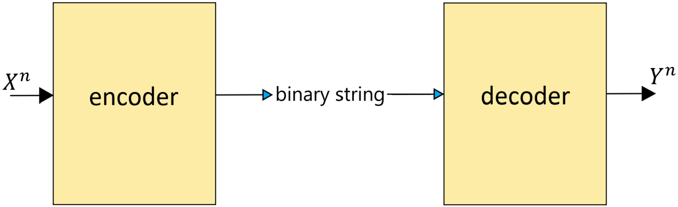
Lossy Compression with Universal Distortion

Adeel Mahmood   Aaron B. Wagner

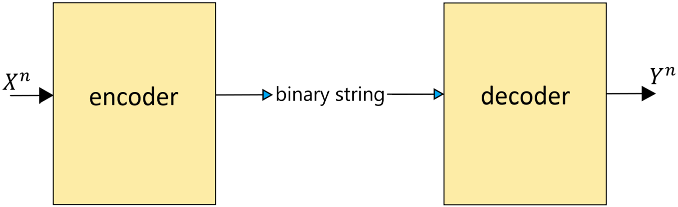
Cornell University, Electrical and Computer Engineering

ISIT 2022

Setup

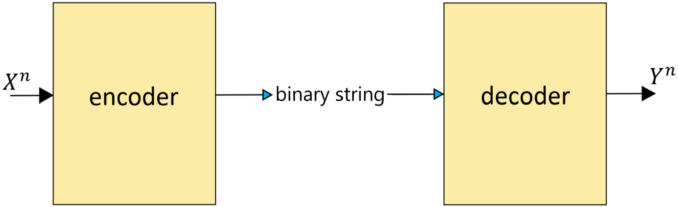


Setup



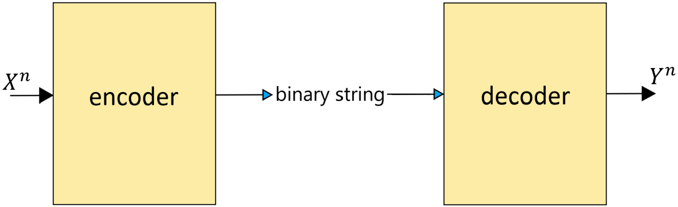
- Source alphabet: A; reconstruction alphabet: B (both finite)

Setup



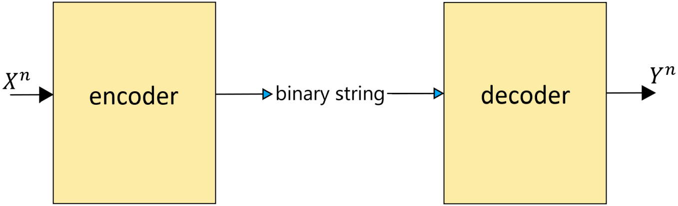
- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with dist. p

Setup



- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with dist. p
- Distortion measure $\rho : A \times B \to [0, \infty)$.

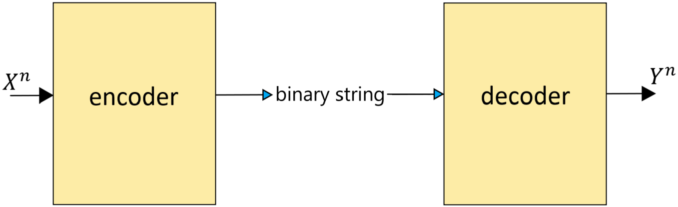
Setup



- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with dist. p
- Distortion measure $\rho : A \times B \to [0, \infty)$.
- A *d-semifaithful code* [Ornstein and Shields '90] satisfies

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

Setup



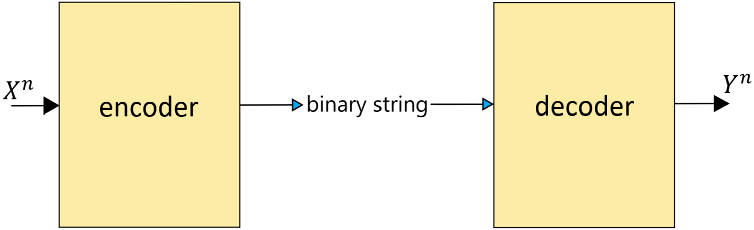
- Source alphabet: A; reconstruction alphabet: B (both finite)
- Source X^n: i.i.d. with dist. p
- Distortion measure $\rho : A \times B \to [0, \infty)$.
- A *d-semifaithful code* [Ornstein and Shields '90] satisfies

