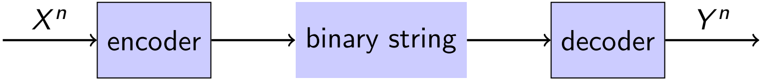
The Minimax Redundancy of Lossy Compression

Adeel Mahmood Aaron Wagner

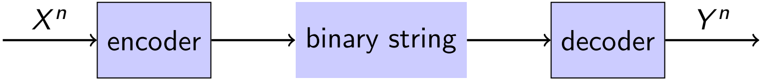
School of Electrical and Computer Engineering
Cornell University

July 2023

Setup

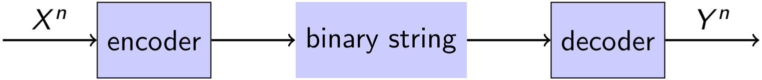


Setup



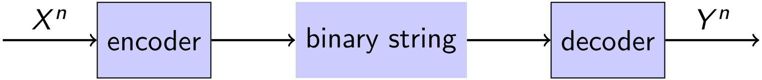
- Source alphabet: A; reconstruction alphabet: B (both finite)

Setup



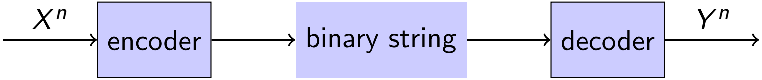
- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with distribution p

Setup



- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with distribution p
- Distortion measure $\rho : A \times B \to [0, \rho_{\max}]$. Assume $\max_{a \in A} \min_{b \in B} \rho(a, b) = 0$.

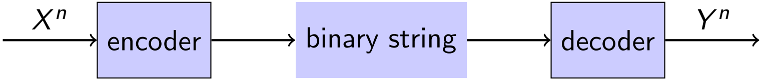
Setup



- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with distribution p
- Distortion measure $\rho : A \times B \to [0, \rho_{\max}]$. Assume $\max_{a \in A} \min_{b \in B} \rho(a, b) = 0$.
- A *d-semifaithful* code [Ornstein and Shields '90] satisfies

$$\rho(X^n, Y^n) := \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, Y_i) \leq d \quad \text{a.s.}$$

Setup

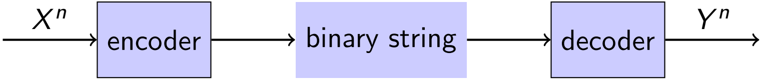


- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with distribution p
- Distortion measure $\rho : A \times B \to [0, \rho_{\max}]$. Assume $\max_{a \in A} \min_{b \in B} \rho(a, b) = 0$.
- A *d-semifaithful* code [Ornstein and Shields '90] satisfies

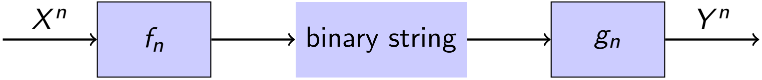
$$\rho(X^n, Y^n) := \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, Y_i) \leq d \quad \text{a.s.}$$

- $\rho =$ Hamming and $d = 0$ recovers the lossless case.

Performance Metric

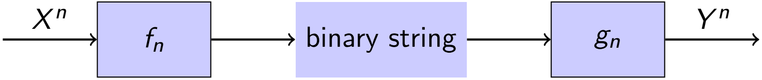


Performance Metric



$$C_n = (f_n, g_n)$$

Performance Metric

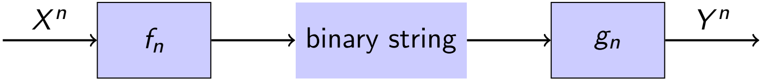


$$C_n = (f_n, g_n)$$

- For prefix-free codes C_n:

$$R(C_n, p) := \frac{\mathbb{E}_p\left[\mathrm{len}(f_n(X^n))\right]}{n}$$

Performance Metric

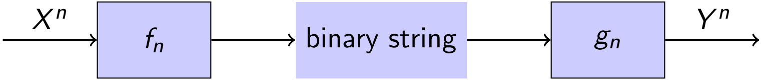


$$C_n = (f_n, g_n)$$

- For prefix-free codes C_n:

$$R(C_n, p) := \frac{\mathbb{E}_p\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{p(y|x):E[\rho(X,Y) \leq d]} I(X;Y)$$

Performance Metric



$$C_n = (f_n, g_n)$$

- For prefix-free codes C_n:

$$R(C_n, p) := \frac{\mathbb{E}_p\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{p(y|x): E[\rho(X,Y) \leq d]} I(X; Y)$$

- Figure-of-merit: the *expected rate redundancy*:

$$R(C_n, p) - R(p, d)$$

Weak and Strong Universality

For any source p, there exists $\{C_n\}$ so that

$$\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0.$$

Weak and Strong Universality

For any source p, there exists $\{C_n\}$ so that

$$\lim_{n \to \infty} R(C_n, p) - R(p, d) = 0.$$

A sequence of codes $\{C_n\}$ is ...

