Redundancy of information: Lowering effective dimension

Joseph S. Miller
University of Wisconsin–Madison

Computability Theory and Applications Online Seminar
in association with
the Midwest Computability Seminar and COVID-19

August 18, 2020

A talk in two parts

- In the first part, we will discuss the 2018 paper "Dimension 1 sequences are close to randoms," by Greenberg, M., Shen, and Westrick (GrMShW 2018).

- The second part, to which the title refers, will focus on recent work of Goh, M., Soskova, and Westrick on lowering effective dimension (GoMSoW).

Effective randomness

Definition. If $U\colon 2^{<\omega} \to 2^{<\omega}$ is a universal prefix-free machine, then

$$K(\sigma) = \min\{|\tau| : U(\tau) = \sigma\}$$

is the *prefix-free (Kolmogorov) complexity* of σ.

Theorem. $X \in 2^{\omega}$ is *(Martin-Löf) random* if and only if

$$(\exists c)(\forall n)\ K(X \upharpoonright n) \geqslant n - c.$$

Definition (Lutz; Mayordomo)

The *(effective Hausdorff) dimension* of a sequence $X \in 2^{\omega}$ is

$$\dim X = \liminf_{n\to\infty} \frac{K(X \upharpoonright n)}{n}.$$

Note that every random has dimension 1.

Hamming and Besicovitch distance

Definition. For each n, the *Hamming distance* between $\sigma, \tau \in 2^n$ is $d(\sigma, \tau) = |\sigma \triangle \tau|$.

This makes $(2^n, d)$ a metric space called n-dimensional *Hamming space*. It is one of the primary objects of study in coding theory.

Definition. The *Besicovitch (pseudo-)distance* between $X, Y \in 2^\omega$ is

$$d(X, Y) = \limsup_{n \to \infty} \frac{|X \upharpoonright n \triangle Y \upharpoonright n|}{n}.$$

In other words, it is the upper density of the symmetric difference of X and Y.

If $d(X, Y) = 0$, we say that X and Y are *coarsely equivalent*.

Dimension 1 sequences

Not every dimension 1 sequence is random.

Example. Let $Y \in 2^\omega$ be random. Define X by

$$X(n) = \begin{cases} 1 & \text{if } n \text{ is a power of 2,} \\ Y(n) & \text{otherwise.} \end{cases}$$

Then X is clearly not random (random sequences must be immune), but we have only slightly lowered the initial segment complexity, so $\dim X = 1$.

But every dimension 1 sequence is close to a random.

Theorem (GrMShW 2018). A sequence $X \in 2^\omega$ has dimension 1 if and only if it is coarsely equivalent to a random sequence.

The proof uses a theorem of Harper about Hamming space... and compactness.

Further questions

We figured out how to efficiently raise the complexity of dimension 1 sequence to get a random. What's next?

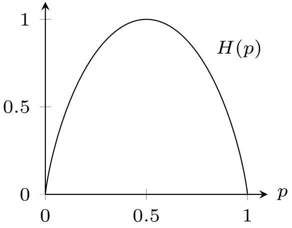
- How hard is it to increase the complexity of a dimension t sequence to get a random? In other words, what is the (Besicovitch) distance from a dimension t sequence to the nearest random sequence? (or equivalently, a dimension 1)?
- What is the distance from a dimension t to the nearest dimension $s \geqslant t$ sequence?
- What about lowering dimension? What is the distance from a dimension 1 to the nearest dimension t sequence?
- What is the distance from a dimension s to the nearest dimension $t \leqslant s$ sequence? *This is Part II.*

Entropy and Hamming balls

Definition. The *Shannon entropy function* is

$$H(p) = -p\log(p) - (1-p)\log(1-p).$$

$H(p)$ is the information content of one coin flip with probabilities p and $1-p$.