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

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

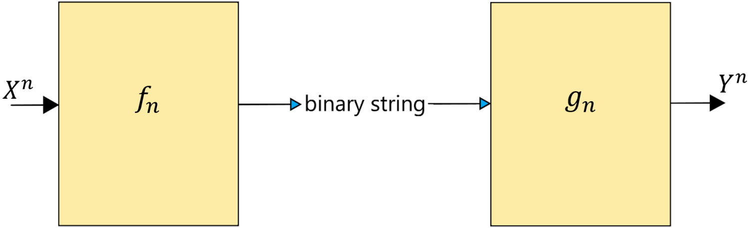
Performance Metric

Operational Rate Redundancy



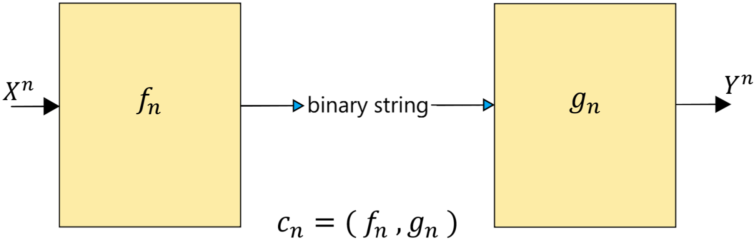
Performance Metric

Operational Rate Redundancy



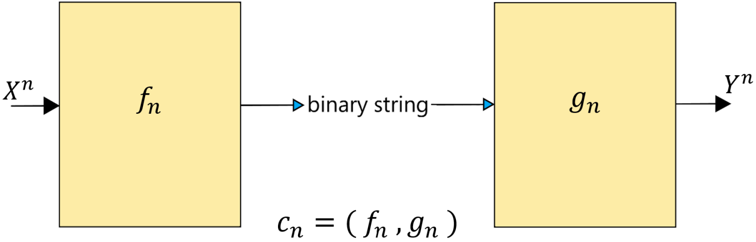
Performance Metric

Operational Rate Redundancy



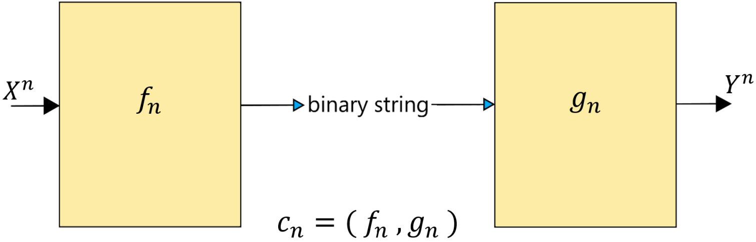
Performance Metric

Operational Rate Redundancy



- The expected rate of c_n is

$$R(c_n, p, d, \rho) \triangleq \frac{\mathbb{E}\left[l(f_n(X^n))\right]}{n}$$



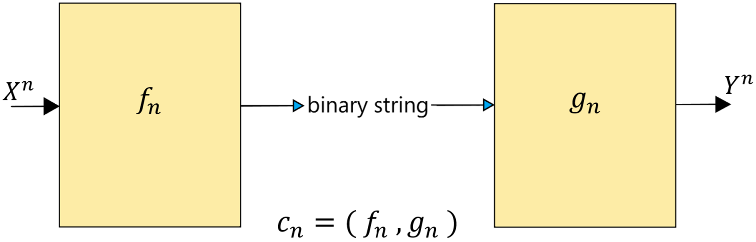
- The expected rate of c_n is

$$R(c_n, p, d, \rho) \triangleq \frac{\mathbb{E}\left[l(f_n(X^n))\right]}{n}$$

- $Y^n = g_n(f_n(X^n))$

Performance Metric

Operational Rate Redundancy



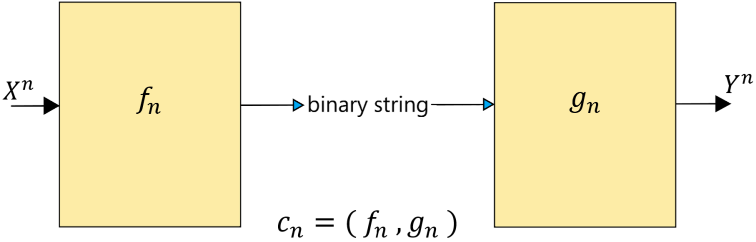
- The expected rate of c_n is

$$R(c_n, p, d, \rho) \triangleq \frac{\mathbb{E}\left[l(f_n(X^n))\right]}{n}$$

- $Y^n = c_n(X^n)$.

Performance Metric

Operational Rate Redundancy



- The expected rate of a prefix-free c_n is

$$R(c_n, p, d, \rho) \triangleq \frac{\mathbb{E}\left[l(f_n(X^n))\right]}{n} \geq \frac{H_{c_n,p}(Y^n)}{n}$$

- $Y^n = c_n(X^n)$.

