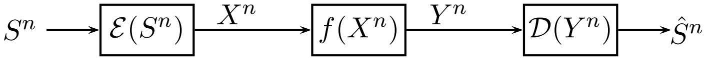
Subset–Universal Lossy Compression

Or Ordentlich
Joint work with Ofer Shayevitz

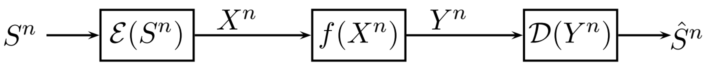
Information Theory Workshop

Jerusalem, Israel
April 28, 2015

Motivation - JSCC over Deterministic Channel

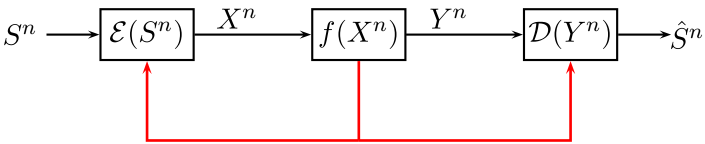


MOTIVATION - JSCC OVER DETERMINISTIC CHANNEL



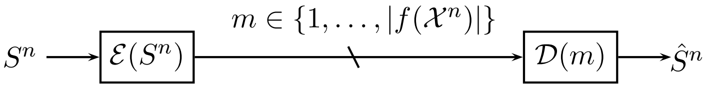
- S^n is a discrete memoryless source with PMF P_S
- $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_+$ a bounded distortion measure
- $f : \mathcal{X}^n \mapsto \mathcal{Y}^n$ is a deterministic channel (NOT memoryless in general)
  Example: Memory block with some cells stuck at 0 or 1, some cells that flip bits and some good cells
- $f(\mathcal{X}^n) \subseteq \mathcal{Y}^n$ is the image of f
- $\mathcal{E}$ and $\mathcal{D}$ are the encoder and decoder

MOTIVATION - JSCC OVER DETERMINISTIC CHANNEL



CSI@Both

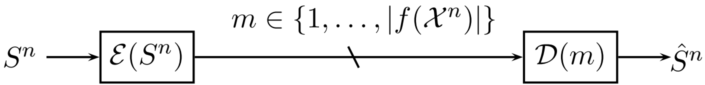
Motivation - JSCC over Deterministic Channel



CSI@Both

- The channel becomes a bit pipe of rate $nR = \log |f(\mathcal{X}^n)|$

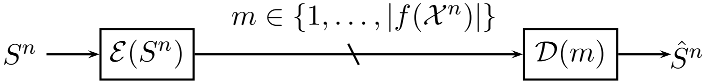
Motivation - JSCC Over Deterministic Channel



CSI@Both

- The channel becomes a bit pipe of rate $nR = \log|f(\mathcal{X}^n)|$
- Separation is optimal

Motivation - JSCC Over Deterministic Channel

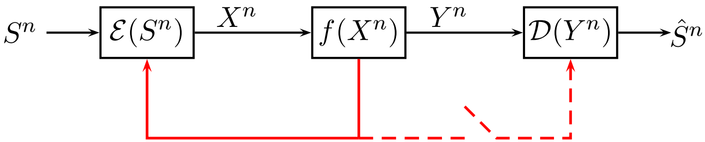


CSI@Both

- The channel becomes a bit pipe of rate $nR = \log|f(\mathcal{X}^n)|$
- Separation is optimal
- Smallest achievable distortion is $D_{P_S}\left(\frac{1}{n}\log|f(\mathcal{X}^n)|\right)$, where

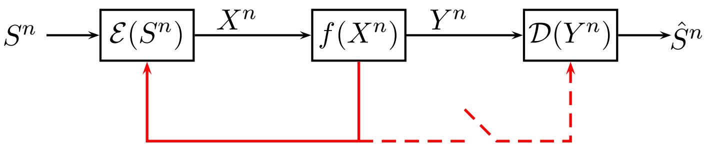
$$D_{P_S}(R) \triangleq \min_{P_{\hat{S}|S}: I(S;\hat{S}) \le R} \sum_{s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}}} P_S(s) P_{\hat{S}|S}(\hat{s}|s) d(s, \hat{s}).$$

