

# Every Sequence is Decompressible from a Random One

David Doty <sup>1</sup>

<sup>1</sup>Department of Computer Science, Iowa State University, Ames, IA 50011 USA. *ddoty at iastate dot edu*. This research was funded in part by grant number 9972653 from the National Science Foundation as part of their Integrative Graduate Education and Research Traineeship (IGERT) program.

## Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant
  - Decompression done with a **universal Turing machine**.

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant
  - Decompression done with a **universal Turing machine**.
- The main result identifies an optimally compressed representation of any **infinite sequence**.

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant
  - Decompression done with a **universal Turing machine**.
- The main result identifies an optimally compressed representation of any **infinite sequence**.
  - “Optimal” = best asymptotic compression ratio achievable by **Turing reductions**

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant
  - Decompression done with a **universal Turing machine**.
- The main result identifies an optimally compressed representation of any **infinite sequence**.
  - “Optimal” = best asymptotic compression ratio achievable by **Turing reductions**
  - Decompression done with a **universal Turing reduction**.

# Results (Intuition)

- Kolmogorov complexity identifies an optimally compressed representation of any **finite string**.
  - “Optimal” = smaller than any **compression algorithm**, within a constant
  - Decompression done with a **universal Turing machine**.
- The main result identifies an optimally compressed representation of any **infinite sequence**.
  - “Optimal” = best asymptotic compression ratio achievable by **Turing reductions**
  - Decompression done with a **universal Turing reduction**.
- Compression direction is not known to be computable.

## Results (Intuition)

- New characterization of constructive dimension:

# Results (Intuition)

- New characterization of constructive dimension:
  - constructive dimension = optimal compression ratio achievable on a sequence with Turing reductions.

# Results (Intuition)

- New characterization of constructive dimension:
  - constructive dimension = optimal compression ratio achievable on a sequence with Turing reductions.
  - Complements Mayordomo's Kolmogorov complexity characterization of constructive dimension.

# Results (Intuition)

- New characterization of constructive dimension:
  - constructive dimension = optimal compression ratio achievable on a sequence with Turing reductions.
  - Complements Mayordomo's Kolmogorov complexity characterization of constructive dimension.
- The compressed sequence can be made Martin-Löf random.

# Results (Intuition)

- New characterization of constructive dimension:
  - constructive dimension = optimal compression ratio achievable on a sequence with Turing reductions.
  - Complements Mayordomo's Kolmogorov complexity characterization of constructive dimension.
- The compressed sequence can be made Martin-Löf random.
  - Kučera and Gács independently showed that every sequence is Turing reducible to a random sequence.

# Results (Intuition)

- New characterization of constructive dimension:
  - constructive dimension = optimal compression ratio achievable on a sequence with Turing reductions.
  - Complements Mayordomo's Kolmogorov complexity characterization of constructive dimension.
- The compressed sequence can be made Martin-Löf random.
  - Kučera and Gács independently showed that every sequence is Turing reducible to a random sequence.
  - In the words of Gács, "it permits us to view even very pathological sequences as the result of the combination of two relatively well-understood processes: the completely chaotic outcome of coin-tossing, and a transducer algorithm."

# Turing Reductions

- $\mathbf{C} = \{0, 1\}^\infty =$  all infinite, binary sequences

# Turing Reductions

- $\mathbf{C} = \{0, 1\}^\infty =$  all infinite, binary sequences
- OTM is the set of all oracle Turing machines.

# Turing Reductions

- $\mathbf{C} = \{0, 1\}^\infty =$  all infinite, binary sequences
- OTM is the set of all oracle Turing machines.
- For  $S, R \in \mathbf{C}$  and  $M \in \text{OTM}$ , we write

$$S \leq_T R \text{ via } M$$

if  $M^R$  computes  $S$ . Write  $M(R) = S$ .

# Turing Reductions

- $\mathbf{C} = \{0, 1\}^\infty =$  all infinite, binary sequences
- OTM is the set of all oracle Turing machines.
- For  $S, R \in \mathbf{C}$  and  $M \in \text{OTM}$ , we write

$$S \leq_T R \text{ via } M$$

if  $M^R$  computes  $S$ . Write  $M(R) = S$ .