Operational Rate Redundancy:

$$R(c_n, p, d, \rho) - \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Operational Rate Redundancy:

$$R(c_n, p, d, \rho) - \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Note that

$$\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} \approx \inf_{c_n \in \mathcal{C}_{d,\rho}} R(c_n, p, d, \rho)$$

Operational Rate Redundancy:

$$R(c_n, p, d, \rho) - \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Note that

$$\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} R(c_n, p, d, \rho) < \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} + \frac{1}{n}$$

Operational Rate Redundancy:

$$R(c_n, p, d, \rho) - \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Note that

$$\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} R(c_n, p, d, \rho) < \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} + \frac{1}{n}$$

Silva and Piantanida '21

Universal d-semifaithful codes

Universal d-semifaithful codes

- The source p is unknown.

Universal d-semifaithful codes

- The source p is unknown.
- For all (ρ, d), there exist (weakly) universal codes $\{c_n\}$ satisfying

$$\lim_{n\to\infty}\left[R(c_n, p, d, \rho) - \inf_{\tilde{c}_n\in\mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n,p}(Y^n)}{n}\right] = 0 \quad \text{for all } p$$

Universal d-semifaithful codes

- The source p is unknown.
- For all (ρ, d), there exist (weakly) universal codes $\{c_n\}$ satisfying

$$R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n, p}(Y^n)}{n} = O\left(\frac{\ln n}{n}\right) \quad \text{for all } p$$

Universal d-semifaithful codes

- The source p is unknown.
- For all (ρ, d), there exist (weakly) universal codes $\{c_n\}$ satisfying

$$R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n, p}(Y^n)}{n} = O\left(\frac{\ln n}{n}\right) \quad \text{for all } p$$

- Strongly (or minimax) universal codes $\{c_n\}$ also exist:

$$\sup_p \left[R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n, p}(Y^n)}{n}\right] = O\left(\frac{\ln n}{n}\right)$$

Universal d-semifaithful codes

- The source p is unknown.
- For all (ρ, d), there exist (weakly) universal codes $\{c_n\}$ satisfying

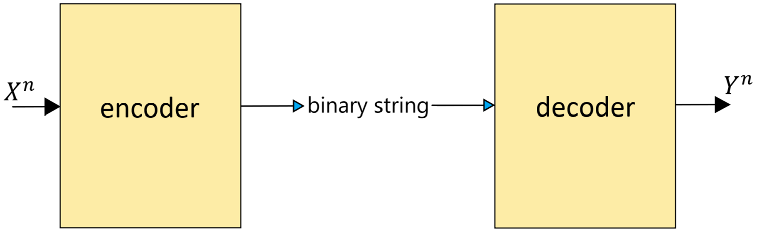
$$R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n,p}(Y^n)}{n} = O\left(\frac{\ln n}{n}\right) \quad \text{for all } p$$

- Strongly (or minimax) universal codes $\{c_n\}$ also exist:

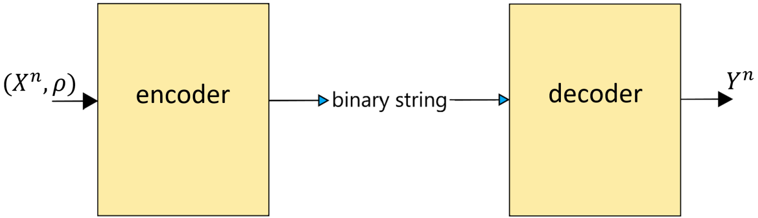
$$\sup_p \left[R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n,p}(Y^n)}{n}\right] = O\left(\frac{\ln n}{n}\right)$$

- We consider a more general universal framework

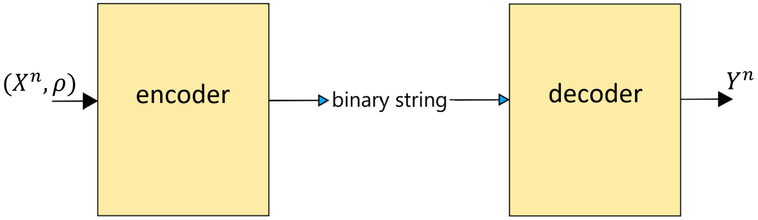
Universal Distortion Framework



Universal Distortion Framework

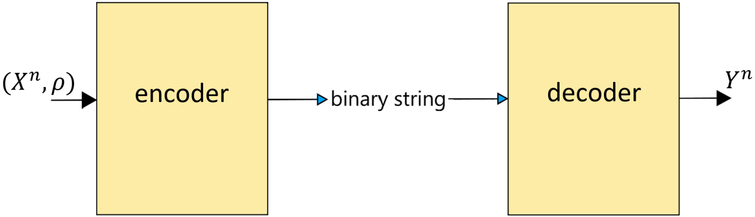


Universal Distortion Framework



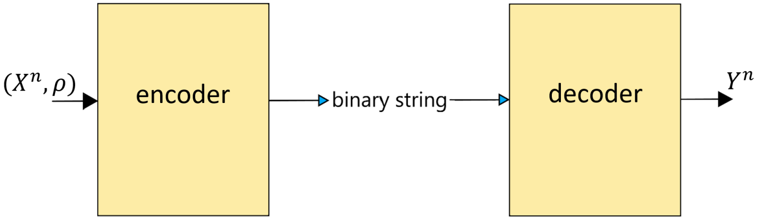
The distortion measure is revealed only to the encoder and only at runtime.