Motivation - JSCC over Deterministic Channel



CSI@Tx Only

Motivation - JSCC over Deterministic Channel

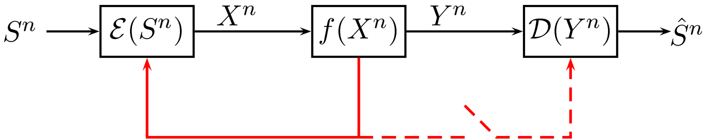


CSI@Tx Only

Channels from a limited class

- In some cases it is possible to learn the channel with small overhead
  - $f(X^n) = [h(X_1), h(X_2), \ldots, h(X_n)]$
  - f may have some other “sparse” structure
- Separation (+training) is optimal and achieves $D_{P_S}\left(\frac{1}{n}\log|f(\mathcal{X}^n)|\right)$

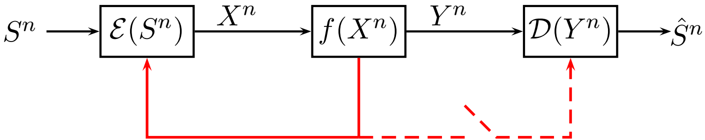
Motivation - JSCC over Deterministic Channel



CSI@Tx Only

Gelfand-Pinsker

Motivation - JSCC over Deterministic Channel



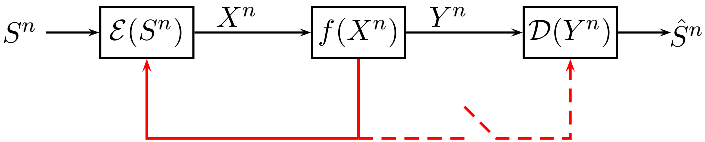
CSI@Tx Only
Gelfand-Pinsker

- If the channel is memoryless with state

$$f(X^n) = [h_{T_1}(X_1), h_{T_2}(X_2), \dots, h_{T_n}(X_n)],$$

where $\{T_i\}$ is an i.i.d. state process, separation (Gelfand-Pinsker + source coding) is optimal (Merhav-Shamai 03) and achieves $D_{P_S}\left(\frac{1}{n}\log|f(\mathcal{X}^n)|\right)$

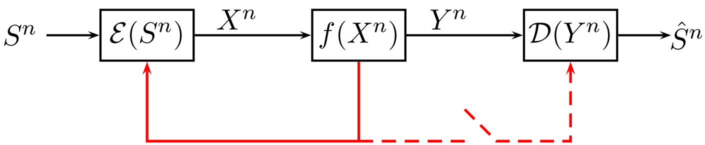
Motivation - JSCC over Deterministic Channel



CSI@Tx Only

What if f is an arbitrary mapping from $\mathcal{X}^n$ to $\mathcal{Y}^n$?

Motivation - JSCC over Deterministic Channel

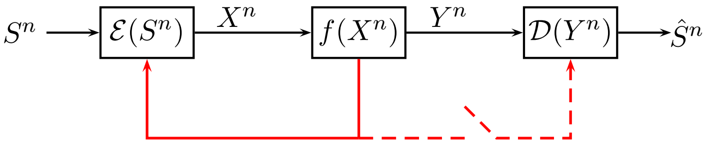


CSI@Tx Only

What if f is an arbitrary mapping from $\mathcal{X}^n$ to $\mathcal{Y}^n$?

