Source Coding with Composite Side Information at the Decoder

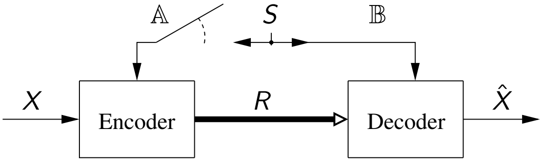
Anatoly Khina

Joint work with: Uri Erez

Tel Aviv University

November 15, 2012
IEEEI, Eilat

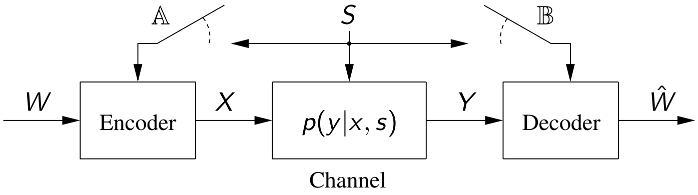
Source Coding with Side-information at the Decoder



- X – Source
- $\hat{X}$ – Reconstructed (distorted) source at the decoder
- S – Side-information
- $d(X, \hat{X})$ – Distortion measure
- **Memorylessness (i.i.d.):**

$$p(\mathbf{x}, \mathbf{s}) = \prod_{i=1}^{n} p(x_i, s_i)$$

Dual Problem: Memoryless State-dependent Channel



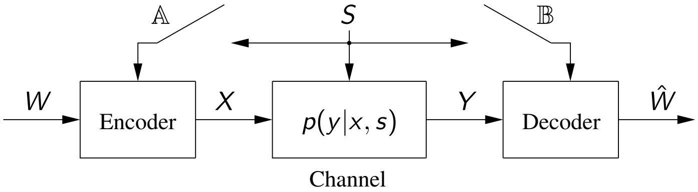
Memoryless channel:

$$p(\mathbf{y}|\mathbf{x}, \mathbf{s}) = \prod_{i=1}^{n} p(y_i|x_i, s_i) \,.$$

"Memoryless" (i.i.d.) state sequence:

$$p(\mathbf{s}) = \prod_{i=1}^{n} p(s_i) \,.$$

State not Known



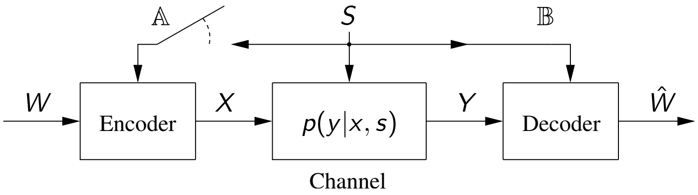
- Problem reduces to "regular" DMC:

$$p(y|x) = \sum_{s \in \mathcal{S}} p(s) p(y|x, s)$$

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i) \,.$$

- **Capacity** [Shannon '48]: $\max_{p_x} I(X;Y)$

State Known at the Decoder

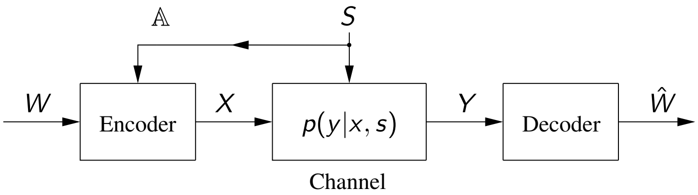


- State (S) available at Rx can be regarded as part of output
- Define $\tilde{Y} = (Y, S)$

  $\Downarrow$

  Channel from X to $\tilde{Y}$ with no state available at Rx
- No special treatment is required for state available at Rx

State Known at the Encoder



Causal state knowledge [Shannon '58]

- $x_i = \text{func}\left(w, s_1^i\right)$ – at time i, only states $s_1, \ldots, s_i$ are known
- **Capacity:** $\max_{p(t)} I(T;Y)$,

  where $t : \mathcal{S} \to \mathcal{X}$, i.e., mappings $x = t(s)$
- Equivalent representation: $\max_{p(u),x(u,s)} I(U;Y)$

  (U is independent of S)

State Known at the Encoder

Non-causal state knowledge [Gel'fand & Pinsker '80]