Universal Distortion Framework



The distortion measure is revealed only to the encoder and only at runtime.

Universal Distortion Framework

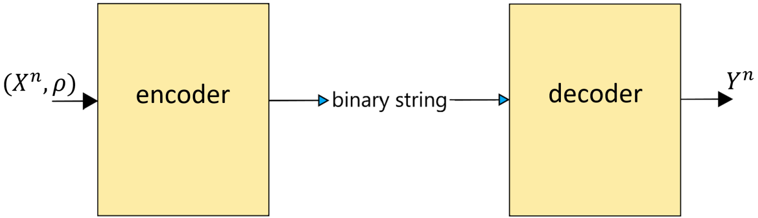


The distortion measure is revealed only to the encoder and only at runtime.

Related Work:

- Martinian, Wornell and Zamir, 2008

Universal Distortion Framework



The distortion measure is revealed only to the encoder and only at runtime.

Related Work:

- Martinian, Wornell and Zamir, 2008
- Merhav, 2022 (arXiv)

Why Universal Distortion?

- Compression systems are typically asked to meet the needs of a variety of end-users who may have discordant notions of distortion.

Why Universal Distortion?

- Compression systems are typically asked to meet the needs of a variety of end-users who may have discordant notions of distortion.

Original Image

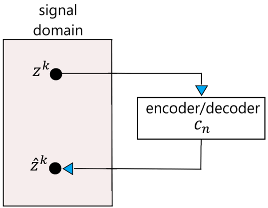
Output Image from Neural Network

Creatism, a deep-learning system for artistic content creation, Fang and Zhang '17

Deep Generative Models for Distribution-Preserving Lossy Compression, Tschannen *et al.* '18

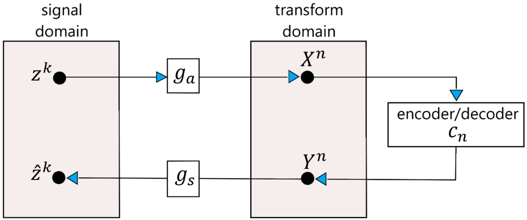
Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



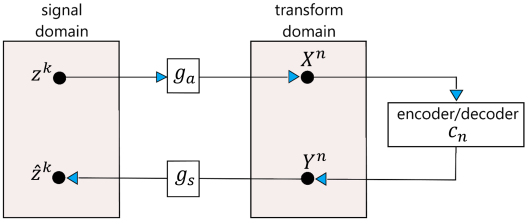
Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



Why Universal Distortion?

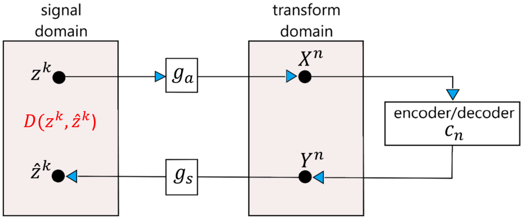
Nonlinear Transform Coding (Ballé *et al.* '20)



$$x^n = g_a(z^k)$$
$$y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

Why Universal Distortion?

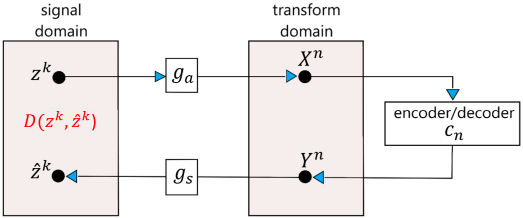
Nonlinear Transform Coding (Ballé *et al.* '20)



$$x^n = g_a(z^k)$$
$$y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)

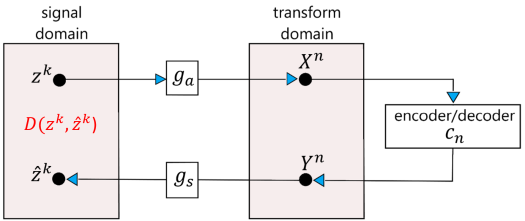


$$x^n = g_a(z^k)$$
$$\rightarrow y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



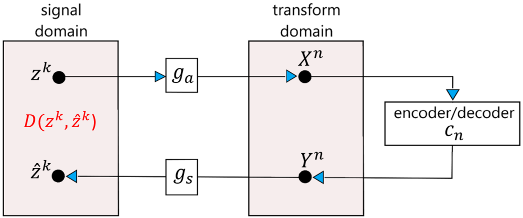
$$x^n = g_a(z^k)$$
$$y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



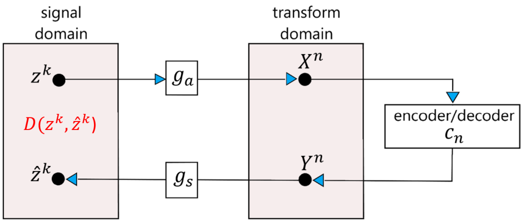
$$x^n = g_a(z^k)$$
$$y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.
- If (g_a, g_s) are nonlinear, then choose y^n according to the local behavior of $D(z^k, \hat{z}^k)$ in the signal domain.