- ... *weakly* or *pointwise* universal if $\{C_n\}$ does not depend on p and

$$\lim_{n \to \infty} R(C_n, p) - R(p, d) = 0 \text{ for all } p.$$

Weak and Strong Universality

For any source p, there exists $\{C_n\}$ so that

$$\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0.$$

A sequence of codes $\{C_n\}$ is ...

- ... *weakly* or *pointwise* universal if $\{C_n\}$ does not depend on p and

$$\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0 \text{ for all } p.$$

- ... *strongly* or *minimax* universal if $\{C_n\}$ does not depend on p and

$$\lim_{n\to\infty} \sup_p \left[R(C_n, p) - R(p, d)\right] = 0.$$

Weak and Strong Universality

For any source p, there exists $\{C_n\}$ so that

$$\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0.$$

A sequence of codes $\{C_n\}$ is ...

- ... *weakly* or *pointwise* universal if $\{C_n\}$ does not depend on p and

$$\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0 \text{ for all } p.$$

- ... *strongly* or *minimax* universal if $\{C_n\}$ does not depend on p and

$$\lim_{n\to\infty} \sup_p \left[R(C_n, p) - R(p, d)\right] = 0.$$

- ... *non-universal* if $\{C_n\}$ depends on p.

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$ [Zhang, Yang, and Wei '97]
Weakly universal	$\Theta\left(\ln n/n\right)$ [Rissanen '84]	$\Theta(\ln n/n)$ [Yang and Zhang (unpublished)]
Strongly universal	$\Theta\left(\ln n/n\right)$ [Davisson *et al.* '81]	

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$ [Zhang, Yang, and Wei '97]
Weakly universal	$\Theta\left(\ln n/n\right)$ [Rissanen '84]	$\Theta(\ln n/n)$ [Yang and Zhang (unpublished)]
Strongly universal	$\Theta\left(\ln n/n\right)$ [Davisson *et al.* '81]	

Other lossy works with $\log n/n$ expected rate redundancy:

- Yu and Speed '93
- Merhav '95

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$ [Zhang, Yang, and Wei '97]
Weakly universal	$\Theta\left(\ln n/n\right)$ [Rissanen '84]	$\Theta(\ln n/n)$ [Yang and Zhang (unpublished)]
Strongly universal	$\Theta\left(\ln n/n\right)$ [Davisson *et al.* '81]	?

Other lossy works with $\log n/n$ expected rate redundancy:

- Yu and Speed '93
- Merhav '95

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta(1/n)$	$\Theta(\ln n/n)$ [Zhang, Yang, and Wei '97]
Weakly universal	$\Theta(\ln n/n)$ [Rissanen '84]	$\Theta(\ln n/n)$ [Yang and Zhang (unpublished)]
Strongly universal	$\Theta(\ln n/n)$ [Davisson *et al.* '81]	?

Other lossy works with $\log n/n$ expected rate redundancy:

- Yu and Speed '93
- Merhav '95

Other works on lossy compression with $\log n/n$ redundancy of some kind: Pilc '68; Linder, G. Lugosi, and K. Zeger '95; Chou, Effros, and Gray '96; Kontoyiannis and Zhang '02; Silva and Piantanida '22 Yang and Zhang '98; Yang and Zhang '99; Mahmood and Wagner '22; Merhav '22

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$ [Zhang, Yang, and Wei '97]
Weakly universal	$\Theta\left(\ln n/n\right)$ [Rissanen '84]	$\Theta(\ln n/n)$ [Yang and Zhang (unpublished)]
Strongly universal	$\Theta\left(\ln n/n\right)$ [Davisson *et al.* '81]	?