- $x_i = \text{func}\,(w, s_1^n)$ – at time i, all states $s_1, \ldots, s_n$ are known
- **Capacity:** $\max_{p(u|s),x(u,s)} I(U;Y) - I(U;S)$
- In terms of "Shannon strategies": $x = t_s(s)$

  (here random strategy t *depends on* s)
- Achievable using "random-binning"

NC knowledge with limited look-ahead [Weissman & El Gamal '06]

- Limited look-ahead k
- $x_i = \text{func}\left(w, s_1^{i+k}\right)$ – at time i, states $s_1, \ldots, s_{i+k}$ are known

Composite State Known at the Encoder

Observations

- Achievables of causal and non-causal cases can be derived in the same way
- Converses of causal and non-causal cases are similar and can be combined

State with causal and non-causal parts [Khina, Kesal, Erez '11]

- $S = ({}^{c}S, {}^{nc}S)$
- ${}^{c}S$ – known *causally* at Tx
- ${}^{nc}S$ – known *non-causally* at Tx
- **Capacity:** $C = \max_{p(u|{}^{nc}s),\, x(u,{}^{nc}s,{}^{c}s)} \left[I(U;Y) - I(U;{}^{nc}S) \right]$
- U is independent of ${}^{c}S$ *given* ${}^{nc}S$

Converse

$$n(R-\epsilon_n) \overset{\text{Fano}}{\leq} I(W;Y_1^n) = \sum_{i=1}^{n} I(W;Y_i|Y_1^{i-1}) \leq \sum_{i=1}^{n} I(W,Y_1^{i-1};Y_i)$$

Shannon (causal)

- $U_i \triangleq (W, Y_1^{i-1})$.
- Causality

  $\Downarrow$

  U_i is independent of S_i.

$$\sum I(W,Y_1^{i-1};Y_i) = \sum I(U_i;Y_i) \leq n \max_{p(u),p(x|u,s)} I(U;Y)$$

Gel'fand–Pinsker (non-causal)

- $U_i \triangleq (W, Y_1^{i-1}, S_{i+1}^n)$
- S_i independent of (W, S_{i+1}^n)
- Chain-rule for mutual informations.

$$\sum I(W,Y_1^{i-1};Y_i) = \cdots = \sum I(U_i;Y_i) - I(U_i;S_i) \leq \max_{p(u|s),p(x|u,s)} \{I(U;Y) - I(U;S)\}$$

Converse

In both converses...

- Achievable rate bounded by $I(U_i; Y_i) - I(U_i; S_i)$
- Auxiliary variable: $U_i = (W, Y_1^{i-1}, S_{i+1}^n)$
  - Causal case: reduces to $U_i = (W, Y_1^{i-1})$, since S_{i+1}^n is independent of Y_i
- **Can be used for the composite causal–non-causal case!**

Converse for the composite causal–non-causal case

- $S = ({}^{\mathrm{c}}S, {}^{\mathrm{nc}}S)$
- Achievable rate is bounded by $I(U_i; Y_i) - I(U_i; {}^{\mathrm{nc}}S_i)$
- Auxiliary variable: $U_i = (W, Y_1^{i-1}, {}^{\mathrm{nc}}S_{i+1}^n)$
- Maximization over $p(u|{}^{\mathrm{nc}}s)$ and $x = \mathrm{func}\,(u, {}^{\mathrm{c}}s, {}^{\mathrm{nc}}s)$
  (U is independent of ${}^{\mathrm{c}}S$ *given* ${}^{\mathrm{nc}}S$)

Achievable

Causal (Shannon) case

- $x = t(s)$
- Strategy t is generated from W (input to equivalent channel)
- Can be thought of as "degenerated random binning" (U independent of S)

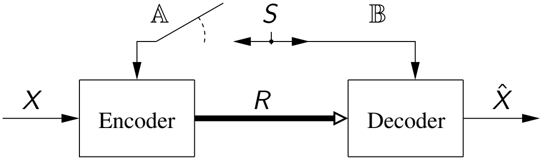
Non-causal (Gel'fand–Pinsker) case

- Use random binning w.r.t. $p(u|s)$ and $x(u,s)$

Composite causal–non-causal case

- Use random binning w.r.t $p(u, {}^{nc}s)$ and $x(u, {}^{c}s, {}^{nc}s)$
- Alternatively, combine random strategies w.r.t ${}^{c}S$ and random binning w.r.t. ${}^{nc}S$