Notation. If $\sigma \in 2^n$ and $r \leqslant n$, then $B_r(\sigma) = \{\tau \in 2^n : d(\sigma,\tau) \leqslant r\}$ is the *Hamming ball* of radius r centered at σ.

Let $V(n,r)$ be the size of a Hamming ball of radius r in 2^n. (They all have the same size.)

Lemma. If $r \leqslant n/2$, then $H(r/n)\,n - o(n) \leqslant \log V(n,r) \leqslant H(r/n)\,n$.

Entropy, density, and Bernoulli randomness

Lemma. If $r \leqslant n/2$, then $\log V(n,r) \approx H(r/n)\,n$.

Prop. If $\sigma \in 2^n$ has pn ones, then $K(\sigma) \leqslant H(p)\,n + o(n)$.

Proof. Note that $\sigma \in B_{pn}(0^n)$. There are $V(n,pn)$ strings in $B_{pn}(0^n)$, so we can give each a description of length $\approx \log V(n,pn) \approx H(p)\,n$ (for $p \leqslant 1/2$). If $p > 1/2$, switch the roles of 0 and 1 and use the fact that $H(1-p) = H(p)$ to get the same bound. □

Corollary. If X has asymptotic density p, then $\dim X \leqslant H(p)$.

Definition. For $p \in [0,1]$, a *Bernoulli p-random* is generated by independently sampling the distribution on $\{0,1\}$ with $\Pr(1) = p$ and $\Pr(0) = (1-p)$. (This can be effectivized.)

Note that Bernoulli p-randoms have density p.

Prop. If $X \in 2^\omega$ is a Bernoulli p-random, then $\dim X = H(p)$.

The best case

Prop. If $d(X, Y) = p$, then $\dim Y \leqslant \dim X + H(p)$.

Proof. Say $Y \upharpoonright n$ and $X \upharpoonright n$ differ on density $\approx p$. To code $Y \upharpoonright n$, it is sufficient to code $X \upharpoonright n$ and the $\approx pn$ changes. Therefore,

$$K(Y \upharpoonright n) \lesssim K(X \upharpoonright n) + H(p)\, n. \qquad \square$$

Corollary. If $\dim X = t$ and $\dim Y = s$, then

$$d(X, Y) \geqslant H^{-1}(|s - t|).$$

(Here, $H^{-1} \colon [0, 1] \to [0, 1/2]$ is an increasing function.)

This bound is achievable (if we get to pick both sequences).

Prop (GrMShW 2018). If $0 \leqslant t \leqslant s \leqslant 1$, then there are $X, Y \in 2^\omega$ such that $\dim X = t$, $\dim Y = s$, and $d(X, Y) = H^{-1}(s - t)$.

A simple obstacle

Question. What is the distance from a dimension t sequence to the nearest dimension $s \geqslant t$ sequence? Is it always $H^{-1}(s-t)$?

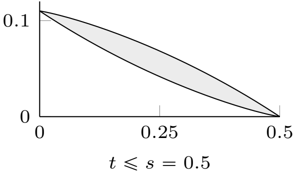
No, and the counterexample is simple.

- Let X be Bernoulli $H^{-1}(t)$-random.
- So $\dim X = t$ and the density of ones in X is $H^{-1}(t)$.
- If $\dim Y = s$, then the density of ones in Y is at least $H^{-1}(s)$.
- Therefore, $d(X,Y) \geqslant H^{-1}(s) - H^{-1}(t)$.

It turns out that

$$H^{-1}(s) - H^{-1}(t) > H^{-1}(s-t),$$

except for trivialities.



Increasing dimension

The previous simple obstacle actually witnesses the worst case.

Thm (GrMShW 2018). Let $0 \leqslant t \leqslant s \leqslant 1$. If $\dim X = t$, then there is a $Y \in 2^\omega$ with $\dim Y = s$ and $d(X, Y) \leqslant H^{-1}(s) - H^{-1}(t)$.