- Compound capacity is zero
  Separation **cannot** achieve distortion $\leq D_{P_S}(0)$ even if f happens to be good

Motivation - JSCC over Deterministic Channel

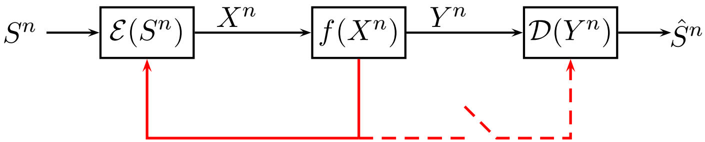


CSI@Tx Only

What if f is an arbitrary mapping from $\mathcal{X}^n$ to $\mathcal{Y}^n$?

- Compound capacity is zero
  Separation **cannot** achieve distortion $\leq D_{P_S}(0)$ even if f happens to be good
- **Is $D_{P_S}(0)$ the best we can do?**

Motivation - JSCC over Deterministic Channel

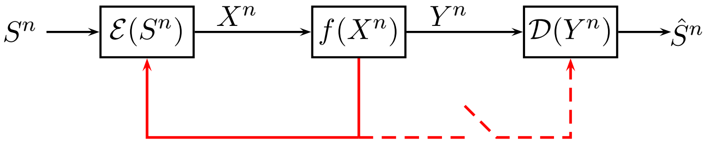


CSI@Tx Only

What if f is an arbitrary mapping from $\mathcal{X}^n$ to $\mathcal{Y}^n$?

- Compound capacity is zero
  Separation **cannot** achieve distortion $\leq D_{P_S}(0)$ even if f happens to be good
- **Is $D_{P_S}(0)$ the best we can do?**
- No! Joint Source-Channel Coding can do better

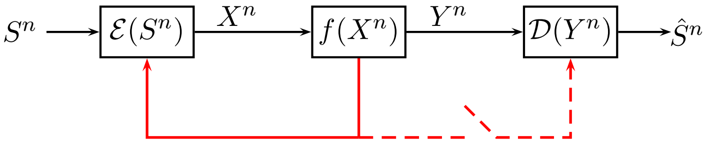
Motivation - JSCC over Deterministic Channel



CSI@Tx Only - Joint Source-Channel Coding

- $\mathcal{D} : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^n$ maps each possible output to a reconstruction sequence
- let $\mathcal{C} = \left\{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_{|\mathcal{Y}|^n}\right\} \subseteq \hat{\mathcal{S}}^n$ be the set of all possible reconstructions. $\mathcal{C}$ is a source code for P_S, where $R = \frac{1}{n} \log |\mathcal{Y}|$

Motivation - JSCC Over Deterministic Channel

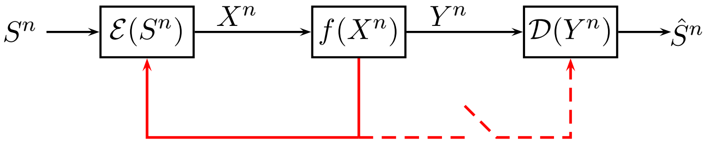


CSI@Tx Only - Joint Source-Channel Coding

- $\mathcal{D} : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^n$ maps each possible output to a reconstruction sequence
- let $\mathcal{C} = \left\{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_{|\mathcal{Y}|^n}\right\} \subseteq \hat{\mathcal{S}}^n$ be the set of all possible reconstructions. $\mathcal{C}$ is a source code for P_S, where $R = \frac{1}{n} \log |\mathcal{Y}|$
- The effect of the channel is diluting $\mathcal{C}$ to the source code

$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f(\mathcal{X}^n)\right) \subseteq \mathcal{C}$$

Motivation - JSCC Over Deterministic Channel



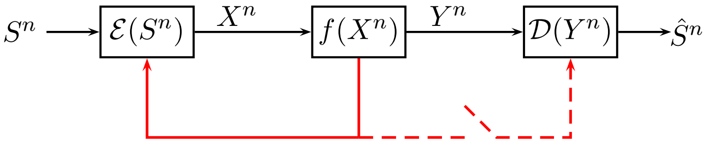
CSI@Tx Only - Joint Source-Channel Coding

