Randomness Extraction Fails for Finite-State Dimension

Subin Pulari , Akhil S[1]

HSE University Moscow, IIT Kanpur

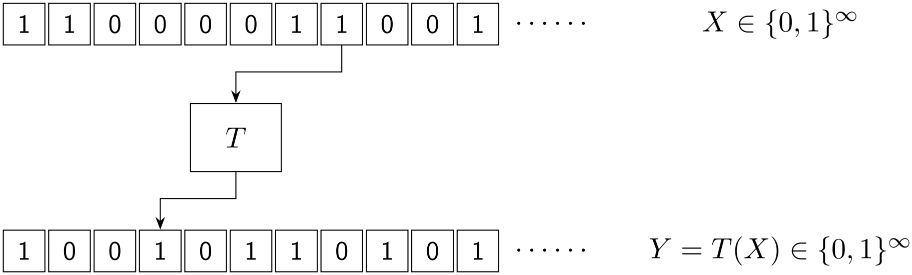
LICS 2026

[1]Supported by "Research-I Foundation" of CSE, IIT Kanpur.

Overview

Randomness Extraction

- Given a weak source of randomness,
  can we "convert" it to a stronger source of randomness ?



Constructive Dimension

- Constructive dimension $(\dim)$ measures the rate of randomness in a sequence.

  For $X \in \Sigma^\infty$, $\dim(X) = \liminf_n \frac{K(X\upharpoonright n)}{n}$

- $K(x)$: Kolmogorov Complexity of x.

Constructive Dimension

- Constructive dimension $(\dim)$ measures the rate of randomness in a sequence.

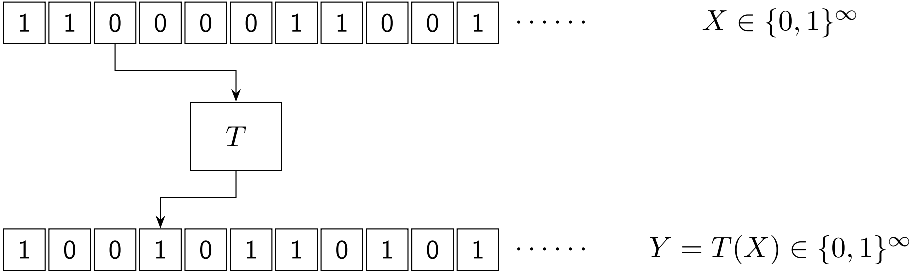
  For $X \in \Sigma^\infty$, $\dim(X) = \liminf_n \frac{K(X \restriction n)}{n}$
- $K(x)$: Kolmogorov Complexity of x.
- $0 \leq \dim \leq 1$.

  $1 \rightarrow$ High rate of randomness (Random strings)

  $0 \rightarrow$ Low rate of randomness (Eg 0^∞).

Randomness Extraction

- Given a low dimension,
  can we "convert" it to a higher dimension ?



Dimension extraction Fails for Turing Machines

- Constructive dimension (dim) : Measures the rate of randomness in a sequence.
- Compressing (Turing) to a higher dimension sequence (let alone Random) is not always possible !

Dimension extraction Fails for Turing Machines

- Constructive dimension $(\dim)$: Measures the rate of randomness in a sequence.
- Compressing (Turing) to a higher dimension sequence (let alone Random) is not always possible !

Theorem (Joseph Miller)

There exists an $A \leq_T \emptyset'$ *such that* $\dim(A) = \frac{1}{2}$, *and for every* $B \leq_T A$,

$$\dim(B) \leq \frac{1}{2}.$$

Finite-state analogue ?

- What if we consider finite-state models of computation ?