Other lossy works with $\log n/n$ expected rate redundancy:

- Yu and Speed '93
- Merhav '95

Other works on lossy compression with $\log n/n$ redundancy of some kind: Pilc '68; Linder, G. Lugosi, and K. Zeger '95; Chou, Effros, and Gray '96; Kontoyiannis and Zhang '02; Silva and Piantanida '22 Yang and Zhang '98; Yang and Zhang '99; Mahmood and Wagner '22; Merhav '22

Another view says that the lossy redundancy is $\tilde{\Theta}(1/\sqrt{n})$ . . . .

A Decomposition

- Consider the *expected rate-distortion function*

$$\mathbb{E}\left[R(P_{X^n}, d)\right],$$

where P_{X^n} is the random type of the i.i.d. source sequence $X^n \sim p$

A Decomposition

- Consider the *expected rate-distortion function*

$$\mathbb{E}\left[R(P_{X^n}, d)\right],$$

where P_{X^n} is the random type of the i.i.d. source sequence $X^n \sim p$

- Conventionally, we upper bound the minimax rate redundancy as

$$\inf_{C_n} \sup_{p} \left[R(C_n, p) - R(p, d)\right] \leq \inf_{C_n} \sup_{p} \left[R(C_n, p) - \mathbb{E}\left[R(P_{X^n}, d)\right]\right] + \sup_{p} \left[\underbrace{\mathbb{E}\left[R(P_{X^n}, d)\right] - R(p, d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \right]$$

Achievability: First Attempt

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

[cf. Kontoyiannis and Zhang '02; Merhav '22]

Achievability: First Attempt

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

   [cf. Kontoyiannis and Zhang '02; Merhav '22]

2. Use acceptance/rejection sampling to identify a subcodebook that is i.i.d. according to the RD-achieving output distribution for P_{x^n}, $Q^{P_{x^n},d,\rho}$.

Achievability: First Attempt

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

   [cf. Kontoyiannis and Zhang '02; Merhav '22]

2. Use acceptance/rejection sampling to identify a subcodebook that is i.i.d. according to the RD-achieving output distribution for P_{x^n}, $Q^{P_{x^n},d,\rho}$.
3. Send the (absolute) index of the first codeword in the subcodebook with distortion at most d.

Achievability: First Attempt

- Classically,

$$-\frac{1}{n} \log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \leq d) \approx R(P_{x^n}, d)$$

Achievability: First Attempt

- Classically,

$$-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \le d) \approx R(P_{x^n}, d)$$

- So the chosen codeword will have an index within the subcodebook of $\approx 2^{nR(P_{x^n},d)}$ ...

Achievability: First Attempt

- Classically,

$$-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \leq d) \approx R(P_{x^n}, d)$$

- So the chosen codeword will have an index within the subcodebook of $\approx 2^{nR(P_{x^n},d)}$ ...
- ... and an index within the overall codebook that is also $\approx 2^{nR(P_{x^n},d)}$ ...

Achievability: First Attempt

- Classically,

$$-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n,Y^n)\le d)\approx R(P_{x^n},d)$$

- So the chosen codeword will have an index within the subcodebook of $\approx 2^{nR(P_{x^n},d)}$ ...
- ... and an index within the overall codebook that is also $\approx 2^{nR(P_{x^n},d)}$ ...
- ... so the binary encoding will have length $nR(P_{x^n},d)$ ...

Achievability: First Attempt

- Classically,

$$-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \leq d) \approx R(P_{x^n}, d)$$

- So the chosen codeword will have an index within the subcodebook of $\approx 2^{nR(P_{x^n},d)}$ ...
- ...and an index within the overall codebook that is also $\approx 2^{nR(P_{x^n},d)}$ ...
- ...so the binary encoding will have length $nR(P_{x^n}, d)$ ...
- ...and the expected rate is $\mathbb{E}[R(P_{X^n}, d)]$.

Achievability: First Attempt

- Classically,

$$-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \le d) \approx R(P_{x^n}, d)$$

- So the chosen codeword will have an index within the subcodebook of $\approx 2^{nR(P_{x^n},d)}$ ...
- ...and an index within the overall codebook that is also $\approx 2^{nR(P_{x^n},d)}$ ...
- ...so the binary encoding will have length $nR(P_{x^n}, d)$ ...
- ...and the expected rate is $\mathbb{E}[R(P_{X^n}, d)]$.