- $\mathcal{D} : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^n$ maps each possible output to a reconstruction sequence
- let $\mathcal{C} = \left\{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_{|\mathcal{Y}|^n}\right\} \subseteq \hat{\mathcal{S}}^n$ be the set of all possible reconstructions. $\mathcal{C}$ is a source code for P_S, where $R = \frac{1}{n} \log |\mathcal{Y}|$
- The effect of the channel is diluting $\mathcal{C}$ to the source code

$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f\left(\mathcal{X}^n\right)\right) \subseteq \mathcal{C}$$

The channel chooses a subset of codewords from the source code

Motivation - JSCC over Deterministic Channel

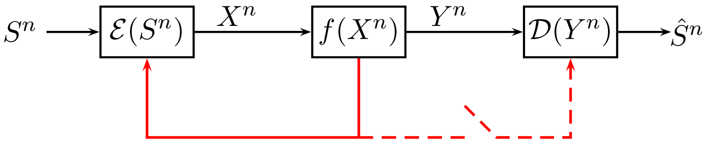


CSI@Tx Only - Joint Source-Channel Coding

$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f(\mathcal{X}^n)\right) \subseteq \mathcal{C} \quad ; \quad R^f_{\text{diluted}} = \frac{1}{n} \log |f(\mathcal{X}^n)|$$

- The encoder knows f and can induce any codeword in $\mathcal{C}^f_{\text{diluted}}$

Motivation - JSCC over Deterministic Channel



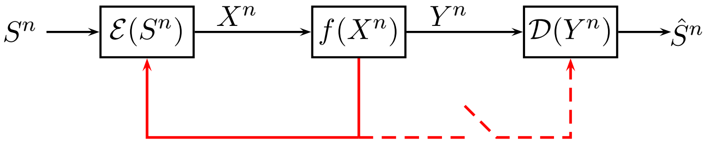
CSI@Tx Only - Joint Source-Channel Coding

$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f(\mathcal{X}^n)\right) \subseteq \mathcal{C} \quad ; \quad R^f_{\text{diluted}} = \frac{1}{n} \log |f(\mathcal{X}^n)|$$

- The encoder knows f and can induce any codeword in $\mathcal{C}^f_{\text{diluted}}$
- The obtained distortion is therefore

$$D\left(\mathcal{C}^f_{\text{diluted}}\right) = \mathbb{E}_{S^n} \min_{\mathbf{c} \in \mathcal{C}^f_{\text{diluted}}} d\left(S^n, \mathbf{c}\right)$$

Motivation - JSCC over Deterministic Channel



CSI@Tx Only - Joint Source-Channel Coding

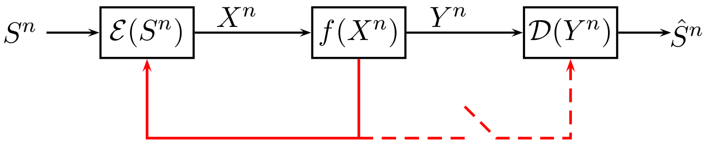
$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f(\mathcal{X}^n)\right) \subseteq \mathcal{C} \quad ; \quad R^f_{\text{diluted}} = \frac{1}{n} \log |f(\mathcal{X}^n)|$$

- The encoder knows f and can induce any codeword in $\mathcal{C}^f_{\text{diluted}}$
- The obtained distortion is therefore

$$D\left(\mathcal{C}^f_{\text{diluted}}\right) = \mathbb{E}_{S^n} \min_{\mathbf{c} \in \mathcal{C}^f_{\text{diluted}}} d\left(S^n, \mathbf{c}\right)$$

- Clearly $D\left(\mathcal{C}^f_{\text{diluted}}\right) \geq D_{P_S}\left(R^f_{\text{diluted}}\right)$

Motivation - JSCC over Deterministic Channel



CSI@Tx Only - Joint Source-Channel Coding

$$\mathcal{C}^f_{\text{diluted}} \triangleq \mathcal{D}\left(f(\mathcal{X}^n)\right) \subseteq \mathcal{C} \quad ; \quad R^f_{\text{diluted}} = \frac{1}{n}\log\left|f(\mathcal{Y}^n)\right|$$

