SUBSET–UNIVERSAL LOSSY COMPRESSION

Or Ordentlich
Joint work with Ofer Shayevitz

Information Theory Workshop

Jerusalem, Israel
April 28, 2015
Motivation - JSCC Over Deterministic Channel

\[ S^n \xrightarrow{\mathcal{E}(S^n)} X^n \xrightarrow{f(X^n)} Y^n \xrightarrow{\mathcal{D}(Y^n)} \hat{S}^n \]
**Motivation - JSCC Over Deterministic Channel**

- $S^n$ is a discrete memoryless source with PMF $P_S$
- $d : S \times \hat{S} \mapsto \mathbb{R}_+$ a bounded distortion measure
- $f : X^n \mapsto Y^n$ 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
- $f(X^n) \subseteq Y^n$ is the image of $f$
- $E$ and $D$ are the encoder and decoder
Motivation - JSCC Over Deterministic Channel

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

- The channel becomes a bit pipe of rate \( nR = \log |f(X^n)| \)
- Separation is optimal
Motivation - JSCC Over Deterministic Channel

CSI@Both

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

$$D_{PS}(R) \triangleq \min_{P_{\hat{S}|S}: I(S;\hat{S}) \leq R} \sum_{s \in S, \hat{s} \in \hat{S}} P_S(s) P_{\hat{S}|S}(\hat{s}|s) d(s, \hat{s}).$$
**Motivation - JSCC Over Deterministic Channel**

\[ S^n \xrightarrow{\mathcal{E}(S^n)} X^n \xrightarrow{f(X^n)} Y^n \xrightarrow{\mathcal{D}(Y^n)} \hat{S}^n \]

CSI@Tx Only
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
  \[ D_{PS} \left( \frac{1}{n} \log |f(X^n)| \right) \]
Motivation - JSCC over Deterministic Channel

\[ S^n \xrightarrow{} E(S^n) \xrightarrow{} X^n \xrightarrow{} f(X^n) \xrightarrow{} Y^n \xrightarrow{} D(Y^n) \xrightarrow{} \hat{S}^n \]

CSI@Tx Only
Gelfand-Pinsker
CSI@Tx Only
Gelfand-Pinsker

- If the channel is memoryless with state

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

where \( \{T_i\} \) is an i.i.d. state process, separation (Gelfand-Pinsker + source coding) is optimal (Merhav-Shamai 03) and achieves

\[ D_{PS} \left( \frac{1}{n} \log |f(\mathcal{X}^n)| \right) \]
What if \( f \) is an arbitrary mapping from \( \mathcal{X}^n \) to \( \mathcal{Y}^n \)?
Motivation - JSCC Over Deterministic Channel

What if $f$ is an arbitrary mapping from $X^n$ to $Y^n$?

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

- Is $D_{PS}(0)$ the best we can do?

- No! Joint Source-Channel Coding can do better
CSI@Tx Only - Joint Source-Channel Coding

- \( D : Y^n \mapsto \hat{S}^n \) maps each possible output to a reconstruction sequence.
- Let \( C = \{\hat{s}_1, \ldots, \hat{s}_{|Y|^n}\} \subseteq \hat{S}^n \) be the set of all possible reconstructions. \( C \) is a source code for \( P_S \), where \( R = \frac{1}{n} \log |Y| \)
CSI@Tx Only - Joint Source-Channel Coding

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

\[
C_{\text{diluted}}^f \triangleq D(f(\mathcal{X}^n)) \subseteq C
\]
CSI@Tx Only - Joint Source-Channel Coding

- \( D : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}^n \) maps each possible output to a reconstruction sequence
- let \( \mathcal{C} = \{\hat{s}_1, \ldots, \hat{s}_{|\mathcal{Y}|^n}\} \subseteq \hat{\mathcal{S}}^n \) 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

\[
\mathcal{C}_{\text{diluted}}^f \triangleq D (f(\mathcal{X}^n)) \subseteq \mathcal{C}
\]

The channel chooses a subset of codewords from the source code
CSI@Tx Only - Joint Source-Channel Coding

\[ \mathcal{C}_{\text{diluted}}^f \triangleq \mathcal{D}(f(X^n)) \subseteq \mathcal{C} ; \quad R_{\text{diluted}}^f = \frac{1}{n} \log |f(X^n)| \]

- The encoder knows \( f \) and can induce any codeword in \( \mathcal{C}_{\text{diluted}}^f \)
CSI@Tx Only - Joint Source-Channel Coding

\[ C_{\text{diluted}}^f \triangleq \mathcal{D}(f(X^n)) \subseteq \mathcal{C} \quad ; \quad R_{\text{diluted}}^f = \frac{1}{n} \log |f(X^n)| \]

- The encoder knows \( f \) and can induce any codeword in \( C_{\text{diluted}}^f \).
- The obtained distortion is therefore

\[
\mathcal{D} \left( C_{\text{diluted}}^f \right) = \mathbb{E}_{S^n} \min_{c \in C_{\text{diluted}}^f} d(S^n, c)
\]
**Motivation - JSCC Over Deterministic Channel**

**CSI@Tx Only - Joint Source-Channel Coding**

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

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

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

- Clearly \( D\left(C^f_{\text{diluted}}\right) \geq D_{PS}\left(R^f_{\text{diluted}}\right) \)
CSI@Tx Only - Joint Source-Channel Coding

\[ \mathcal{C}_{\text{diluted}}^f \triangleq \mathcal{D} (f(X^n)) \subseteq \mathcal{C} ; \quad R_{\text{diluted}}^f = \frac{1}{n} \log |f(Y^n)| \]

