Thesis Outline

1. Source Coding with Feed-forward
   - Point to Point Source Coding (Chap. 2)
     - Problem Statement
     - Rate-Distortion Function
     - Error Exponents
   - Computation (Chap. 4)
   - Multiple Descriptions (Chap. 5)

2. Channel Coding with Feedback (Chap. 3, 4)
Thesis Outline

1 Source Coding with Feed-forward
   ▶ Point to Point Source Coding (Chap. 2)
      - Problem Statement
      - Rate-Distortion Function
      - Error Exponents
   ▶ Computation (Chap. 4)
   ▶ Multiple Descriptions (Chap. 5)

2 Channel Coding with Feedback (Chap. 3,4)
Lossy Data Compression

- Source $X$, reconstruction $\hat{X}$
- Distortion measure $d(X, \hat{X})$

**Rate-distortion function**

Minimum rate $R$ for distortion level $D$ (Shannon ’59)
Lossy Data Compression

Source $X$, reconstruction $\hat{X}$

Distortion measure $d(X, \hat{X})$

Rate-distortion function

Minimum rate $R$ for distortion level $D$ (Shannon '59)
Source Coding with Side-Information

- $X, Y$ correlated random variables
- *Example*: Temperature at nearby cities
- Presence of $Y \Rightarrow$ lower rate for given distortion $D$

Rate-distortion function $R_{WZ}(D)$ [Wyner, Ziv ’76]
• $X$, $Y$ correlated random variables

• *Example*: Temperature at nearby cities

• Presence of $Y \Rightarrow$ lower rate for given distortion $D$

Rate-distortion function $R_{WZ}(D)$ [Wyner, Ziv ’76]
Source Coding with Side-Information

Example: Block length $N = 5$

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_7$</td>
<td>$X_8$</td>
<td>$X_9$</td>
<td>$X_{10}$</td>
</tr>
<tr>
<td>Encoder</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>W</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Side Info</td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$Y_3$</td>
<td>$Y_4$</td>
<td>$Y_5$</td>
<td>$Y_6$</td>
<td>$Y_7$</td>
<td>$Y_8$</td>
<td>$Y_9$</td>
<td>$Y_{10}$</td>
</tr>
<tr>
<td>Decoder</td>
<td>$\hat{X}_1$</td>
<td>$\hat{X}_2$</td>
<td>$\hat{X}_3$</td>
<td>$\hat{X}_4$</td>
<td>$\hat{X}_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Source Coding with Side-Information

Example: Block length $N = 5$

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_7$</td>
<td>$X_8$</td>
<td>$X_9$</td>
<td>$X_{10}$</td>
</tr>
<tr>
<td>Encoder</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>W</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Side Info</td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$Y_3$</td>
<td>$Y_4$</td>
<td>$Y_5$</td>
<td>$Y_6$</td>
<td>$Y_7$</td>
<td>$Y_8$</td>
<td>$Y_9$</td>
<td>$Y_{10}$</td>
</tr>
<tr>
<td>Decoder</td>
<td>$\hat{X}_1$</td>
<td>$\hat{X}_2$</td>
<td>$\hat{X}_3$</td>
<td>$\hat{X}_4$</td>
<td>$\hat{X}_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Source Coding with Side-Information

Example: Block length $N = 5$

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_7$</td>
<td>$X_8$</td>
<td>$X_9$</td>
<td>$X_{10}$</td>
</tr>
<tr>
<td>Encoder</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>W</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Side Info</td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$Y_3$</td>
<td>$Y_4$</td>
<td>$Y_5$</td>
<td>$Y_6$</td>
<td>$Y_7$</td>
<td>$Y_8$</td>
<td>$Y_9$</td>
<td>$Y_{10}$</td>
</tr>
<tr>
<td>Decoder</td>
<td>$\hat{X}_1$</td>
<td>$\hat{X}_2$</td>
<td>$\hat{X}_3$</td>
<td>$\hat{X}_4$</td>
<td>$\hat{X}_5$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Suppose there is a delay in the side info available at the decoder.
Suppose there is a delay in the side info available at the decoder.

\[
\begin{array}{cccccccccc}
\text{Time} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\text{Source} & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 & X_{10} \\
\text{Side Info} & - & - & - & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 \\
\text{Decoder} & \hat{X}_1 & \hat{X}_2 & \hat{X}_3 & \hat{X}_4 & \hat{X}_5
\end{array}
\]
Suppose there is a delay in the side info available at the decoder.