- $\#(M^R, S \upharpoonright n)$  is the number of bits of  $R$  queried by  $M$  when computing  $S \upharpoonright n$ .

# Infinite Sequence Compression

## Definition

$$\rho_M^-(S, R) = \liminf_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}$$

# Infinite Sequence Compression

## Definition

$$\rho_M^-(S, R) = \liminf_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}, \quad \rho_M^+(S, R) = \limsup_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}$$

# Infinite Sequence Compression

## Definition

$$\rho_M^-(S, R) = \liminf_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}, \quad \rho_M^+(S, R) = \limsup_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}$$

Lower compression ratio of  $S$ :

$$\rho^-(S) = \min_{\substack{R \in \mathbf{C} \\ M \in \mathbf{OTM}}} \{ \rho_M^-(S, R) \mid S \leq_T R \text{ via } M \}$$

# Infinite Sequence Compression

## Definition

$$\rho_M^-(S, R) = \liminf_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}, \quad \rho_M^+(S, R) = \limsup_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}$$

*Lower compression ratio of S:*

$$\rho^-(S) = \min_{\substack{R \in \mathbf{C} \\ M \in \mathbf{OTM}}} \{ \rho_M^-(S, R) \mid S \leq_T R \text{ via } M \}$$

*Upper compression ratio of S:*

$$\rho^+(S) = \min_{\substack{R \in \mathbf{C} \\ M \in \mathbf{OTM}}} \{ \rho_M^+(S, R) \mid S \leq_T R \text{ via } M \}$$

# Infinite Sequence Compression

## Definition

$$\rho_M^-(S, R) = \liminf_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}, \quad \rho_M^+(S, R) = \limsup_{n \rightarrow \infty} \frac{\#(M^R, S \upharpoonright n)}{n}$$

Lower compression ratio of  $S$ :

$$\rho^-(S) = \min_{\substack{R \in \mathbf{C} \\ M \in \text{OTM}}} \{ \rho_M^-(S, R) \mid S \leq_T R \text{ via } M \}$$

Upper compression ratio of  $S$ :

$$\rho^+(S) = \min_{\substack{R \in \mathbf{C} \\ M \in \text{OTM}}} \{ \rho_M^+(S, R) \mid S \leq_T R \text{ via } M \}$$

For  $S \in \mathbf{C}$ ,  $\rho^-(S)$  and  $\rho^+(S)$  are the optimal best- and worst-case compression ratios achievable with **Turing reductions**.

# Kolmogorov Complexity (Finite String Compression)

Let  $U$  be a self-delimiting, universal Turing machine.

# Kolmogorov Complexity (Finite String Compression)

Let  $U$  be a self-delimiting, universal Turing machine.

## Definition

For  $w \in \{0, 1\}^*$ , the *Kolmogorov complexity* of  $w$  is

$$K(w) = \min\{ |\pi| \mid U(\pi) = w \}.$$

# Kolmogorov Complexity (Finite String Compression)

Let  $U$  be a self-delimiting, universal Turing machine.

## Definition

For  $w \in \{0, 1\}^*$ , the *Kolmogorov complexity* of  $w$  is

$$K(w) = \min\{ |\pi| \mid U(\pi) = w \}.$$

## Fact

For any compression algorithm  $A : \{0, 1\}^* \rightarrow \{0, 1\}^*$ , there is a constant  $c_A \in \mathbb{N}$  such that, for all  $w \in \{0, 1\}^*$ ,

$$K(w) \leq |A(w)| + c_A.$$

# Kolmogorov Complexity (Finite String Compression)

Let  $U$  be a self-delimiting, universal Turing machine.

## Definition

For  $w \in \{0, 1\}^*$ , the *Kolmogorov complexity* of  $w$  is

$$K(w) = \min\{ |\pi| \mid U(\pi) = w \}.$$

## Fact

For any compression algorithm  $A : \{0, 1\}^* \rightarrow \{0, 1\}^*$ , there is a constant  $c_A \in \mathbb{N}$  such that, for all  $w \in \{0, 1\}^*$ ,

$$K(w) \leq |A(w)| + c_A.$$

For  $w \in \{0, 1\}^*$ ,  $\frac{K(w)}{|w|}$  is the optimal compression ratio achievable with an **algorithm**.