Can we find $\mathcal{C}$ such that for almost every f

$$D\left(\mathcal{C}^f_{\text{diluted}}\right) \approx D_{P_S}\left(R^f_{\text{diluted}}\right)?$$

Subset–Universal Source Codes

Definition: Subset–Universal Source Code

A source code $\mathcal{C}$ with rate R is called *subset–universal* w.r.t. P_S and distortion measure d if for every $0 < R' < R$ almost every subset* of $2^{nR'}$ of its codewords achieve average distortion close to $D_{P_S}(R')$

$*$ The fraction of subsets for which this does not hold vanishes with n

SUBSET–UNIVERSAL SOURCE CODES

Definition: Subset–Universal Source Code

A source code $\mathcal{C}$ with rate R is called *subset–universal* w.r.t. P_S and distortion measure d if for every $0 < R' < R$ almost every subset* of $2^{nR'}$ of its codewords achieve average distortion close to $D_{P_S}(R')$

$*$ The fraction of subsets for which this does not hold vanishes with n

Main Result

For every DMS P_S, bounded distortion measure $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_+$ and rate $R > 0$, there exist a subset–universal source code

Subset–Universal Source Codes

Definition: Subset–Universal Source Code

A source code $\mathcal{C}$ with rate R is called *subset–universal* w.r.t. P_S and distortion measure d if for every $0 < R' < R$ almost every subset* of $2^{nR'}$ of its codewords achieve average distortion close to $D_{P_S}(R')$

$*$ The fraction of subsets for which this does not hold vanishes with n

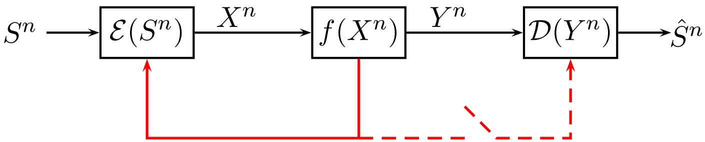
Main Result

For every DMS P_S, bounded distortion measure $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_+$ and rate $R > 0$, there exist a subset–universal source code

Extension

For every bounded distortion measure $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_+$ and rate $R > 0$, there exist a code $\mathcal{C}$ that is subset–universal w.r.t. all PMFs on $\mathcal{S}$

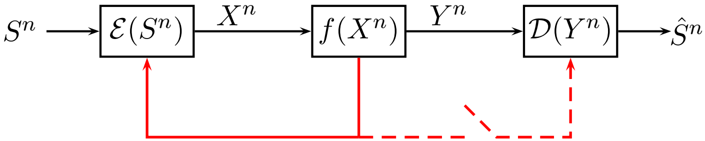
Back to the Motivating Example



Corollary

There exists a JSCC scheme that achieves average distortion $D_{P_S}\left(\frac{1}{n}\log|f(\mathcal{X}^n)|\right)$ for almost every deterministic channel f

Back to the Motivating Example



Corollary

There exists a JSCC scheme that achieves average distortion $D_{P_S}\left(\frac{1}{n}\log|f(\mathcal{X}^n)|\right)$ for almost every deterministic channel f

Remarks:

- There is no loss due to the receiver's ignorance
- The scheme does not need to depend on P_S
- The result holds for any deterministic channel if common randomness is allowed

Related Work

Ziv 72

There exists a codebook with rate R that universally achieves the distortion-rate function $D(R)$ for any stationary source, and even for a certain class of nonstationary sources.

Related Work

Ziv 72

There exists a codebook with rate R that universally achieves the distortion-rate function $D(R)$ for any stationary source, and even for a certain class of nonstationary sources.