Unfortunate Fact (Mahmood and Wagner '22)

For $|A| = 2$, $|B| = 3$, $\rho_{max} = 3$, $d = 1$ and any even n,

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}(\rho(x^n, Y^n) \le d) - R(P_{x^n}, d, \rho)\right] = \infty.$$

Mitigating Fact (Mahmood and Wagner '22)

Fix $d > 0$. Then

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Achievability

Mitigating Fact (Mahmood and Wagner '22)

Fix $d > 0$. Then

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Proof Idea:

Achievability

Mitigating Fact (Mahmood and Wagner '22)

Fix $d > 0$. Then

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Proof Idea:

1. Change of measure $Q^{P_{x^n},d,\rho}\to Q^*_{Y|X}$ (optimal test channel) using exponential tilting.

Achievability

Mitigating Fact (Mahmood and Wagner '22)

Fix $d > 0$. Then

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le\frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Proof Idea:

1. Change of measure $Q^{P_{x^n},d,\rho}\to Q^*_{Y|X}$ (optimal test channel) using exponential tilting.
2. Split analysis according to variance of $\rho(x^n,Y^n)$ under $Q^*_{Y|X}$:
   - High-variance: Berry-Esseen Theorem
   - Low-variance: Chebyshev's Inequality

Achievability

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

[cf. Kontoyiannis and Zhang '02; Merhav '22]

Achievability

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

   [cf. Kontoyiannis and Zhang '02; Merhav '22]

2. Use acceptance/rejection sampling to identify a subcodebook that is i.i.d. according to the RD-achieving output distribution for P_{x^n}.

Achievability

Given x^n:

1. Generate an infinite, shared codebook with codewords drawn according to the *normalized ML distribution*:

$$Q^{NML}(y^n) = \frac{\max_{q \in \mathcal{P}(B)} q^n(y^n)}{\sum_{z^n \in B^n} \max_{q \in \mathcal{P}(B)} q^n(z^n)}$$

   [cf. Kontoyiannis and Zhang '02; Merhav '22]

2. Use acceptance/rejection sampling to identify a subcodebook that is i.i.d. according to the RD-achieving output distribution for P_{x^n}.
3. Send the (absolute) index of the first codeword in the subcodebook that is within distortion $d + \frac{\text{const}}{n^{5/8}}$.

Post-Correction

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Yields $O(1/n^{5/8})$ redundancy w.r.t. $\mathbb{E}[R(P_{X^n},d)]$. But this is only $d+\frac{\text{const}}{n^{5/8}}$ semifaithful

Post-Correction

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\text{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\text{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Yields $O(1/n^{5/8})$ redundancy w.r.t. $\mathbb{E}[R(P_{X^n},d)]$. But this is only $d+\frac{\text{const}}{n^{5/8}}$ semifaithful

- Recall the assumption

$$\max_{a\in A}\min_{b\in B}\rho(a,b)=0.$$

Post-Correction

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Yields $O(1/n^{5/8})$ redundancy w.r.t. $\mathbb{E}[R(P_{X^n},d)]$. But this is only $d+\frac{\text{const}}{n^{5/8}}$ semifaithful

- Recall the assumption
$$\max_{a\in A}\min_{b\in B}\rho(a,b)=0.$$
- We replace
$$M\propto n^{3/8}$$
symbols in y^n, say $y_{l_1},\ldots,y_{l_M}$, with $\hat{y}_{l_1},\ldots,\hat{y}_{l_M}$ so that $\rho(x_{l_i},\hat{y}_{l_i})=0$ for all $i=1,\ldots,M$.

Post-Correction

$$\sup_{x^n,\rho}\left[-\frac{1}{n}\log Q^{P_{x^n},d,\rho}\left(\rho(x^n,Y^n)\le d+\frac{\mathrm{const}}{n^{5/8}}\right)-R(P_{x^n},d,\rho)\right]$$
$$\le \frac{\mathrm{const}}{n^{5/8}}+O\left(\frac{\log n}{n}\right)$$

Yields $O(1/n^{5/8})$ redundancy w.r.t. $\mathbb{E}[R(P_{X^n},d)]$. But this is only $d+\frac{\text{const}}{n^{5/8}}$ semifaithful

- Recall the assumption

$$\max_{a\in A}\min_{b\in B}\rho(a,b)=0.$$

- We replace

$$M\propto n^{3/8}$$

  symbols in y^n, say $y_{I_1},\ldots,y_{I_M}$, with $\hat{y}_{I_1},\ldots,\hat{y}_{I_M}$ so that $\rho(x_{I_i},\hat{y}_{I_i})=0$ for all $i=1,\ldots,M$.
- Requires sending $n^{3/8}\left[\log n+\log|B|\right]$ bits. Also yields $\approx\frac{1}{n^{5/8}}$ redundancy.
- Meets d constraint a.s.

