that

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

Definition 15. The *effective topological entropy* of T on Y is defined by

$$h^{\text{eff}}_{\text{top}}(T, Y) = \inf\{s \mid Y \text{ has an effective } s - \text{null cover.}\}.$$

Theorem 16. *[GHR10] Let X be a computable compact metric space, and $T : X \to X$ be a computable map. Then*

$$h_{top}(T) = \sup_{x\in X} \overline{OC}(x, T) = \sup_{x\in X} \underline{OC}(x, T).$$

Effective Ornstein Theorem

The partition α of X is called a *generator* if the σ-algebra on X is generated by $\cdots \vee T^{-2}\alpha \vee T^{-1}\alpha \vee \alpha \vee T\alpha \vee T^2\alpha \ldots$.

Theorem 17. *If α is a generator, then $h(\alpha, T) = h(T)$.*

(α is a "natural" partition induced by T.)

Definition 18. An *isomorphism* $\phi : A \to B$ is a function such that $\phi \circ T_A = T_B \circ \phi$.

Theorem 19. *[Kolmogorov 64, Sinai 68] If two dynamical systems are isomorphic to each other, then they have the same Kolmogorov-Sinai entropy.*

Converse of the KS theorem

Let Σ_A and Σ_B be two finite alphabets.

Let $A = (\Sigma_A^\infty, \mathcal{B}(\Sigma_A^\infty), P_A, T_A)$ and $B = (\Sigma_B^\infty, \mathcal{B}(\Sigma_B^\infty), P_B, T_B)$ be two Bernoulli systems with the same KS entropy.

Are the two systems necessarily isomorphic?

(Note: Σ_A and Σ_B need not have the same cardinality.)

Answer: Yes [Orn70]. In fact, there is a finitary isomorphism between them [KS79].

Setting

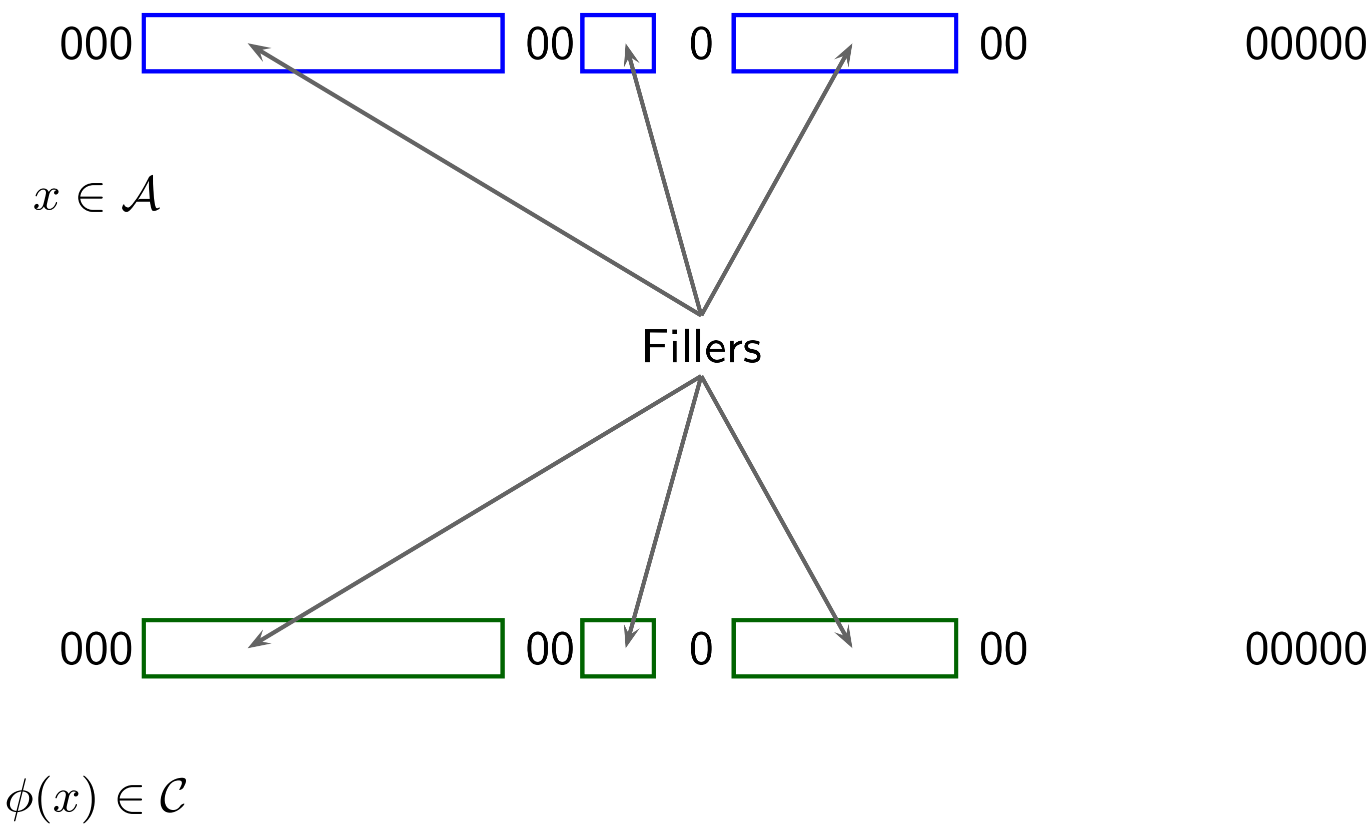
The finite portions $x[\,-m\ldots 0\ldots m\,]$ of an infinite sequence x are the *cylinders* of x.