Source Coding with Side-information at the Decoder



- X – Source
- $\hat{X}$ – Reconstructed (distorted) source at the decoder
- S – Side-information
- $d(X, \hat{X})$ – Distortion measure

Side-information at the encoder ($\mathbb{A}$ is closed)

- S can be regarded as part of the source: $\tilde{X} \triangleq (X, S)$
- Distortion measure is w.r.t. X (and not $\tilde{X}$)

Source Coding with Side-Information at the Decoder

Non-causal side-information [Wyner–Ziv '76]

$$R(D) = \min \left[I(U;X) - I(U;S) \right] ,$$

where minimum is over all $f : \mathcal{U} \times \mathcal{S} \to \hat{\mathcal{X}}$,
s.t. $E\left[d(X, f(U,S)) \right] \leq D$.

Causal side-information [Weissman–El Gamal '06]

$$R(D) = \min I(U;X)$$

minimum over the **same set as in the non-causal problem**.

Composite causal–non-causal side-information

$$R(D) = \min \left[I(U;X) - I(U;{}^{\text{nc}}S) \right]$$

minimum over the same set.

Separation Principle

Separation between channel and source with side-informations (SIs) [Merhav–Shamai '03]

Separation holds between channel and source coding with

- Non-causal channel SI (Ge'fand–Pinsker)
  $+$
  Non-causal source SI (Wyner–Ziv)
- Causal channel SI (Shannon)
  $+$
  Non-causal source SI (Wyner–Ziv)

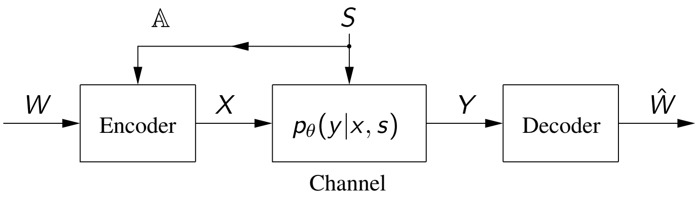
Separation between channel and source with **composite** SIs

Separation holds between channel and source coding with

Composite channel SI
$+$
Composite source SI

When do Both Converses Diverge?

Compound state-dependent channel:



- $\theta \in \Theta$ – "Compound parameter": constant, unknown to Tx

Compound channel with no state knowledge ($\mathbb{A}$ is open)

- Worst-case capacity – Maximal rate for all θ **simultaneously**
- **Capacity:** [Blackwell et al. '59; Dobrushin '59; Wolfowitz '60]

$$C^{\text{wc}} = \max_{p(x)} \min_{\theta \in \Theta} I(X;Y)$$

When do Both Converses Diverge?

Compound Channel with state S known *causally* at Tx

Trying to generalize Shannon's converse to compound case:

- $U_i \triangleq (W, Y_1^{i-1})$
- U_i depends on the statistics of Y which is unknown! (since θ not known)

$$n(R - \epsilon_n) \leq \sum_{i=1}^{n} I_\theta(W, Y_1^{i-1}; Y_i) \leq \sum_{i=1}^{n} I_\theta(W, \cancel{Y_1^{i-1}}, S_1^{i-1}; Y_i)$$

causality

- $U_i \triangleq (W, S_1^{i-1})$ – No knowledge of θ is assumed! (Original U_i used by Shannon)
- **(Worst-case) Capacity:** [Khina, Erez '10]

$$C^{\text{wc}} = \max_{p(u), x(u,s)} \min_{\theta \in \Theta} I_\theta(U; Y)$$

When do Both Converses Diverge?

Compound Channel with state S known *non-causally* at Tx

- Similarly one would expect:

$$C^{\text{wc}} = \max_{p(u|s),x(u,s)} \min_{\theta \in \Theta} I_\theta(U;Y)$$

- This is **achievable** but **not optimal**!
  [Piantanida, Shamai '10; Nair, El Gamal, Chia '10]
  - Marton's broadcast technique exceeds this rate

Summary

- Different side-information scenarios can be treated similarly
- Recognizing these similarities allows to:
  - Solve several different scenarios at once
  - Combine results for composite/mixed scenarios
- However, in certain more complex scenarios similar/combined treatments might diverge