# Randomness and Constructive Dimension

## Definition

A sequence  $S \in \mathbf{C}$  is *Martin-Löf random*, and we write  $S \in \text{RAND}$ , if there is a constant  $c$  such that, for all  $n \in \mathbb{N}$ ,

$$K(S \upharpoonright n) \geq n - c.$$

# Randomness and Constructive Dimension

## Definition

A sequence  $S \in \mathbf{C}$  is *Martin-Löf random*, and we write  $S \in \text{RAND}$ , if there is a constant  $c$  such that, for all  $n \in \mathbb{N}$ ,

$$K(S \upharpoonright n) \geq n - c.$$

## Definition

For all  $S \in \mathbf{C}$ , the *constructive dimension* and the *strong constructive dimension* of  $S$  are respectively

$$\begin{aligned} \dim(S) &= \liminf_{n \rightarrow \infty} \frac{K(S \upharpoonright n)}{n}, \\ \text{Dim}(S) &= \limsup_{n \rightarrow \infty} \frac{K(S \upharpoonright n)}{n}. \end{aligned}$$

# Every sequence is reducible to a random one (Kuřera 1985/1989, Gács 1986)

## Theorem

*There is an OTM  $M$  such that, for all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  such that*

- 1  $S \leq_T R$  via  $M$ .
- 2  $\rho_M^+(S, R) = 1$ .

# Every sequence is reducible to a random one (Kučera 1985/1989, Gács 1986)

## Theorem

*There is an OTM  $M$  such that, for all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  such that*

- 1  $S \leq_T R$  via  $M$ .
- 2  $\rho_M^+(S, R) = 1$ .

- Random sequences have no structure detectable by any algorithm.
- Information about *any* sequence  $S$  can be hidden in a sequence  $R$ , without imparting any detectable structure on  $R$ .

## Compression via Turing reductions (Ryabko 1986)

## Theorem

For every  $S \in \mathbf{C}$ ,

$$\dim(S) = \inf_{M_e, M_d \in \text{OTM}} \left\{ \rho_{M_d}^-(S, M_e(S)) \mid M_d(M_e(S)) = S \right\}.$$

## Compression via Turing reductions (Ryabko 1986)

## Theorem

For every  $S \in \mathbf{C}$ ,

$$\dim(S) = \inf_{M_e, M_d \in \text{OTM}} \left\{ \rho_{M_d}^-(S, M_e(S)) \mid M_d(M_e(S)) = S \right\}.$$

- PRO: The compression direction is computable by  $M_e$ .
- CON: Must optimize over all OTMs.
- CON: No  $\rho^+$  or Dim.

# Every sequence is decompressible from a random one

## Lemma

For all  $S \in \mathbf{C}$ ,  $\rho^-(S) \geq \dim(S)$ , and  $\rho^+(S) \geq \text{Dim}(S)$ .

# Every sequence is decompressible from a random one

## Lemma

For all  $S \in \mathbf{C}$ ,  $\rho^-(S) \geq \dim(S)$ , and  $\rho^+(S) \geq \text{Dim}(S)$ .

## Theorem

*There is an OTM  $M$  such that, for every  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  such that*

- 1  $S \leq_T R$  via  $M$ .
- 2  $\rho_M^-(S, R) = \dim(S)$ .
- 3  $\rho_M^+(S, R) = \text{Dim}(S)$ .

# Every sequence is decompressible from a random one

## Lemma

For all  $S \in \mathbf{C}$ ,  $\rho^-(S) \geq \dim(S)$ , and  $\rho^+(S) \geq \text{Dim}(S)$ .

## Theorem

*There is an OTM  $M$  such that, for every  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  such that*

- 1  $S \leq_T R$  via  $M$ .
- 2  $\rho_M^-(S, R) = \dim(S)$ .
- 3  $\rho_M^+(S, R) = \text{Dim}(S)$ .

## Corollary

For all  $S \in \mathbf{C}$ ,  $\rho^-(S) = \dim(S)$ , and  $\rho^+(S) = \text{Dim}(S)$ .