"Proof":

- Split a source sequence of length $n = k\ell$ to ℓ consecutive subsequences of length k, where $\ell \gg k$
- Find the source code with 2^{kR} codewords of length k that achieves the smallest empirical distortion
- Send the codebook first, and then compress each of the ℓ blocks using it
- For $\ell \gg k$ the overhead becomes negligible

Related Work

Ziv 72

There exists a codebook with rate R that universally achieves the distortion-rate function $D(R)$ for any stationary source, and even for a certain class of nonstationary sources.

"Proof":

- Split a source sequence of length $n = k\ell$ to ℓ consecutive subsequences of length k, where $\ell \gg k$
- Find the source code with 2^{kR} codewords of length k that achieves the smallest empirical distortion
- Send the codebook first, and then compress each of the ℓ blocks using it
- For $\ell \gg k$ the overhead becomes negligible

Although more general than our result w.r.t. source statistics, this construction is not subset–universal

Proof Sketch - Mixture Distributions

- Let $\mathcal{P}^{|\hat{\mathcal{S}}|}$ denote the simplex containing all PMFs on $\hat{\mathcal{S}}$
- For $\theta \in \mathcal{P}^{|\hat{\mathcal{S}}|}$, let $P_\theta(\hat{s})$ be the corresponding pmf evaluated at $\hat{s}$
- Let $w(\theta)$ be the uniform probability density function on $\mathcal{P}^{|\hat{\mathcal{S}}|}$
- Define the mixture distribution

$$Q(\hat{\mathbf{s}}^n) = \int_{\theta \in \mathcal{P}^{|\hat{\mathcal{S}}|}} w(\theta) \prod_{i=1}^{n} P_\theta(\hat{s}_i) d\theta$$

Proof Sketch - Mixture Distributions

- Let $\mathcal{P}^{|\hat{\mathcal{S}}|}$ denote the simplex containing all PMFs on $\hat{\mathcal{S}}$
- For $\theta \in \mathcal{P}^{|\hat{\mathcal{S}}|}$, let $P_\theta(\hat{s})$ be the corresponding pmf evaluated at $\hat{s}$
- Let $w(\theta)$ be the uniform probability density function on $\mathcal{P}^{|\hat{\mathcal{S}}|}$
- Define the mixture distribution

$$Q(\hat{\mathbf{s}}^n) = \int_{\theta \in \mathcal{P}^{|\hat{\mathcal{S}}|}} w(\theta) \prod_{i=1}^{n} P_\theta(\hat{s}_i) d\theta$$

Joint Typicality Lemma for Mixture Distribution

Let $\hat{\mathbf{S}}^n \sim Q(\hat{\mathbf{s}}^n)$. Let $P_{S\hat{S}}$ be some pmf on $\mathcal{S} \times \hat{\mathcal{S}}$, and let $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$, for some $\varepsilon' < \varepsilon$. For n large enough

$$\Pr\left(\hat{\mathbf{S}}^n \in \mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}|\mathbf{s}^n)\right) \geq 2^{-n\left(I(S;\hat{S})+\delta(\varepsilon)\right)},$$

where $\delta(\epsilon) \to 0$ for $\epsilon \to 0$.

Proof Sketch

- **Codebook Generation:** Draw 2^{nR} codewords $\mathcal{C} = \{\hat{\mathbf{s}}_1^n, \ldots, \hat{\mathbf{s}}_{2^{nR}}^n\}$ independently from $Q(\hat{\mathbf{s}}^n)$

Proof Sketch

- **Codebook Generation:** Draw 2^{nR} codewords $\mathcal{C} = \{\hat{\mathbf{s}}_1^n, \ldots, \hat{\mathbf{s}}_{2^{nR}}^n\}$ independently from $Q(\hat{\mathbf{s}}^n)$
- **Encoding:** Given $\hat{\mathbf{s}}^n$, $\mathcal{C}$, and an index set $\mathcal{I} \subseteq [2^{nR}]$ with $|\mathcal{I}| = 2^{nR'}$, send
$$m = \operatorname*{argmin}_{m \in \mathcal{I}} d\left(\mathbf{s}, \hat{\mathbf{s}}^n(m)\right)$$