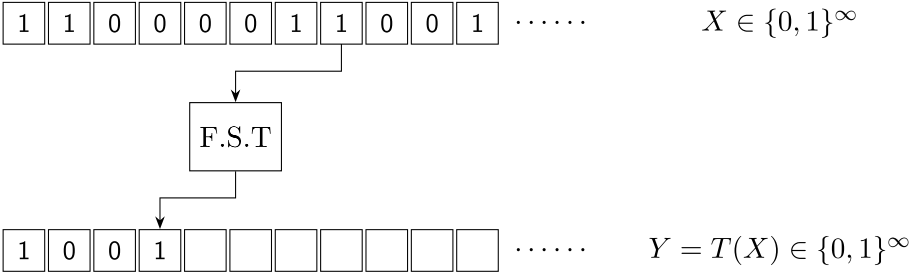
Finite-state analogue ?

- What if we consider finite-state models of computation ?
- Finite-state transducers : Transducers (T) which are finite-state computable.
- Finite-state dimension ($\dim_{\mathrm{FS}}$) : Density of randomness visible to finite-state automata.

Randomness Extraction

Question [Jun Le Goh]

Given a sequence with low finite-state dimension,
Can we use finite-state transducers and convert it to
A sequence with higher finite-state dimension ?



F.S Compression : Known Results

Theorem (Doty '08)

For every $\varepsilon > 0$ *and every* $X \in \Sigma^\infty$ *with* $\dim_{\mathrm{FS}}(X) > 0$, *there exists* $Y \in \Sigma^\infty$ *s.t*

$$Y \leq_{\mathrm{FS}} X, \qquad \dim_{\mathrm{FS}}(Y) \geq \frac{\dim_{\mathrm{FS}}(X)}{\mathrm{Dim}_{\mathrm{FS}}(X)} - \varepsilon, \qquad \mathrm{Dim}_{\mathrm{FS}}(Y) \geq 1 - \varepsilon.$$

- Compression to $\dim_{\mathrm{FS}}/\mathrm{Dim}_{\mathrm{FS}}$ is always possible.

Randomness extraction fails for f.s.d

Theorem

For any $s \in (0,1)$, there exists a binary sequence $X \in \{0,1\}^\infty$ such that

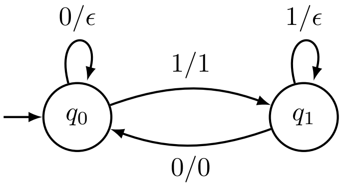
$$\dim_{\mathrm{FS}}(X) = s \quad \textit{and} \quad \forall \textit{ binary FSTs } T, \ \ \dim_{\mathrm{FS}}(T(X)) \leq s.$$

- Finite-state compression to a higher finite-state dimension sequence (let alone Normal) is not always possible !

Finite-state notions of Randomness

Finite-State Transducers

- Finite-state transducer (FST) : deterministic finite-state automaton with an output symbol per transition.
- Output : Concatenation of all emitted strings, over the run of input sequence on the FST.



Finite-State Notions of Randomness

- A sequence $X \in \Sigma^\infty$ is normal if

$\forall \ell \in \mathbb{N}, \forall u \in \Sigma^\ell$

$$\lim_{n\to\infty} P(u, X \upharpoonright n) = 2^{-\ell}$$

Finite-State Notions of Randomness

- A sequence $X \in \Sigma^\infty$ is normal if

  $\forall \ell \in \mathbb{N}, \forall u \in \Sigma^\ell$
  $$\lim_{n\to\infty} P(u, X \upharpoonright n) = 2^{-\ell}$$

- Every finite block $u \in \Sigma^\ell$ occurs with the expected limiting frequency $(2^{-\ell})$.

Finite-State Notions of Randomness

- A sequence $X \in \Sigma^\infty$ is normal if

  $\forall \ell \in \mathbb{N}, \forall u \in \Sigma^\ell$

  $$\lim_{n\to\infty} P(u, X \upharpoonright n) = 2^{-\ell}$$

- Every finite block $u \in \Sigma^\ell$ occurs with the expected limiting frequency $(2^{-\ell})$.

$$P(u,w) \ := \frac{1}{k}\Big|\{0 \le i < k : \ w[i\ell : (i+1)\ell] = u\}\Big|, \qquad k = \left\lfloor \frac{|w|}{\ell} \right\rfloor,$$

Finite-State Notions of Randomness

- Normal Sequences : Random to automata !.
- Finite-State Dimension : Quantifies rate/density of randomness visible to finite-state automata.