A Decomposition

- Consider the *expected rate-distortion function*

$$\mathbb{E}\left[R(P_{X^n}, d)\right],$$

where P_{X^n} is the random type of the i.i.d.-p source string X^n

- Conventionally, we upper bound the rate redundancy as

$$\inf_{C_n} \sup_p \left[R(C_n, p) - R(p, d)\right] \leq \inf_{C_n} \sup_p \Big[\overbrace{R(C_n, p) - \mathbb{E}\left[R(P_{X^n}, d)\right]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)}\Big] + \sup_p \Big[\underbrace{\mathbb{E}\left[R(P_{X^n}, d)\right] - R(p, d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}}\Big]$$

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$
Strongly universal	$\Theta\left(\ln n/n\right)$	?

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$
Strongly universal	$\Theta\left(\ln n/n\right)$	$\tilde{O}(1/\sqrt{n})$

Converse

- For the achievability, we used

$$\inf_{C_n} \sup_p [R(C_n, p) - R(p, d)] \le \inf_{C_n} \sup_p \Big[\overbrace{R(C_n, p) - \mathbb{E}[R(P_{X^n}, d)]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)} \Big] + \sup_p \Big[\underbrace{\mathbb{E}[R(P_{X^n}, d)] - R(p, d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \Big]$$

Converse

- For the achievability, we used

$$\inf_{C_n} \sup_p [R(C_n,p) - R(p,d)] \le \inf_{C_n} \sup_p \Big[\overbrace{R(C_n,p) - \mathbb{E}[R(P_{X^n},d)]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)} \Big] + \sup_p \Big[\underbrace{\mathbb{E}[R(P_{X^n},d)] - R(p,d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \Big]$$

- Ignoring $\log n/n$ terms, we have, for any p, d, ρ [Mahmood and Wagner '22]

$$\inf_{C_n} R(C_n,p) \ge \mathbb{E}[R(P_{X^n},d)]$$

Converse

- For the achievability, we used

$$\inf_{C_n} \sup_p [R(C_n,p) - R(p,d)] \le \inf_{C_n} \sup_p \Big[\overbrace{R(C_n,p) - \mathbb{E}[R(P_{X^n},d)]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)} \Big] + \sup_p \Big[\underbrace{\mathbb{E}[R(P_{X^n},d)] - R(p,d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \Big]$$

- Ignoring $\log n/n$ terms, we have, for any p, d, ρ [Mahmood and Wagner '22]

$$\inf_{C_n} R(C_n,p) \ge \mathbb{E}[R(P_{X^n},d)]$$

$$\inf_{C_n} R(C_n,p) - R(p,d) \ge \mathbb{E}[R(P_{X^n},d)] - R(p,d)$$

Converse

- For the achievability, we used

$$\inf_{C_n} \sup_p [R(C_n, p) - R(p, d)] \le \inf_{C_n} \sup_p \Big[\overbrace{R(C_n, p) - \mathbb{E}[R(P_{X^n}, d)]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)} \Big] + \sup_p \Big[\underbrace{\mathbb{E}[R(P_{X^n}, d)] - R(p, d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \Big]$$

- Ignoring $\log n / n$ terms, we have, for any p, d, ρ [Mahmood and Wagner '22]

$$\inf_{C_n} R(C_n, p) \ge \mathbb{E}[R(P_{X^n}, d)]$$

$$\sup_p \inf_{C_n} R(C_n, p) - R(p, d) \ge \sup_p \left[\mathbb{E}[R(P_{X^n}, d)] - R(p, d) \right]$$

Converse

- For the achievability, we used

$$\inf_{C_n} \sup_p \left[R(C_n,p) - R(p,d)\right] \le \inf_{C_n} \sup_p \Big[\overbrace{R(C_n,p) - \mathbb{E}\left[R(P_{X^n},d)\right]}^{\tilde{O}\left(\frac{1}{n^{5/8}}\right)}\Big] + \sup_p \Big[\underbrace{\mathbb{E}\left[R(P_{X^n},d)\right] - R(p,d)}_{\tilde{O}\left(\frac{1}{\sqrt{n}}\right) \text{ using Palaiyanur and Sahai '08}} \Big]$$