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

\[ D \left( \mathcal{C}_{\text{diluted}}^f \right) \approx D_{\text{PS}} \left( R_{\text{diluted}}^f \right) ? \]
Definition: Subset–Universal Source Code

A source code $C$ 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 $n$
**Definition: Subset–Universal Source Code**

A source code $C$ 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 $d : S \times \hat{S} \mapsto \mathbb{R}_+$ and rate $R > 0$, there exist a subset–universal source code
<table>
<thead>
<tr>
<th>Definition: Subset–Universal Source Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A source code $C$ with rate $R$ is called <em>subset–universal</em> w.r.t. $P_S$ and distortion measure $d$ if for every $0 &lt; R' &lt; R$ almost every subset* of $2^{nR'}$ of its codewords achieve average distortion close to $D_{P_S}(R')$</td>
</tr>
<tr>
<td>* The fraction of subsets for which this does not hold vanishes with $n$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Main Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>For every DMS $P_S$, bounded distortion measure $d : S \times \hat{S} \mapsto \mathbb{R}_+$ and rate $R &gt; 0$, there exist a subset–universal source code</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>For every bounded distortion measure $d : S \times \hat{S} \mapsto \mathbb{R}_+$ and rate $R &gt; 0$, there exist a code $C$ that is subset–universal w.r.t. all PMFs on $S$</td>
</tr>
</tbody>
</table>
Corollary

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

There exists a JSCC scheme that achieves average distortion $D_{PS} \left( \frac{1}{n} \log |f(X^n)| \right)$ 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.
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.
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 $\ell$ 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 $\ell$ blocks using it
- For $\ell \gg k$ the overhead becomes negligible
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 $\ell$ 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 $\ell$ 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 $\mathcal{P}^{\hat{S}}$ denote the simplex containing all PMFs on $\hat{S}$
- For $\theta \in \mathcal{P}^{\hat{S}}$, let $P_\theta(\hat{s})$ be the corresponding pmf evaluated at $\hat{s}$
- Let $w(\theta)$ be the uniform probability density function on $\mathcal{P}^{\hat{S}}$
- Define the mixture distribution

\[
Q(\hat{s}^n) = \int_{\theta \in \mathcal{P}^{\hat{S}}} w(\theta) \prod_{i=1}^{n} P_\theta(\hat{s}_i) d\theta
\]
Proof Sketch - Mixture Distributions

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

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

Joint Typicality Lemma for Mixture Distribution

Let $\hat{S}^n \sim Q(\hat{s}^n)$. Let $P_{S\hat{S}}$ be some pmf on $S \times \hat{S}$, and let $s^n \in T_{\epsilon'}(P_S)$, for some $\epsilon' < \epsilon$. For $n$ large enough

$$\Pr \left( \hat{S}^n \in T_{\epsilon}(P_{S\hat{S}}|s^n) \right) \geq 2^{-n\left(I(S;\hat{S})+\delta(\epsilon)\right)},$$

where $\delta(\epsilon) \to 0$ for $\epsilon \to 0$. 
**Proof Sketch**

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

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

- **Encoding:** Given $\hat{s}^n$, $C$, and an index set $\mathcal{I} \subseteq [2^{nR}]$ with $|\mathcal{I}| = 2^{nR'}$, send

$$m = \arg\min_{m \in \mathcal{I}} d(s, \hat{s}^n(m))$$
**Proof Sketch**

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

- **Encoding:** Given $\hat{s}^n$, $C$, and an index set $\mathcal{I} \subseteq [2^{nR}]$ with $|\mathcal{I}| = 2^{nR'}$, send

  $$m = \arg\min_{m \in \mathcal{I}} d(s, \hat{s}^n(m))$$

**Analyze a suboptimal encoder:**

For some small $\delta > 0$, find

$$P^R_{\hat{s}|S} = \arg\min_{P_{\hat{s}|S}:I(S;\hat{s}) \leq R' - \delta} \sum_{s \in S, \hat{s} \in \hat{S}} P_S(s)P_{\hat{s}|S}(\hat{s} | s)d(s, \hat{s})$$

and set $P^R_{\hat{s}S} = P_S P^R_{\hat{s}|S}$. Send the smallest index $m \in \mathcal{I}$ such that

$$(s^n, \hat{s}^n(m)) \in \mathcal{T}_\varepsilon(n)(P^R_{\hat{s}S}).$$

Note: If such an index is found the distortion is $\approx D_{PS}(R')$
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 $s^n \in T_{\varepsilon'}(P_S)$
**Proof Sketch**

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

- With high probability $s^n \in T_{\epsilon'}^n(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{s}^n)$
Proof Sketch

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

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

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

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

- With high probability $s^n \in T_{\epsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{s}^n)$
- Assuming $s^n \in T_{\epsilon'}^{(n)}(P_S)$, for each one of them
  \[
  \Pr \left( \hat{S}^n \in T_{\epsilon'}^{(n)}(P_{S\hat{S}}|s^n) \right) \geq 2^{-n(I(S;\hat{S}^{R'})+\delta(\epsilon))},
  \]

- The probability that none of them is in $T_{\epsilon'}^{(n)}(P_{S\hat{S}}|s^n)$ 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 $s^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$
- We have $2^{nR'}$ codewords drawn from $Q(\hat{s}^n)$
- Assuming $s^n \in \mathcal{T}_{\varepsilon'}^{(n)}(P_S)$, for each one of them

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

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

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

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