Finite-State Dimension

Inofrmation Lossless (IL)-FST : FST whose input can be reconstructed from the output and final state reached on that input.

Definition

If T is an FST and $X \in \{0,1\}^\infty$, then the *compression ratio* of T on X is

$$\rho_T(X) = \liminf_{n\to\infty} \frac{|T(X \upharpoonright n)|}{n}.$$

Definition

The *finite-state dimension* of a sequence $X \in \{0,1\}^\infty$ is

$$\dim_{\mathrm{FS}}(X) = \inf\{\rho_T(X) \mid T \text{ is an ILFST}\}.$$

Finite-State Dimension : Entropy

Definition

For $w \in \Sigma^*$ with $|w| \geq \ell$, define

$$H_\ell(w) = -\frac{1}{\ell \log |\Sigma|} \sum_{u \in \Sigma^\ell} P(u, w) \log P(u, w).$$

Finite-State Dimension : Entropy

Definition

For $w \in \Sigma^*$ with $|w| \geq \ell$, define

$$H_\ell(w) = -\frac{1}{\ell \log |\Sigma|} \sum_{u \in \Sigma^\ell} P(u, w) \, \log P(u, w).$$

For $X \in \Sigma^\infty$, $\ell \in \mathbb{N}$, define

$$H_\ell(X) = \liminf_{k \to \infty} H_\ell(X \upharpoonright k\ell),$$

Finite-State Dimension : Entropy

Definition

For $w \in \Sigma^*$ with $|w| \geq \ell$, define

$$H_\ell(w) = -\frac{1}{\ell \log |\Sigma|} \sum_{u \in \Sigma^\ell} P(u,w) \log P(u,w).$$

For $X \in \Sigma^\infty$, $\ell \in \mathbb{N}$, define

$$H_\ell(X) = \liminf_{k \to \infty} H_\ell(X \upharpoonright k\ell),$$

[BHV05] For $X \in \Sigma^\infty$,

$$\dim_{\mathrm{FS}}(X) = \inf_{\ell \geq 1} H_\ell(X).$$

Finite-State Dimension: Examples

- For Normal sequences, $N = \{n_i\}$, f.s.d $= 1$
- For sequences with 2x dilution, $X = \{n_i.n_i\}$, $\{n_i.0\}$, $\{n_i.n_{i+1}.0.0\}$ f.s.d $= 0.5$
- For regular sequences, say $X = 0^\infty$, f.s.d $= 0$

Randomness extraction fails for f.s.d

Theorem

For any $s \in (0,1)$, there exists a binary sequence $X \in \{0,1\}^\infty$ such that

$$\dim_{\mathrm{FS}}(X) = s \quad \textit{and} \quad \forall \textit{ binary FSTs } T, \;\; \dim_{\mathrm{FS}}(T(X)) \leq s.$$

- Randomness extraction (Finite-state) to a higher finite-state dimension sequence (let alone Normal) is not always possible !

Proof Outline

Goal

Given $s \in (0,1)$, find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = s$ and

for all F.S.T T, $\dim_{\mathrm{FS}} T(X) \leq s$.

Simpler question

- Diagonalise against a fixed Transducer T.
- Set $s = 1/2$.

Even Simpler question

- Diagonalise against a fixed Transducer T.
- Set $s = 1/2$.
- Assume T is a fixed length encoder

$$T : \Sigma^n \to \Sigma^*.$$

Given T, find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = 1/2$ and $\dim_{\mathrm{FS}} T(X) \geq 1/2$.

Even Simpler question

- Diagonalise against a fixed Transducer T.
- Set $s = 1/2$.
- Assume T is a fixed length encoder

$$T : \Sigma^n \to \Sigma^*.$$

Given T, find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = 1/2$ and

$\dim_{\mathrm{FS}} T(X) \geq 1/2$.