- Ignoring $\log n/n$ terms, we have, for any p, d, ρ [Mahmood and Wagner '22]

$$\inf_{C_n} R(C_n,p) \ge \mathbb{E}\left[R(P_{X^n},d)\right]$$

$$\sup_p \inf_{C_n} R(C_n,p) - R(p,d) \ge \sup_p \left[\mathbb{E}\left[R(P_{X^n},d)\right] - R(p,d)\right] \ge \Theta\left(\frac{1}{\sqrt{n}}\right)$$

Converse

- Consider the binary-Hamming case:

$$R(p, d) = (H(p) - H_b(d))^+ \qquad 0 \le d \le 1/2,$$

where $H_b(\cdot)$ denotes binary entropy.

Converse

- Consider the binary-Hamming case:

$$R(p, d) = (H(p) - H_b(d))^+ \qquad 0 \leq d \leq 1/2,$$

where $H_b(\cdot)$ denotes binary entropy.

- Suppose the source is Bernoulli(d). Then $R(p, d) = 0$, but

$$\mathbb{E}\left[R(P_{X^n}, d)\right] = \mathbb{E}\left[\left(H_b\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}_{d \pm \frac{1}{\sqrt{n}}} - H_b(d)]\right)^+\right] = \Theta\left(\frac{1}{\sqrt{n}}\right)$$

Converse

- Consider the binary-Hamming case:

$$R(p,d) = (H(p) - H_b(d))^+ \qquad 0 \le d \le 1/2,$$

where $H_b(\cdot)$ denotes binary entropy.

- Suppose the source is Bernoulli(d). Then $R(p,d) = 0$, but

$$\mathbb{E}\left[R(P_{X^n}, d)\right] = \mathbb{E}\left[\left(H_b\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}_{d \pm \frac{1}{\sqrt{n}}} - H_b(d)\right)^+\right] = \Theta\left(\frac{1}{\sqrt{n}}\right)$$

- End-to-end, for the binary-Hamming case, ignoring log terms:

$$\sup_p \inf_{C_n} R(C_n, p) - R(p,d) \ge \sup_p \left[\mathbb{E}\left[R(P_{X^n}, d)\right] - R(p,d)\right] \ge \Theta\left(\frac{1}{\sqrt{n}}\right)$$

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$
Strongly universal	$\Theta\left(\ln n/n\right)$	$\tilde{O}(1/\sqrt{n})$

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$
Strongly universal	$\Theta\left(\ln n/n\right)$	$\tilde{\Theta}(1/\sqrt{n})$

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$
Strongly universal	$\Theta\left(\ln n/n\right)$	$\tilde{\Theta}(1/\sqrt{n})$

But the binary-Hamming example is non-universal . . . .

$$\sup_{p} \inf_{C_n} R(C_n, p) - R(p, d) \geq \Theta\left(\frac{1}{\sqrt{n}}\right)$$

Redundancy Bounds

	Lossless	Lossy
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$ 🤔
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$ 🤔
Strongly universal	$\Theta\left(\ln n/n\right)$	$\tilde{\Theta}(1/\sqrt{n})$

But the binary-Hamming example is non-universal . . . .

$$\sup_{p} \inf_{C_n} R(C_n, p) - R(p, d) \geq \Theta\left(\frac{1}{\sqrt{n}}\right)$$

Redundancy Bounds

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta(1/n)$	$\Theta(\ln n/n)$	
Weakly universal	$\Theta(\ln n/n)$	$\Theta(\ln n/n)$	
Strongly universal	$\Theta(\ln n/n)$		

Redundancy Bounds

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta(1/n)$	$\Theta(\ln n/n)$	$\Omega(1/\sqrt{n})$
Weakly universal	$\Theta(\ln n/n)$	$\Theta(\ln n/n)$	
Strongly universal	$\Theta(\ln n/n)$		

Redundancy Bounds

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\Omega(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	
Strongly universal	$\Theta(\ln n/n)$		$\tilde{O}(1/\sqrt{n})$

Redundancy Bounds

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$		$\tilde{\Theta}(1/\sqrt{n})$

Redundancy Bounds

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{O}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.
2. The matrix $\mathcal{E}(\lambda^*)$, defined by $[\mathcal{E}(\lambda^*)]_{j,k} = e^{-\lambda^* \rho(j,k)}$ is full column rank.

