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.):**

$$p(x, s) = \prod_{i=1}^{n} p(x_i, s_i)$$
Dual Problem: Memoryless State-dependent Channel

Memoryless channel:

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

“Memoryless” (i.i.d.) state sequence:

\[ p(s) = \prod_{i=1}^{n} p(s_i). \]
State not Known

Problem reduces to “regular” DMC:

\[ p(y|x) = \sum_{s \in S} p(s) p(y|x, s) \]

\[ p(y|x) = \prod_{i=1}^{n} p(y_i|x_i) . \]

- **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]

- \( x_i = \text{func}(w, s_1^i) \) – at time \( i \), only states \( s_1, \ldots, s_i \) are known

- **Capacity:** \( \max_{p(t)} I(T; Y) \),

  where \( t: S \to X \), i.e., mappings \( x = t(s) \)

- Equivalent representation: \( \max_{p(u), x(u,s)} I(U; Y) \)

  \( (U \text{ is independent of } S) \)
### State Known at the Encoder

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

- $x_i = \text{func}(w, s^n_1)$ – 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 $k$

- $x_i = \text{func}(w, s^{i+k}_1)$ – 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]
- $S = (cS, ncS)$
- $cS$ – known causally at Tx
- $ncS$ – known non-causally at Tx
- **Capacity:** $C = \max_{p(u|ncS), x(u, ncS, cS)} [I(U; Y) - I(U; ncS)]$
- $U$ is independent of $cS$ given $ncS$
Converse

\[ n(R - \epsilon_n) \leq Fano I(W; Y^n_1) = \sum_{i=1}^{n} I(W; Y_i | Y_{1}^{i-1}) \leq \sum_{i=1}^{n} I(W, Y_{1}^{i-1}; Y_i) \]

**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.

\[
\sum I(W, Y_{1}^{i-1}; Y_i) = \cdots = \sum I(U_i; Y_i) - I(U_i; S_i) \leq \max_{p(u|s), p(x|u,s)} \{ I(U; Y) - I(U; S) \}
\]
In both converses...

- Achievable rate bounded by $I(U_i; Y_i) - I(U_i; S_i)$
- Auxiliary variable: $U_i = (W, Y_{i-1}^i, S_{i+1}^n)$
  - Causal case: reduces to $U_i = (W, Y_{i-1}^i)$, 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

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

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

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

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

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

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:**

\[
\begin{align*}
&W \\
&\xrightarrow{\text{Encoder}} X \\
&\xrightarrow{\text{Channel}} Y \\
&\xrightarrow{\text{Decoder}} \hat{W}
\end{align*}
\]

- \( \theta \in \Theta \) – “Compound parameter”: constant, unknown to Tx

**Compound channel with no state knowledge (A is open)**

- Worst-case capacity – Maximal rate for all \( \theta \) **simultaneously**

**Capacity:** [Blackwell et al. ’59; Dobrushin ’59; Wolfowitz ’60]

\[
C^{wc} = \max_{p(x)} \min_{\theta \in \Theta} I(X; Y)
\]
When do Both Converses Diverge?

### Compound Channel with state $S$ known causally at Tx

Trying to generalize Shannon’s converse to compound case:

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

\[
n(R - \epsilon_n) \leq \sum_{i=1}^{n} I_{\theta}(W, Y_i^{i-1}; Y_i) \leq \sum_{i=1}^{n} I_{\theta}(W, Y_i^{i-1}, S_i^{i-1}; Y_i)^{\text{causality}}
\]

- $U_i \triangleq (W, S_i^{i-1})$ – No knowledge of $\theta$ is assumed! (Original $U_i$ used by Shannon)

- **(Worst-case) Capacity:** [Khina, Erez ’10]

\[
C^{wc} = \max_{p(u),x(u,s)} \min_{\theta \in \Theta} I_{\theta}(U; Y)
\]
When do Both Converses Diverge?

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

- Similarly one would expect:

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

- 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.