- There are similarities to the proof that dimension 1 sequences are close to random sequences.
  - The analogue for finite strings follows from Harper's theorem.
  - The construction is done blockwise, then we use compactness.
- But there is a new difficulty: X can have regions of complexity higher than t followed by regions of complexity lower than t.
- There are actually two constructions, conditioned on which of $(1-t)(H^{-1}(t))'$ and $(1-s)(H^{-1}(s))'$ is larger.
- Each construction is proved to work, under its respective assumption, using a somewhat delicate convexity argument.

Increasing dimension: $s = 0.99$

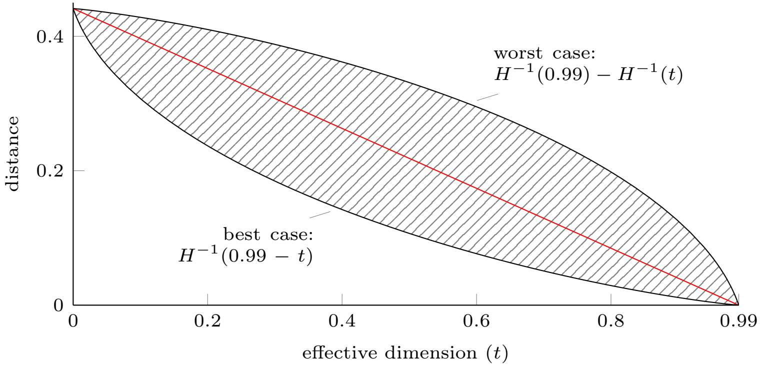


Figure: Best and worst cases for the distance from a dimension $t \leqslant 0.99$ sequence to the nearest dimension 0.99 sequence. Mysteriously, this is rotationally symmetric, and would be for any dimension s in place of 0.99. (It is not symmetric under reflection.)

Decreasing dimension

Lemma (Delsarte and Piret). For each $r \leqslant n$, the Hamming space 2^n can be covered by $\approx 2^n/V(n,r)$ Hamming balls of radius r.

- The collection of centers is called a *covering code* of radius r.
- By an easy volume argument, the lemma is (essentially) optimal.
- The lemma is proved using the probabilistic method. But we can find such a code via exhaustive search.
- For $p \leqslant 1/2$, there is a covering code of radius pn and size $\approx 2^n/V(n,pn)$. For every center τ in that code,

$$K(\tau) \lesssim \log\left(2^n/V(n,pn)\right) = n - \log V(n,pn) \approx (1-H(p))n.$$

Proposition. Every $\sigma \in 2^n$ is within $H^{-1}(1-t)n$ bits of a string τ such that $K(\tau) \lesssim tn$.

Decreasing dimension, cont.

Proposition. Every $\sigma \in 2^n$ is within $H^{-1}(1-t)n$ bits of a string τ such that $K(\tau) \lesssim tn$.

Theorem (GrMShW 2018). For any $Y \in 2^\omega$ and $t \in [0,1]$, there is an $X \in 2^\omega$ such that $\dim X = t$ and $d(X,Y) \leqslant H^{-1}(1-t)$.

Proof. Simply apply the proposition blockwise to Y. The blocks should grow, but not too quickly; it's sufficient to let the nth block of Y have size n. □

Corollary (GrMShW 2018). If $\dim Y = 1$ and $t \in [0,1]$, then there is an $X \in 2^\omega$ such that $\dim X = t$ and $d(X,Y) = H^{-1}(1-t)$.

Starting from a dimension 1, the best case can always achieved!

Every taco truck is near a corner, but...