There is a slow noisy path from the encoder to the decoder. The encoder receives data at times 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. The side information is received at times 1, 2, 3, 4, 5, and 6. The decoder estimates the source at times 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10.
Side-Information with Delay

Suppose there is a delay in the side info available at the decoder.

![Diagram](image)

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_7$</td>
<td>$X_8$</td>
<td>$X_9$</td>
<td>$X_{10}$</td>
</tr>
<tr>
<td>Encoder</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$W$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Side Info</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$Y_3$</td>
<td>$Y_4$</td>
<td>$Y_5$</td>
<td>$Y_6$</td>
<td>$Y_7$</td>
</tr>
<tr>
<td>Decoder</td>
<td>$\hat{X}_1$</td>
<td>$\hat{X}_2$</td>
<td>$\hat{X}_3$</td>
<td>$\hat{X}_4$</td>
<td>$\hat{X}_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is feed-forward?

The source field itself available with delay at decoder.

Time  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---
Source | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$
Encoder | - | - | - | - | - | W | - | - | - | - | W
Extra info | - | - | - | - | - | - | - | X_1 | X_2 | X_3 | X_4
Decoder | $\hat{X}_1$ | $\hat{X}_2$ | $\hat{X}_3$ | $\hat{X}_4$ | $\hat{X}_5$

Here, block length = 5, delay = 6 time units.
What is feed-forward?

The source field itself available with delay at decoder.

Time

1  2  3  4  5  6  7  8  9  10

Source

\( X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9 \quad X_{10} \)

Encoder

- - - - - W - - - - - W

Extra info

- - - - - - - - X_1 X_2 X_3 X_4

Decoder

\( \hat{X}_1 \quad \hat{X}_2 \quad \hat{X}_3 \quad \hat{X}_4 \quad \hat{X}_5 \)

Here, block length = 5, delay = 6 time units.
What is feed-forward?

The source field itself available with delay at decoder.

Here, block length= 5, delay= 6 time units.
What is feed-forward?

The source field itself available with delay at decoder.

Encoder -> Rate R -> Decoder

slow noiseless path

Rate R

Time 1 2 3 4 5 6 7 8 9 10
Source \( X_1 \) \( X_2 \) \( X_3 \) \( X_4 \) \( X_5 \) \( X_6 \) \( X_7 \) \( X_8 \) \( X_9 \) \( X_{10} \)
Encoder - - - - - W - - - - - W
Extra info - - - - - - - - X_1 X_2 X_3 X_4
Decoder \( \hat{X}_1 \) \( \hat{X}_2 \) \( \hat{X}_3 \) \( \hat{X}_4 \) \( \hat{X}_5 \)

Here, block length = 5, delay = 6 time units.
Feed-forward ⇒ Decoder knows some of the past source samples.

FF with delay $k$, block length $N$.

To produce $\hat{X}_n$, the decoder knows index $W$ and $(X_1, \ldots, X_{n-k})$. 
Stock Market Example

- Behavior of a particular stock over an $N$-day period.
- Stock price modeled as a $k + 1$-state Markov chain.
- Value of stock: Markov source $\{X_n\}$
Insider and Investor

- Insider - *a priori* knowledge about behavior of stock
- Investor has stock for $N$ days, needs to know when value drops.
- Insider: gives information to investor at rate $R$. 
Reconstruction

- Decision of investor on day $n$: $\hat{X}_n$
  - $\hat{X}_n = 1 \Rightarrow$ price drop from day $n - 1$ to $n$
  - Otherwise $\hat{X}_n = 0$
  - Distortion $= 1$ if $\hat{X}_n$ is wrong
Reconstruction

- Decision of investor on day \( n \): \( \hat{X}_n \)
- \( \hat{X}_n = 1 \) ⇒ price drop from day \( n - 1 \) to \( n \)
- Otherwise \( \hat{X}_n = 0 \)
- Distortion = 1 if \( \hat{X}_n \) is wrong
Feed-forward!

- Before day $n$, investor knows previous stock values $X^{n-1}$, makes decision $\hat{X}_n$.
- Minimum info (in bits/sample) needed to predict drops with distortion $D$?
Feed-Forward: A Formal Definition

Previous: [Weissman et al 03], [Pradhan 04], [Martinian et al 04]

- **Source** $X$: Alphabet $\mathcal{X}$, reconstruction alphabet $\hat{\mathcal{X}}$
- **Encoder**: Rate $R$, $e : \mathcal{X}^N \rightarrow \{1, \ldots, 2^{NR}\}$
- **Decoder**: knows all the past $(n - k)$ source samples to reconstruct $n$th sample.

$$g_n : \{1, \ldots, 2^{NR}\} \times \mathcal{X}^{n-k} \rightarrow \hat{\mathcal{X}}, \quad n = 1, \ldots, N.$$
A Formal Definition (contd.)