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



$$x^n = g_a(z^k)$$
$$y^n = c_n(x^n)$$
$$\hat{z}^k = g_s(y^n)$$

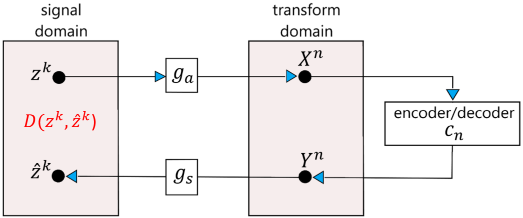
Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.
- If (g_a, g_s) are nonlinear, then choose y^n according to the local behavior of $D(z^k, \hat{z}^k)$ in the signal domain.

$$D(z^k, \hat{z}^k) = D(z^k, g_s(y^n))$$

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



$$
\begin{aligned}
x^n &= g_a(z^k)\\
y^n &= c_n(x^n)\\
\hat{z}^k &= g_s(y^n)
\end{aligned}
$$

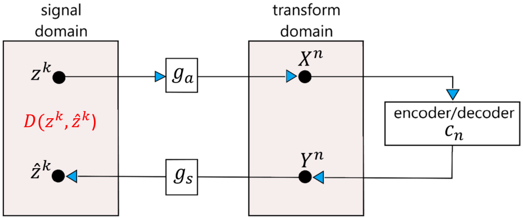
Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.
- If (g_a, g_s) are nonlinear, then choose y^n according to the local behavior of $D(z^k, \hat{z}^k)$ in the signal domain.

$$
\begin{aligned}
D(z^k, \hat{z}^k) &= D(z^k, g_s(y^n))\\
&\approx D(z^k, g_s(x^n)) + \nabla D(z^k, g_s(x^n))(y^n - x^n) + \frac{1}{2}(y^n - x^n)^T \nabla^2 D(z^k, g_s(x^n))(y^n - x^n)
\end{aligned}
$$

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



$$\begin{aligned} x^n &= g_a(z^k) \\ y^n &= c_n(x^n) \\ \hat{z}^k &= g_s(y^n) \end{aligned}$$

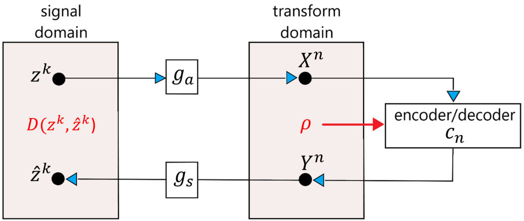
Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.
- If (g_a, g_s) are nonlinear, then choose y^n according to the local behavior of $D(z^k, \hat{z}^k)$ in the signal domain.

$$\begin{aligned} D(z^k, \hat{z}^k) &= D(z^k, g_s(y^n)) \\ &\approx D(z^k, g_s(x^n)) + \cancelto{0}{\nabla D(z^k, g_s(x^n))(y^n - x^n)} + \frac{1}{2}(y^n - x^n)^T \nabla^2 D(z^k, g_s(x^n))(y^n - x^n) \end{aligned}$$

Why Universal Distortion?

Nonlinear Transform Coding (Ballé *et al.* '20)



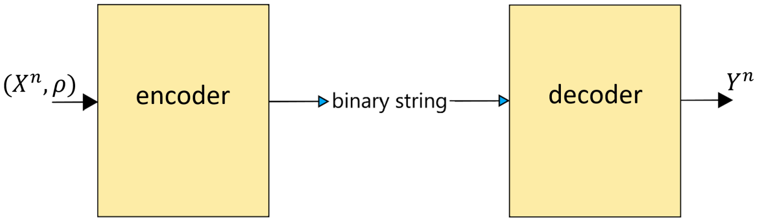
$$
\begin{aligned}
x^n &= g_a(z^k) \\
y^n &= c_n(x^n) \\
\hat{z}^k &= g_s(y^n)
\end{aligned}
$$

Goal: For a fixed (g_a, g_s) and a fixed codebook for c_n, minimize $D(z^k, \hat{z}^k)$ over the choice of y^n.