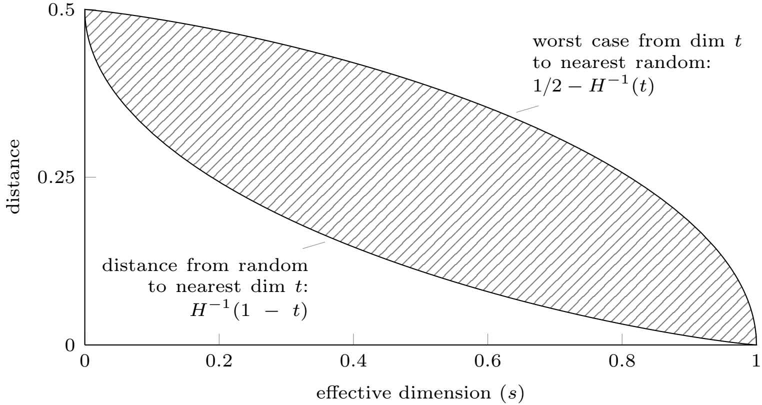


Figure: Every random is close to a dimension t sequence, but not every dimension t sequence is close to a random.

Another obstacle

Question. What is the distance from a dimension s to the nearest dimension $t \leqslant s$ sequence? Is it always $H^{-1}(s-t)$?

No: if information is stored *redundantly*, it is harder to erase. (Note that this can't happen in dimension 1.)

Let's look at an example (GrMShW 2018).

- Let $Z \in 2^\omega$ be random and $Y = Z \oplus Z$. So $\dim Y = 1/2$.
- For a contradiction, fix an $X \in 2^\omega$ of dimension 0 such that $d(X,Y) = H^{-1}(1/2)$.
- We can code $Y \upharpoonright 2n$ by giving:
  - A description of $X \upharpoonright 2n$,
  - For each $i < n$ such that $X(2i) \neq X(2i+1)$, the value $Y(2i)$, and
  - A description of $\{i < n \colon X(2i) = X(2i+1) \neq Y(2i)\}$.

Another obstacle, cont.

- We can code $Y \upharpoonright 2n$ by giving:
  - A description of $X \upharpoonright 2n$: $K(X \upharpoonright 2n)$.
  - For each $i < n$ such that $X(2i) \neq X(2i+1)$, the value $Y(2i)$: There are $\lesssim H^{-1}(1/2)2n$ such i.
  - A description of $\{i < n \colon X(2i) = X(2i+1) \neq Y(2i)\}$: This is a subset of n of size $\lesssim H^{-1}(1/2)n$, so it has a description of length $\lesssim H(H^{-1}(1/2))n = n/2$.
- Putting this all together,

$$n \approx K(Y \upharpoonright 2n) \lesssim K(X \upharpoonright 2n) + H^{-1}(1/2)2n + n/2.$$

- So $K(X \upharpoonright 2n) \gtrsim n/2 - H^{-1}(1/2)2n \approx 0.28n$, which contradicts the assumption that X has dimension 0.

Decreasing dimension: the worst case

Theorem (GoMSoW). If $Y \in 2^\omega$ has dimension s and $0 \leqslant t < s$, then there is an $X \in 2^\omega$ with $\dim X = t$ and

$$d(X,Y) \leqslant \mathrm{Worst}(s,t) := \begin{cases} H^{-1}(1-t) & \text{if } t \leqslant 1 - H(2^{s-1}), \\ \dfrac{t-s}{\log(2^{1-s}-1)} & \text{if } t > 1 - H(2^{s-1}). \end{cases}$$

Furthermore, for any $s \in (0,1]$, there is a sequence Y_s of dimension s such that these bounds are tight.

Observations.

- If $t \leqslant s$ is small enough, then we can't lower dimension from Y_s any better than if it were a *random* sequence!
- In particular, for any $s \in (0,1]$, the distance from Y_s to the nearest dimension 0 is $1/2$.