Incremental Proof outline

- Diagnolise against a Fixed length encoder
- Diagnolise against a Fixed F.S.T
- Diagnolise against all F.S.T s

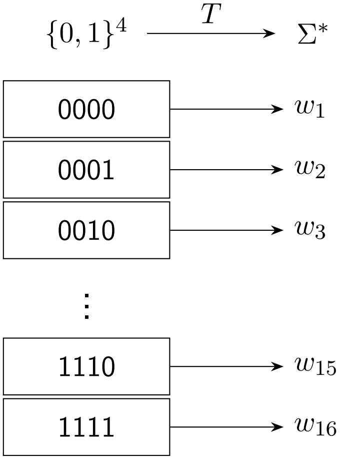
Diagnolise against a Fixed length encoder

Find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = 1/2$

and $\dim_{\mathrm{FS}} T(X) \geq 1/2$.

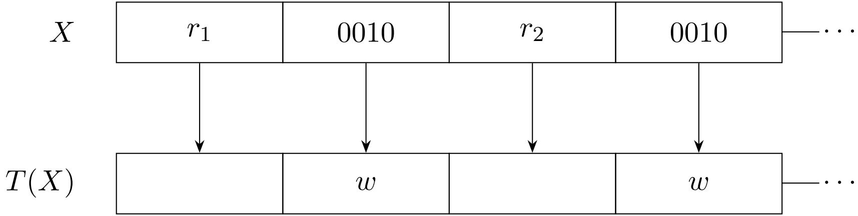
Diagonalisation

- Let w be the longest among the outputs.
- Say $u = 0010$ be the corresponding input.



Diagnolise against a Fixed length encoder

Take X by diluting a normal $N = r_1.r_2\ldots$ with u. $|r_i| = |u|$.



- Equal dilution $\implies$ $\dim_{\mathrm{FS}}(X) = 1/2$.
- w is longest among outputs $\implies$ $\dim_{\mathrm{FS}}(T(X)) \leq 1/2$.

Incremental Proof outline

- Diagnolise against a Fixed length encoder
- Diagnolise against a Fixed F.S.T
- Diagnolise against all F.S.T s

Single transducer

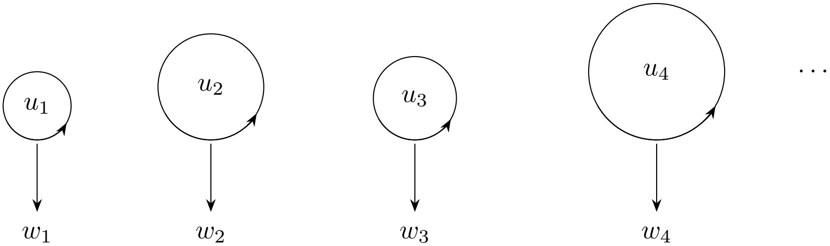
- Diagonalise against a fixed Transducer T.
- Set $s = 1/2$.
- T is an arbitrary F.S.T

Given T, find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = 1/2$ and

$\dim_{\mathrm{FS}} T(X) \geq 1/2$.

Diagnolise against a single F.S.T

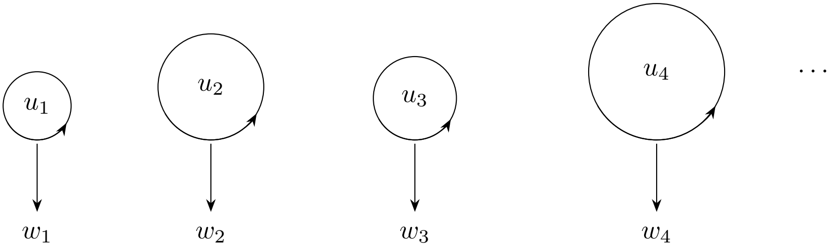
- Fix a terminal strongly connected component (S.C.C), enumerate all the cycles.