# "Proof"

## Theorem

*For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that*

$$S \leq_T R \text{ via } M, \quad \rho_M^-(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_T R \text{ via } M, \quad \rho_M^-(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_T R \text{ via } M, \quad \rho_M^-(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .
- To compute the  $(i + 1)^{\text{th}}$  block of  $k_i$  bits, let  $A_i \subseteq \{0, 1\}^{k_i}$  be the set of length- $k_i$  strings  $u$  for which  $\mathbf{d}((S \upharpoonright n_i)u) \geq \mathbf{d}(S \upharpoonright n_i + k_i)$ .

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_{\text{T}} R \text{ via } M, \quad \rho_{\overline{M}}(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .
- To compute the  $(i+1)^{\text{th}}$  block of  $k_i$  bits, let  $A_i \subseteq \{0, 1\}^{k_i}$  be the set of length- $k_i$  strings  $u$  for which  $\mathbf{d}((S \upharpoonright n_i)u) \geq \mathbf{d}(S \upharpoonright n_i + k_i)$ .
- Then  $(\exists^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{dim}(S)$  and  $(\forall^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{Dim}(S)$ .

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_T R \text{ via } M, \quad \rho_M^-(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .
- To compute the  $(i + 1)^{\text{th}}$  block of  $k_i$  bits, let  $A_i \subseteq \{0, 1\}^{k_i}$  be the set of length- $k_i$  strings  $u$  for which  $\mathbf{d}((S \upharpoonright n_i)u) \geq \mathbf{d}(S \upharpoonright n_i + k_i)$ .
- Then  $(\exists^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{dim}(S)$  and  $(\forall^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{Dim}(S)$ .
- Set the next  $\log |A_i|$  bits of  $P \in \mathbf{C}$  to be an index into  $A_i$ .

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_T R \text{ via } M, \quad \rho_M^-(S, R) = \dim(S), \quad \rho_M^+(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .
- To compute the  $(i + 1)^{\text{th}}$  block of  $k_i$  bits, let  $A_i \subseteq \{0, 1\}^{k_i}$  be the set of length- $k_i$  strings  $u$  for which  $\mathbf{d}((S \upharpoonright n_i)u) \geq \mathbf{d}(S \upharpoonright n_i + k_i)$ .
- Then  $(\exists^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{dim}(S)$  and  $(\forall^\infty i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{Dim}(S)$ .
- Set the next  $\log |A_i|$  bits of  $P \in \mathbf{C}$  to be an index into  $A_i$ , along with  $d_i \in \mathbb{Q}$ , slightly smaller than  $\mathbf{d}(S \upharpoonright n_i + k_i)$ , using  $O(\log k_i)$  bits.

# "Proof"

## Theorem

For all  $S \in \mathbf{C}$ , there is a sequence  $R \in \text{RAND}$  and an OTM  $M$  such that

$$S \leq_{\text{T}} R \text{ via } M, \quad \rho_{\bar{M}}(S, R) = \dim(S), \quad \rho_{M^+}(S, R) = \text{Dim}(S).$$

## Proof.

- Suppose we have computed  $i$  blocks of  $S$ :  $S \upharpoonright n_i$ .
- To compute the  $(i + 1)^{\text{th}}$  block of  $k_i$  bits, let  $A_i \subseteq \{0, 1\}^{k_i}$  be the set of length- $k_i$  strings  $u$  for which  $\mathbf{d}((S \upharpoonright n_i)u) \geq \mathbf{d}(S \upharpoonright n_i + k_i)$ .
- Then  $(\exists^{\infty} i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{dim}(S)$  and  $(\forall^{\infty} i) \sum_{j=1}^i \log |A_j| \leq n_i \cdot \text{Dim}(S)$ .
- Set the next  $\log |A_i|$  bits of  $P \in \mathbf{C}$  to be an index into  $A_i$ , along with  $d_i \in \mathbb{Q}$ , slightly smaller than  $\mathbf{d}(S \upharpoonright n_i + k_i)$ , using  $O(\log k_i)$  bits.
- Then  $S \leq_{\text{T}} P$ . Reduce  $P$  to  $R \in \text{RAND}$  via the construction of Gács.

# Questions?