- Distortion measure $d_N(X^N, \hat{X}^N)$.

GOAL

Given any source $X$, find the least $R$ such that

$$E[d_N(x^N, \hat{x}^N)] \leq D.$$ 

- Rate-Distortion function with Feed-forward!
• Inspired by [Marko ’73]: work on bidirectional communication
• [Massey ’90] The directed information flowing from $A^N$ to $B^N$

$$I(A^N \rightarrow B^N) = \sum_{n=1}^{N} I(A^n; B_n|B^{n-1}).$$

• Interestingly:

$$I(A^N \rightarrow B^N) = I(A^N; B^N) - \sum_{n=2}^{N} I(B^{n-1}; A_n|A^{n-1})$$

$$I(A^N \rightarrow B^N) = I(A^N; B^N) - I(0B^{N-1} \rightarrow A^N)$$
Directed Information

- Inspired by [Marko '73]: work on bidirectional communication
- [Massey '90] The directed information flowing from $A^N$ to $B^N$

\[ I(A^N \to B^N) = \sum_{n=1}^{N} I(A^n; B_n|B^{n-1}). \]

- Interestingly:

\[ I(A^N \to B^N) = I(A^N; B^N) - \sum_{n=2}^{N} I(B^{n-1}; A_n|A^{n-1}) \]

\[ I(A^N \to B^N) = I(A^N; B^N) - I(0B^{N-1} \to A^N) \]
Causality in Information Flow

\[ I(A^N; B^N) = I(A^N \rightarrow B^N) + I(0B^{N-1} \rightarrow A^N) \]

- \( I(A^N \rightarrow B^N) \): How causal knowledge of \( A_n \)'s reduces the uncertainty in \( B_n \)
- Example: GNP vs Money Supply [Geweke '82]
Without FF, need $I(\hat{X}^N; X^N)$ bits to represent $X^N$ with $\hat{X}^N$.

- With feed-forward, to produce $\hat{X}_n$, the decoder knows $X^{n-1}$.
- Number of bits required is reduced by $I(\hat{X}_n; X^{n-1} | \hat{X}^{n-1})$. 
Without FF, need $I(\hat{X}^N; X^N)$ bits to represent $X^N$ with $\hat{X}^N$.

- With feed-forward, to produce $\hat{X}_n$, the decoder knows $X^{n-1}$.
- Number of bits required is reduced by $I(\hat{X}_n; X^{n-1}|\hat{X}^{n-1})$. 

Diagram: 
- Encoder to Decoder with Rate $R$
- Delay $N+1$ feedback to Encoder
With delay 1 feed-forward, we need

\[ I(\hat{X}^N; X^N) - \sum_{n=2}^{N} I(\hat{X}_n; X^{n-1} | \hat{X}^{n-1}) \]  \text{bits.} \]

Directed information from $\hat{X}^N$ to $X^N$!
With delay $k$ feed-forward, to produce $\hat{X}_n$, the decoder knows $X^{n-k}$.

No. of bits: $I(\hat{X}^N; X^N) - \sum_{n=k+1}^N I(\hat{X}_n; X^{n-k} | \hat{X}^{n-1})$

Not Directed Information - will denote it $I_k(\hat{X}^N \rightarrow X^N)$
- ‘$k$–directed information’.

Rate-distortion function
- Optimize $I_k$ over $P_{\hat{X}^N|X^N}$ that satisfy distortion
- No savings for discrete memoryless source w/ single-letter distortion measure
Delay $k$ feed-forward

With delay $k$ feed-forward, to produce $\hat{X}_n$, the decoder knows $X^{n-k}$.

No. of bits: $I(\hat{X}^N; X^N) - \sum_{n=k+1}^{N} I(\hat{X}_n; X^{n-k} | \hat{X}^{n-1})$

Not Directed Information- will denote it $I_k(\hat{X}^N \rightarrow X^N)$
- ‘$k$–directed information’.

Rate-distortion function

Optimize $I_k$ over $P_{\hat{X}^N|X^N}$ that satisfy distortion

No savings for discrete memoryless source w/ single-letter distortion measure
Even when source is stationary, ergodic:
- with feed-forward, optimal joint distribution may not be.