Decreasing dimension: the worst case

Theorem (GoMSoW). If $Y \in 2^\omega$ has dimension s and $0 \leqslant t < s$, then there is an $X \in 2^\omega$ with $\dim X = t$ and

$$d(X,Y) \leqslant \mathrm{Worst}(s,t) := \begin{cases} H^{-1}(1-t) & \text{if } t \leqslant 1 - H(2^{s-1}), \\ \dfrac{t-s}{\log(2^{1-s}-1)} & \text{if } t > 1 - H(2^{s-1}). \end{cases}$$

Furthermore, for any $s \in (0,1]$, there is a sequence Y_s of dimension s such that these bounds are tight.

Observations.

- Even limited to few changes, we can always lower the dimension. I.e., for all $s > 0$ and $\varepsilon > 0$, there is a $t < s$ with $\mathrm{Worst}(s,t) \leqslant \varepsilon$.
- The function is continuous, and even differentiable. *Is there a simple reason this has to be the case?*
- The second case is linear in t. *Why?*

Decreasing dimension: $s = 1/2$

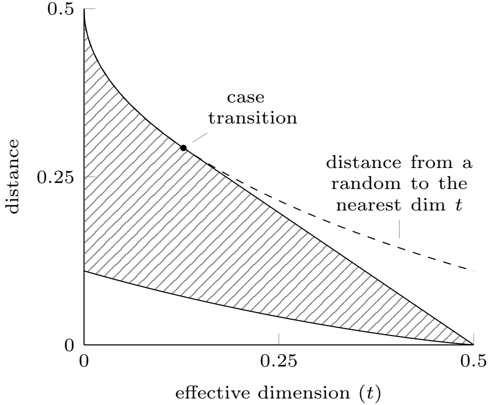


Figure: The distance from a dimension $1/2$ sequence to the nearest dimension $t \leqslant 1/2$ sequence.

Decreasing dimension: $t = 1/2$

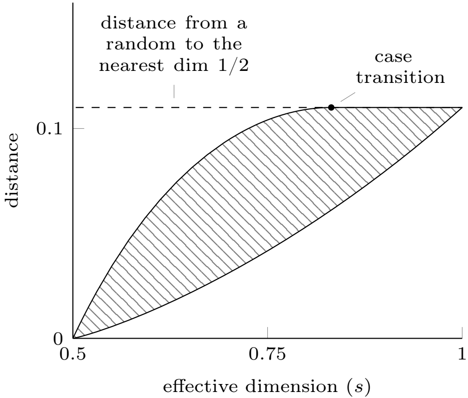


Figure: The distance from a dimension $s \geqslant 1/2$ sequence to the nearest dimension $1/2$ sequence.

Covering codes revisited

How do we code information *robustly*? We need to better understand covering codes. Recall:

Lemma (Delsarte and Piret). For each $r \leqslant n$, there is a covering code $C \subseteq 2^n$ of radius r such that $|C| \approx 2^n / V(n, r)$.

- For $\tau \in 2^n$ and $r \leqslant q \leqslant n$, how many centers from C should we expect to be in the Hamming ball $B_q(\tau)$?
- Each σ is in $V(n, q)$ balls of radius q, so each has a probability of $V(n, q)/2^n$ to be in a randomly chosen Hamming ball of radius q.
- Therefore, on average, we should expect $B_q(\tau)$ to contain around

$$|C| \frac{V(n, q)}{2^n} \approx \frac{V(n, q)}{V(n, r)} \text{ centers from } C.$$

- We want a covering code that is "evenly distributed", i.e., never much worse that this average behavior.

Covering codes revisited, cont.

Such codes exist.

Lemma. For $r \leqslant n$, there is an covering code $C \subseteq 2^n$ of radius r such that $|C| \approx 2^n/V(n,r)$. Furthermore, for every $q \geqslant r$ and every $\tau \in 2^n$, we have

$$|B_q(\tau) \cap C| \lesssim \frac{V(n,q)}{V(n,r)}.$$

(This can be proved using the probabilistic method.)