Proof Sketch

- **Codebook Generation:** Draw 2^{nR} codewords $\mathcal{C} = \{\hat{\mathbf{s}}_1^n, \ldots, \hat{\mathbf{s}}_{2^{nR}}^n\}$ independently from $Q(\hat{\mathbf{s}}^n)$
- **Encoding:** Given $\hat{\mathbf{s}}^n$, $\mathcal{C}$, and an index set $\mathcal{I} \subseteq [2^{nR}]$ with $|\mathcal{I}| = 2^{nR'}$, send

$$m = \underset{m \in \mathcal{I}}{\operatorname{argmin}}\, d\left(\mathbf{s}, \hat{\mathbf{s}}^n(m)\right)$$

Analyze a suboptimal encoder:

For some small $\delta > 0$, find

$$P_{\hat{S}|S}^{R'} = \underset{P_{\hat{S}|S}: I(S;\hat{S}) \leq R' - \delta}{\operatorname{argmin}} \sum_{s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}}} P_S(s) P_{\hat{S}|S}(\hat{s}|s) d(s, \hat{s})$$

and set $P_{S\hat{S}}^{R'} = P_S P_{\hat{S}|S}^{R'}$. Send the smallest index $m \in \mathcal{I}$ such that

$$(\mathbf{s}^n, \hat{\mathbf{s}}^n(m)) \in \mathcal{T}_\varepsilon^{(n)}(P_{S\hat{S}}^{R'}).$$

Note: If such an index is found the distortion is $\approx D_{P_S}(R')$

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

- With high probability $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

- With high probability $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{\mathbf{s}}^n)$

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

- With high probability $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{\mathbf{s}}^n)$
- Assuming $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$, for each one of them

$$\Pr\left(\hat{\mathbf{S}}^n \in \mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}^{R'}|\mathbf{s}^n)\right) \geq 2^{-n\left(I(S;\hat{S}^{R'})+\delta(\varepsilon)\right)},$$

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

- With high probability $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{\mathbf{s}}^n)$
- Assuming $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$, for each one of them

$$\Pr\left(\hat{\mathbf{S}}^n \in \mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}^{R'}|\mathbf{s}^n)\right) \geq 2^{-n\left(I(S;\hat{S}^{R'})+\delta(\varepsilon)\right)},$$

- The probability that none of them is in $\mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}|\mathbf{s}^n)$ is upper bounded by

$$\exp\left\{-2^{n(R'-I(S;\hat{S}^{R'})-\delta(\epsilon))}\right\} = \exp\left\{-2^{n(\delta-\delta(\epsilon))}\right\}$$

Proof Sketch

Fix $\mathcal{I}$ and R'. We will show that the average probability that no index is found is small

- With high probability $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{\mathbf{s}}^n)$
- Assuming $\mathbf{s}^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$, for each one of them

$$\Pr\left(\hat{\mathbf{S}}^n \in \mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}^{R'}|\mathbf{s}^n)\right) \geq 2^{-n\left(I(S;\hat{S}^{R'})+\delta(\varepsilon)\right)},$$

- The probability that none of them is in $\mathcal{T}_{\varepsilon}^{(n)}(P_{S\hat{S}}|\mathbf{s}^n)$ is upper bounded by

$$\exp\left\{-2^{n(R'-I(S;\hat{S}^{R'})-\delta(\epsilon))}\right\} = \exp\left\{-2^{n(\delta-\delta(\epsilon))}\right\}$$

This is true for any $\mathcal{I}$ and R'. By Markov's inequality and continuity of $D_{P_S}(R)$ this is true for any $R' < R$ and almost any $\mathcal{I}$ with cardinality $2^{nR'}$

Summary

- We defined the notion of subset–universal lossy source codes
- We proved that for any PMF and d such codes exist
- We further showed that there exist a code that is simultaneously subset–universal for all PMFs on the same alphabet
- Our motivation was JSCC for an unknown deterministic channels
- There should be more applications...