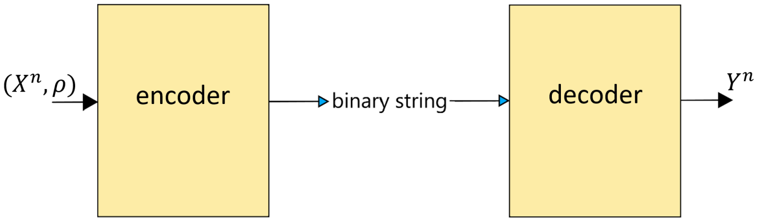
- If (g_a, g_s) are linear, orthogonal transforms, then choose the y^n closest to x^n.
- If (g_a, g_s) are nonlinear, then choose y^n according to the local behavior of $D(z^k, \hat{z}^k)$ in the signal domain.

$$
\begin{aligned}
D(z^k, \hat{z}^k) &= D(z^k, g_s(y^n)) \\
&\approx D(z^k, g_s(x^n)) + \cancelto{0}{\nabla D(z^k, g_s(x^n))}(y^n - x^n) + \frac{1}{2}(y^n - x^n)^T \nabla^2 D(z^k, g_s(x^n))(y^n - x^n)
\end{aligned}
$$

Universal Distortion Framework



Universal Distortion Framework



- Quantization Approach
- VC Dimension Approach

The Naive Quantization Approach . . .

The Naive Quantization Approach . . .

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.

The Naive Quantization Approach ...

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The Naive Quantization Approach . . .

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The quantization $[\rho]$ of ρ will satisfy for all $a \in A$, $b \in B$

- $\big|[\rho](a, b) - \rho(a, b)\big| < \frac{d}{n}$.

The Naive Quantization Approach ...

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The quantization $[\rho]$ of ρ will satisfy for all $a \in A$, $b \in B$

- $\big|[\rho](a,b) - \rho(a,b)\big| < \frac{d}{n}$.

There are $\left(\frac{\rho_{\mathsf{max}}}{d} n\right)^{|A||B|}$ quantized distortion measures.

The Naive Quantization Approach . . .

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The quantization $[\rho]$ of ρ will satisfy for all $a \in A$, $b \in B$

- $\big|[\rho](a, b) - \rho(a, b)\big| < \frac{d}{n}$.

There are $\left(\frac{\rho_{\mathsf{max}}}{d} n\right)^{|A||B|}$ quantized distortion measures.

The Naive Quantization Approach . . .

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The quantization $[\rho]$ of ρ will satisfy for all $a \in A$, $b \in B$

- $\big|[\rho](a, b) - \rho(a, b)\big| < \frac{d}{n}$.
-

There are $\left(\frac{\rho_{\mathsf{max}}}{d} n\right)^{|A||B|}$ quantized distortion measures.

The Naive Quantization Approach ...

- Assume $\rho : A \times B \to [0, \rho_{\mathsf{max}}]$.
- We implement $\frac{d}{n}$-quantization of ρ.

The quantization $[\rho]$ of ρ will satisfy for all $a \in A$, $b \in B$

- $\big|[\rho](a, b) - \rho(a, b)\big| < \frac{d}{n}$.
- $[\rho](a, b) \leq \rho(a, b)$.

There are $\left(\frac{\rho_{\mathsf{max}}}{d} n\right)^{|A||B|}$ quantized distortion measures.

Convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$.

Three-Stage Coding Scheme

Given an input (x^n, ρ)

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log \left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n \right)^{|A||B|} \right) \quad \text{bits.}$$

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log\left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n\right)^{|A||B|} \right) \quad \text{bits.}$$

2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n}$$

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log \left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n \right)^{|A||B|} \right) \quad \text{bits.}$$

2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using
$$\approx \log \left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n \right)^{|A||B|} \right) \quad \text{bits.}$$
2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying
$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$
$$\mathbb{E}_p[R(c_n^*, t, d, [\rho])] \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log\left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n\right)^{|A||B|} \right) \quad \text{bits.}$$

2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

$$\mathbb{E}_p[R(c_n^*, t, d, [\rho])] \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log\left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n\right)^{|A||B|}\right) \quad \text{bits.}$$

2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

$$\mathbb{E}_p[R(c_n^*, t, d, [\rho])] \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

3. The encoding is not yet d-semifaithful!

Given an input (x^n, ρ)

1. Identify the type t and quantized ρ to the decoder using

$$\approx \log \left(n^{|A|} \times \left(\frac{\rho_{\max}}{d} n \right)^{|A||B|} \right) \quad \text{bits.}$$

2. For the t and $[\rho]$, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, [\rho]) \approx \inf_{c_n \in \mathcal{C}_{d,[\rho]}} \frac{H_{c_n,t}(Y^n)}{n} \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

$$\mathbb{E}_p[R(c_n^*, t, d, [\rho])] \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

3. The encoding is not yet d-semifaithful! Post-correction

Convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$.

Post-Correction

$$[\rho](x^n, c_n^*(x^n)) \leq d \;\not\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \leq d$$

Convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$.