- Fix s and n. Let C be as in the lemma for $r = H^{-1}(1-s)n$.
- Note that $|C| \approx 2^n/V(n,r) \approx 2^n/2^{(1-s)n} = 2^{sn}$.
- Pick $\sigma \in C$ randomly. In particular, $K(\sigma) \approx sn$.

Claim. This $\sigma \in 2^n$ is robust in the following sense: if we change σ on density $H^{-1}(1-t)$ to get a string τ, where $t \leqslant s$, then $K(\tau) \gtrsim tn$.

Robust coding

Claim. If we change the σ from the previous slide on density $H^{-1}(1-t)$ to get a string τ, where $t \leqslant s$, then $K(\tau) \gtrsim tn$.

Proof. Let $q = H^{-1}(1-t)$. We can determine σ by giving a description of τ and the index of σ in $B_q(\tau) \cap C$. But

$$|B_q(\tau) \cap C| \lesssim V(n,q)/V(n,r) \approx 2^{(1-t)n}/2^{(1-s)n} = 2^{(s-t)n}.$$

Therefore,

$$sn \approx K(\sigma) \lesssim K(\tau) + (s-t)n,$$

hence $K(\tau) \gtrsim tn$. □

Proposition. There is a $\sigma \in 2^n$ such that $K(\sigma) \approx sn$ and if $d(\sigma,\tau) \leqslant H^{-1}(1-t)n$, for any $t \leqslant s$, then $K(\tau) \gtrsim tn$.

Robust coding, cont.

Proposition. There is a $\sigma \in 2^n$ such that $K(\sigma) \approx sn$ and if $d(\sigma, \tau) \leqslant H^{-1}(1-t)n$, for any $t \leqslant s$, then $K(\tau) \gtrsim tn$.

Recall. Every $\sigma \in 2^n$ is within $H^{-1}(1-t)n$ bits of a string τ such that $K(\tau) \lesssim tn$.

> There is a $\sigma \in 2^n$ of complexity sn such that, for $t < s$, it is just as hard to lower the complexity to tn as if σ were random.

As already stated, things are different for infinite sequences, at least when $t < s$ is not too small.

> This is the only case where the result for infinite sequences is not an analogue of the result for strings.

Decreasing dimension: the optimal strategy

Theorem (GoMSoW). If $Y \in 2^\omega$ has dimension s and $0 \leqslant t < s$, then there is an $X \in 2^\omega$ with $\dim X = t$ and

$$d(X, Y) \leqslant \text{Worst}(s, t) := \begin{cases} H^{-1}(1-t) & \text{if } t \leqslant 1 - H(2^{s-1}), \\ \dfrac{t-s}{\log(2^{1-s}-1)} & \text{if } t > 1 - H(2^{s-1}). \end{cases}$$

Furthermore, for any $s \in (0, 1]$, there is a sequence Y_s of dimension s such that these bounds are tight.

Notes.

- The optimal strategy alternates between leaving an interval unchanged to save up changes, then making a lot of changes.
- The ratio of the length of a "savings block" to the corresponding "spending block" and the density of changes needed is a simple optimization problem.

Decreasing dimension: proving optimality

Theorem (GoMSoW). If $Y \in 2^\omega$ has dimension s and $0 \leqslant t < s$, then there is an $X \in 2^\omega$ with $\dim X = t$ and

$$d(X,Y) \leqslant \text{Worst}(s,t) := \begin{cases} H^{-1}(1-t) & \text{if } t \leqslant 1 - H(2^{s-1}), \\ \dfrac{t-s}{\log(2^{1-s}-1)} & \text{if } t > 1 - H(2^{s-1}). \end{cases}$$

Furthermore, for any $s \in (0,1]$, there is a sequence Y_s of dimension s such that these bounds are tight.

Notes.

- To show tightness, we need to construct Y_s.
- This is done by concatenating randomly chosen "robust" strings of dimension s and increasing lengths. (I.e., use the finite result.)
- A convexity argument is used to prove that we can't do better than claimed.

— THANK YOU —