A *finitary* map $\phi : A \to B$ is one where for every $x \in A$ such that $\phi(x)$ is defined, there is an N such that $\phi(x[\,-N\ldots 0\ldots N\,])$ determines $\phi(x)[0]$.

This N, in general, depends on the x.

Further, $\phi(x)$ may not be defined on some x.

Overview of the Proof



Computability of ϕ

Let us assume that A and B are computable systems.

Does this make ϕ computable?

No! ϕ is undefined at several points - it is defined on some measure 1 proper subset, but may be undefined on a measure 0, nonempty set.
Where exactly is the isomorphism well-defined?

Answer:(Ghosh, Nandakumar, Pal 2014) (At least) over the Martin-Löf random points in the systems. ϕ is *layerwise computable*.

Future Directions

Open Questions/Directions

1. Establish effective versions of results in topological dynamics. For example, what kinds of recurrence holds on an effectively co-meager set? (Ongoing work with Lutz, Jindal and Vijayvargiya.)
2. Establish effective versions of other recurrence theorems. Does, for example, the Furstenberg Multiple Recurrence hold for all Martin-Löf random points in effective dynamical systems?
3. Do effective versions give “easier” proofs of classical theorems? e.g. Look at what happens on typical points. (e.g. Day 2014)
4. Resource-bounded recurrence and ergodicity. We have to develop tools first. (Miyabe?)
5. See if current results extend to Schnorr randomness (Jason Rute?)
6. Effectivization of physical systems with maximal entropy - can this yield new examples of pseudorandomness?
7. Reverse Math of Dynamical System Theorems?
8. Connections to Descriptive Set Theory?

Thank You.

References

[AKM65] R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy and equivalence of dynamical systems. *Transactions of the AMS*, 114:309–319, 1965.

[Din70] E. I. Dinaburg. The relation between topological entropy and metric entropy. *Soviet Math.*, 11:13–16, 1970.

[GHR10] S. Galotolo, M. Hoyrup, and C. Rojas. Effective symbolic dynamics, random points, statistical behavior, complexity and entropy. *Information and Computation*, 208(1):23–41, 2010.

[Hoc09] Michael Hochman. Upcrossing inequalities for stationary sequences and applications. *Annals of Probability*, 37(6):2135–2149, 2009.

[Hoy12] Mathieu Hoyrup. The dimension of ergodic random sequences. In *Symposium on Theoretical Aspects of Computer Science*, pages 567–576, 2012.

[KS79] Michael Keane and Meier Smorodinsky. Bernoulli schemes of the same entropy are finitarily isomorphic. *Annals of Mathematics*, 109(2):397–406, 1979.

[Orn70] Donald Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Mathematics*, 4(3):337 – 352, 1970.