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.
2. The matrix $\mathcal{E}(\lambda^*)$, defined by $[\mathcal{E}(\lambda^*)]_{j,k} = e^{-\lambda^* \rho(j,k)}$ is full column rank.
3. The source distribution and optimal reconstruction distribution are both full support.

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.
2. The matrix $\mathcal{E}(\lambda^*)$, defined by $[\mathcal{E}(\lambda^*)]_{j,k} = e^{-\lambda^* \rho(j,k)}$ is full column rank.
3. The source distribution and optimal reconstruction distribution are both full support.
4. The determinant of the Jacobian,

$$\frac{\partial F(p, \lambda^*)}{\partial p_{j_1} \partial p_{j_2} \cdots p_{j_{|B|}} \partial \lambda^*},$$

is nonzero for some $1 \leq j_1 < j_2 < \cdots < j_{|B|} \leq |A|$, where F is the vector-valued function

$$F(p, \lambda^*) = \begin{bmatrix} Q^{p,d,\rho}(1) & Q^{p,d,\rho}(2) & \ldots & Q^{p,d,\rho}(|B|) & d \end{bmatrix}.$$

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.
2. The matrix $\mathcal{E}(\lambda^*)$, defined by $[\mathcal{E}(\lambda^*)]_{j,k} = e^{-\lambda^* \rho(j,k)}$ is full column rank.
3. The source distribution and optimal reconstruction distribution are both full support.
4. The determinant of the Jacobian,

$$\frac{\partial F(p, \lambda^*)}{\partial p_{j_1} \partial p_{j_2} \cdots p_{j_{|B|}} \partial \lambda^*},$$

   is nonzero for some $1 \leq j_1 < j_2 < \cdots < j_{|B|} \leq |A|$, where F is the vector-valued function

$$F(p, \lambda^*) = \begin{bmatrix} Q^{p,d,\rho}(1) & Q^{p,d,\rho}(2) & \ldots & Q^{p,d,\rho}(|B|) & d \end{bmatrix}.$$

The binary-Hamming instance is not in this set . . .

Regularity Conditions [Yang & Zhang]

1. $0 < \frac{\partial R(p,d)}{\partial d} := \lambda^* < \infty$.
2. The matrix $\mathcal{E}(\lambda^*)$, defined by $[\mathcal{E}(\lambda^*)]_{j,k} = e^{-\lambda^* \rho(j,k)}$ is full column rank.
3. The source distribution and optimal reconstruction distribution are both full support.
4. The determinant of the Jacobian,

$$\frac{\partial F(p, \lambda^*)}{\partial p_{j_1} \partial p_{j_2} \cdots p_{j_{|B|}} \partial \lambda^*},$$

   is nonzero for some $1 \leq j_1 < j_2 < \cdots < j_{|B|} \leq |A|$, where F is the vector-valued function

$$F(p, \lambda^*) = \begin{bmatrix} Q^{p,d,\rho}(1) & Q^{p,d,\rho}(2) & \ldots & Q^{p,d,\rho}(|B|) & d \end{bmatrix}.$$

The binary-Hamming instance is not in this set ... but it is in the closure.

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{O}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

- A new approach to the study of universal lossy compression . . .

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta(1/n)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta(\ln n/n)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

- A new approach to the study of universal lossy compression ...
  - ... without regularity assumptions

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

- A new approach to the study of universal lossy compression ...
  - ...without regularity assumptions
  - ...with strong universality (cf. lossless)

Conclusion

	Lossless	Lossy (d-semifaithful)	
		with regularity cond.	without regularity cond.
Non-universal	$\Theta\left(1/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Weakly universal	$\Theta\left(\ln n/n\right)$	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$
Strongly universal	$\Theta(\ln n/n)$	$\tilde{\Theta}(1/\sqrt{n})$	$\tilde{\Theta}(1/\sqrt{n})$

- A new approach to the study of universal lossy compression ...
  - ... without regularity assumptions
  - ... with strong universality (cf. lossless)
- Such an approach yields different redundancies and gives rise to new schemes.