- Let $\beta_i = |w_i|/|u_i|$.
- Let u be the cucle with maximum value of β_i. Let w be the corresponding output,

Diagnolise against a single F.S.T

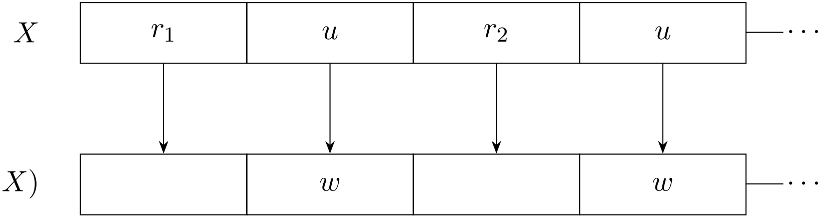
- Fix a terminal strongly connected component (S.C.C), enumerate all the cycles.



- Let $\beta_i = |w_i|/|u_i|$.
- Let u be the cucle with maximum value of β_i. Let w be the corresponding output,

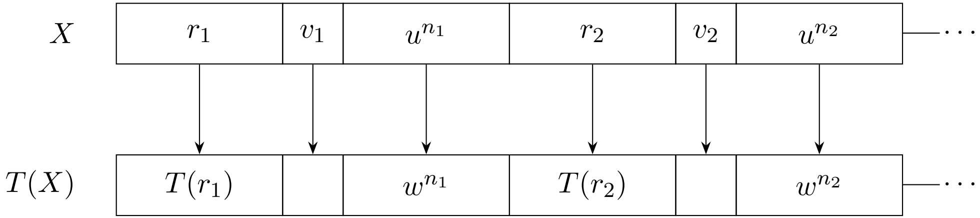
Diagnolise against a single F.S.T

Attempt : Take X by diluting a normal $N = r_1.r_2\dots$ with u.



Diagnolise against a single F.S.T

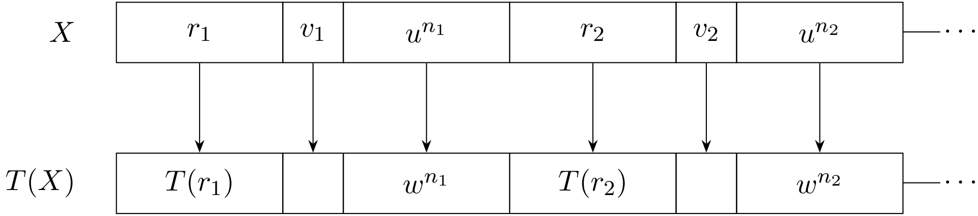
Take X by diluting a normal $N = r_1.r_2\ldots$ with u^{n_i}. $|r_i| = |u^{n_i}|$.



- Add (finite) pads v_i to bring to cycle u after r_i
- Ensure $\lim_i n_i \to \infty$.

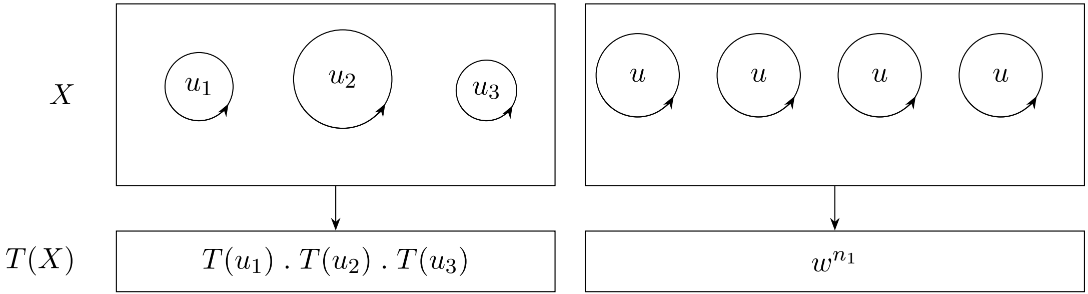
Diagnolise against a single F.S.T

Take X by diluting a normal $N = r_1.r_2\ldots$ with u^{n_i}. $|r_i| = |u^{n_i}|$.



- (Roughly) Equal dilution $\implies \dim_{\mathrm{FS}}(X) = 1/2$.