Post-Correction

$$[\rho](x^n, c_n^*(x^n)) \le d \;\not\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \le d$$

However,

$$[\rho](x^n, c_n^*(x^n)) \le d \;\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \le d + \frac{d}{n}$$

Convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$.

Post-Correction

$$[\rho](x^n, c_n^*(x^n)) \leq d \;\not\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \leq d$$

However,

$$[\rho](x^n, c_n^*(x^n)) \leq d \;\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \leq d + \frac{d}{n}$$

$$\text{Recall}: \; \max_{a \in A} \min_{b \in B} \rho(a, b) = 0$$

$$[\rho](x^n, c_n^*(x^n)) \leq d \;\not\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \leq d$$

However,

$$[\rho](x^n, c_n^*(x^n)) \leq d \;\Longrightarrow\; \rho(x^n, c_n^*(x^n)) \leq d + \frac{d}{n}$$

$$\text{Recall}: \; \max_{a \in A} \min_{b \in B} \rho(a, b) = 0$$

One symbol post-correction takes $\log(n) + \log(|B|)$ bits.

Convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$.

Theorem (Mahmood and Wagner '22)

There exists a universal distortion d-semifaithful code c_n satisfying

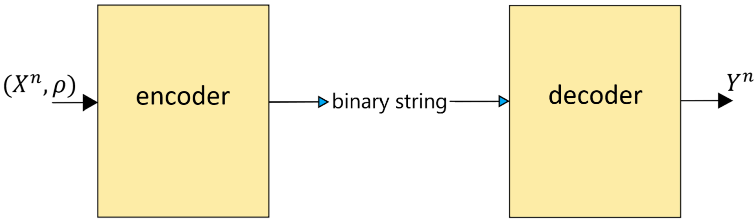
$$\sup_{p,\rho} \left[R(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n,p}(Y^n)}{n} \right] \leq O\left(\frac{\ln n}{n}\right).$$

Problems with the Quantization Approach

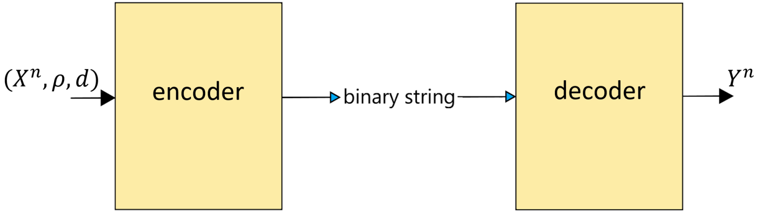
- Works only for uniformly bounded distortion measures.
- Convergence not uniform over d.

The VC dimension approach

The VC dimension approach



The VC dimension approach



- For sequences x^n and y^n, we have

$$\rho(x^n, y^n) = \sum_{a \in A, b \in B} s(a, b)\rho(a, b),$$

where s is the joint n-type.

Distortion Measures as Linear Classifiers

- For sequences x^n and y^n, we have

$$\rho(x^n, y^n) = \sum_{a \in A, b \in B} s(a, b)\rho(a, b),$$

  where s is the joint n-type.

- For any given ρ and d, define a linear classifier

$$\mathcal{L}_{\rho,d} : \mathcal{P}_n(A \times B) \to \{-1, +1\}$$

  as

Distortion Measures as Linear Classifiers

- For sequences x^n and y^n, we have

$$\rho(x^n, y^n) = \sum_{a \in A, b \in B} s(a, b)\rho(a, b),$$

  where s is the joint n-type.

- For any given ρ and d, define a linear classifier

$$\mathcal{L}_{\rho,d} : \mathcal{P}_n(A \times B) \to \{-1, +1\}$$

  as

$$\mathcal{L}_{\rho,d}(s) = \begin{cases} +1 & \text{if} \quad \sum_{a,b} s(a, b)\rho(a, b) \le d \\ -1 & \text{if} \quad \sum_{a,b} s(a, b)\rho(a, b) > d. \end{cases}$$

Equivalence Relation

Equivalence Relation

- We say $(\rho, d) \sim (\rho', d')$ if

$$\mathcal{L}_{\rho,d}(s) = \mathcal{L}_{\rho',d'}(s) \quad \text{for all joint types } s.$$

Equivalence Relation

- We say $(\rho, d) \sim (\rho', d')$ if

$$\mathcal{L}_{\rho,d}(s) = \mathcal{L}_{\rho',d'}(s) \quad \text{for all joint types } s.$$

- If $(\rho, d) \sim (\rho', d')$, then for all x^n and y^n,

$$\rho(x^n, y^n) \leq d \iff \rho'(x^n, y^n) \leq d'$$

Equivalence Relation

- We say $(\rho, d) \sim (\rho', d')$ if

$$\mathcal{L}_{\rho,d}(s) = \mathcal{L}_{\rho',d'}(s) \quad \text{for all joint types } s.$$

- If $(\rho, d) \sim (\rho', d')$, then for all x^n and y^n,

$$\rho(x^n, y^n) \leq d \iff \rho'(x^n, y^n) \leq d'$$

- A universal code designed for (ρ, d) will work equally well for any other (ρ', d') in the same equivalence class.

Equivalence Relation

- We say $(\rho, d) \sim (\rho', d')$ if

$$\mathcal{L}_{\rho,d}(s) = \mathcal{L}_{\rho',d'}(s) \quad \text{for all joint types } s.$$

- If $(\rho, d) \sim (\rho', d')$, then for all x^n and y^n,

$$\rho(x^n, y^n) \leq d \iff \rho'(x^n, y^n) \leq d'$$

- A universal code designed for (ρ, d) will work equally well for any other (ρ', d') in the same equivalence class.

A different notion of equivalence in Stjernvall '83

How many equivalence classes?

The VC dimension of the set of linear classifiers in $|A||B|$-dimensional space is equal to

$$|A||B| + 1$$

How many equivalence classes?

The VC dimension of the set of linear classifiers in $|A||B|$-dimensional space is equal to

$$|A||B| + 1 \qquad \text{(Vapnik '98)}$$

How many equivalence classes?

The VC dimension of the set of linear classifiers in $|A||B|$-dimensional space is equal to

$$|A||B| + 1 \qquad\qquad \text{(Vapnik '98)}$$

How many equivalence classes?

The VC dimension of the set of linear classifiers in $|A||B|$-dimensional space is equal to

$$|A||B| + 1 \qquad \text{(Vapnik '98)}$$

$$\#\text{ of Equivalence classes} \leq |\mathcal{P}_n(A \times B)|^{|A||B|+1}$$

How many equivalence classes?

The VC dimension of the set of linear classifiers in $|A||B|$-dimensional space is equal to

$$|A||B| + 1 \qquad\qquad (\text{Vapnik '98})$$

$$\begin{aligned}\#\text{ of Equivalence classes} &\leq |\mathcal{P}_n(A \times B)|^{|A||B|+1} \\ &\leq (n+1)^{|A|^2|B|^2-1}\end{aligned}$$

Two-Stage Coding Scheme

Given an input (x^n, ρ, d)

Two-Stage Coding Scheme

Given an input (x^n, ρ, d)

1. Identify the type t and equivalence class to the decoder using

$$\approx \log\left(n^{|A|} \times n^{|A|^2|B|^2-1}\right) \quad \text{bits.}$$

Two-Stage Coding Scheme

Given an input (x^n, ρ, d)

1. Identify the type t and equivalence class to the decoder using

$$\approx \log \left(n^{|A|} \times n^{|A|^2|B|^2-1} \right) \quad \text{bits.}$$

2. For the given type and equivalence class, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, \rho) \approx \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

Two-Stage Coding Scheme

Given an input (x^n, ρ, d)

1. Identify the type t and equivalence class to the decoder using

$$\approx \log\left(n^{|A|} \times n^{|A|^2|B|^2-1}\right) \quad \text{bits.}$$

2. For the given type and equivalence class, use a near-optimal code c_n^* satisfying

$$R(c_n^*, t, d, \rho) \approx \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,t}(Y^n)}{n}$$

$$\mathbb{E}_p[R(c_n^*, t, d, \rho)] \leq \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$$

Minimax convergence to $\inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n}$

Theorem (Mahmood and Wagner)

In the generalized universal distortion setting,

$$\inf_{c_n} \sup_{p,d,\rho} \left[R_n(c_n, p, d, \rho) - \inf_{\tilde{c}_n \in \mathcal{C}_{d,\rho}} \frac{H_{\tilde{c}_n,p}(Y^n)}{n} \right] \leq O\left(\frac{\ln n}{n} \right),$$

where the infimum is over all codes which meet the input distortion constraint with respect to the input distortion measure.

Final Thoughts

Where is the Rate-Distortion Function?

- $R(p, d, \rho)$ is the single-letter information theoretic limit of the multi-letter entropy target:

$$\lim_{n \to \infty} \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} = R(p, d, \rho) \quad \text{(Kieffer '78)}$$

Final Thoughts

Where is the Rate-Distortion Function?

- $R(p, d, \rho)$ is the single-letter information theoretic limit of the multi-letter entropy target:

$$\lim_{n\to\infty} \inf_{c_n \in \mathcal{C}_{d,\rho}} \frac{H_{c_n,p}(Y^n)}{n} = R(p, d, \rho) \quad \text{(Kieffer '78)}$$

- Convergence to RD function that is minimax over both p and ρ is possible using different techniques - next talk!

References

A. Mahmood and A. B. Wagner, "Lossy compression with universal distortion," 2021. [Online]. Available: arXiv:2110.07022

Thank you for listening!

Adeel Mahmood

am2384@cornell.edu