Diagnolise against a single F.S.T



- Each r_i can be decomposed into cycles.
- $|w|/|u|$ is largest $\implies |T(r_i)| \leq |w^{n_i}|$.
- $\implies \dim_{\mathrm{FS}}(T(X)) \leq 1/2$.

Incremental Proof outline

- Diagnolise against a Fixed length encoder
- Diagnolise against a Fixed F.S.T
- Diagnolise against all F.S.T s

Diagonalise against all transducers

- Diagnolise against all Transducers T.

Find an $X \in \Sigma^{\infty}$, such that $\dim_{\mathrm{FS}}(X) = 1/2$ and

for all F.S.T T, $\dim_{\mathrm{FS}} T(X) \geq 1/2$.

Diagonalise against all transducers

- Diagnolise against all Transducers T.

Find an $X \in \Sigma^\infty$, such that $\dim_{\mathrm{FS}}(X) = 1/2$ and

for all F.S.T T, $\dim_{\mathrm{FS}} T(X) \geq 1/2$.

Diagonalise against all transducers

- Take an enumeration of F.S.T s $T_1, T_2, \ldots$
- Ensure each transducer occurs infinitely often in the enumeration (Enumerate $\{T\} \times \{T\}$).
- At stage i, diagonalise against transducer T_i.

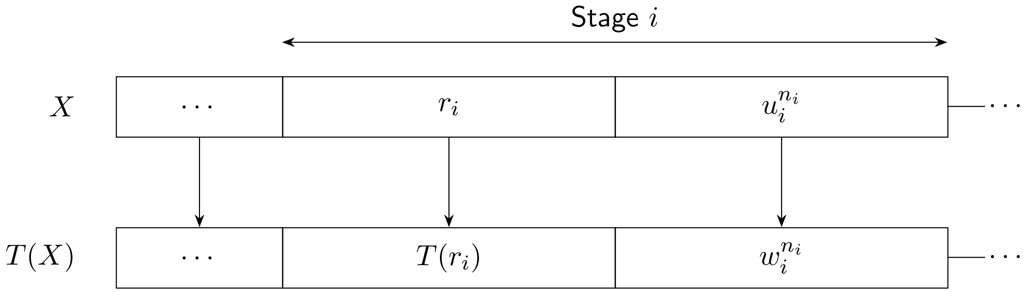
Diagonalise against all transducers

- Take an enumeration of F.S.T s $T_1, T_2, \ldots$
- Ensure each transducer occurs infinitely often in the enumeration (Enumerate $\{T\} \times \{T\}$).
- At stage i, diagonalise against transducer T_i.

Diagonalise against all transducers

- Take an enumeration of F.S.T s $T_1, T_2, \ldots$
- Ensure each transducer occurs infinitely often in the enumeration (Enumerate $\{T\} \times \{T\}$).
- At stage i, diagonalise against transducer T_i.

Diagonalise against all transducers



- Stage i: Diagonalise against T_i.
- Ensure $n_{i-1} = o(n_i)$.

Conclusion

- Randomness extraction is not always possible in the finite-state world.
- For every $s \in (0,1)$, there exists a sequence X with

$$\dim_{\mathrm{FS}}(X) = s$$

such that no finite-state transducer can increase its finite-state dimension.

Open Problems

- Is Doty's $\dim_{\mathrm{FS}} / \mathrm{Dim}_{\mathrm{FS}}$ bound the best for all sequences ?
- Can Turing machines always increase finite-state dimension ?
- Can we diagonalise against dimensions w.r.t all bases b?

References

- J. J. Dai, J. I. Lathrop, J. H. Lutz, E. Mayordomo. *Finite-state dimension*. Theoretical Computer Science, 310(1–3), 2004.
- J. S. Miller. *Extracting information is hard: a Turing degree of non-integral effective Hausdorff dimension*. Advances in Mathematics, 226(1):373–384, 2011.
- D. Doty. *Dimension Extractors and Optimal Decompression*. Theory Comput Syst 43, 425–463 (2008).