- Source could be non-stationary, non-ergodic
- Sequence of distortion functions $d_n(.,.)$

- Need information-spectrum methods [Han, Verdu '93, '95]
Even when source is stationary, ergodic:
- with feed-forward, optimal joint distribution may not be.

- Source could be non-stationary, non-ergodic
- Sequence of distortion functions $d_n(.,.)$

- Need information-spectrum methods [Han, Verdu ’93, ’95]
Definitions

\( a_1, a_2, \ldots : \text{random sequence} \)

- \( \lim \sup_{\text{in prob}} a_n = \bar{a} \): Smallest number \( \alpha \) such that
  \[
  \lim_{n \to \infty} \Pr(a_n > \alpha) = 0.
  \]

We will need

\[
i_k(\hat{x}^n \rightarrow x^n) = \frac{1}{n} \log \frac{P(x^n, \hat{x}^n)}{P(x^n) \cdot \prod_{i=1}^{n} P(\hat{x}_i|\hat{x}^{i-1}, x^{i-k})}
\]
Definitions

\( a_1, a_2, \ldots : \text{random sequence} \)

\( \limsup_{\text{in prob}} a_n = \overline{a} \): Smallest number \( \alpha \) such that

\[
\lim_{n \to \infty} \Pr(a_n > \alpha) = 0.
\]

We will need

\[
i_k(\hat{x}^n \rightarrow x^n) = \frac{1}{n} \log \frac{P(x^n, \hat{x}^n)}{P(x^n) \cdot \prod_{i=1}^{n} P(\hat{x}_i | \hat{x}^{i-1}, x^{i-k})}
\]

\[
\text{pdf}
\]

\[
i(\hat{x} \rightarrow x)
\]
Theorem for a general source

\[ \mathbf{P}_X = \{P_{X_1}, P_{X_2}, \ldots, P_{X_N}, \ldots\} \]

\[ \mathbf{P}_{\hat{X}|X} = \{P_{\hat{X}_1|X_1}, P_{\hat{X}_2|X_2}, \ldots, P_{\hat{X}_N|X_N}, \ldots\} \]

Theorem

For arbitrary source \( X \) with distribution \( \mathbf{P}_X \), the rate-distortion function with feed-forward delay \( k \), the infimum of all achievable rates at probability-1 distortion \( D \), is given by

\[
R_{ff}(D) = \inf_{\mathbf{P}_{\hat{X}|X}: \rho(\mathbf{P}_{\hat{X}|X}) \leq D} \bar{I}_k(\hat{X} \rightarrow X),
\]

where

\[
\rho(\mathbf{P}_{\hat{X}|X}) \triangleq \limsup_{inprob} d_n(x^n, \hat{x}^n)
\]
Rate-Distortion Theorem for a general source

- $P_X = \{P_{X_1}, P_{X^2}, \ldots, P_{X^N}, \ldots\}$
- $P_{\hat{X}|X} = \{P_{\hat{X}_1|X_1}, P_{\hat{X}^2|X^2}, \ldots, P_{\hat{X}^N|X^N}, \ldots\}$

**Theorem**

For arbitrary source $X$ with distribution $P_X$, the rate-distortion function with feed-forward delay $k$, the infimum of all achievable rates at probability-1 distortion $D$, is given by

$$R_{ff}(D) = \inf_{P_{\hat{X}|X}: \rho(P_{\hat{X}|X}) \leq D} I_k(\hat{X} \rightarrow X),$$

where

$$\rho(P_{\hat{X}|X}) \triangleq \limsup_{n \to \infty} \frac{d_n(x^n, \hat{x}^n)}{\text{inprob}}$$
What is feed-forward in source coding?
Directed information and why it occurs
Feed-forward rate-distortion result for general sources, distortions

Next ...

How to compute the rate-distortion function?
The story so far...

- What is feed-forward in source coding?
- Directed information and why it occurs
- Feed-forward rate-distortion result for general sources, distortions

Next...

How to compute the rate-distortion function?
Source $X$ with distribution $P_X$.

Find $P_{\hat{X}|X}$ to minimize $\bar{T}_k(\hat{X} \rightarrow X)$ s.t

$$\limsup d_n(X^n, \hat{X}^n) \leq D$$

in prob

Multi-letter optimization - difficult!
Source $X$ with distribution $P_X$.

Find $P_{\hat{X}|X}$ to minimize $\bar{T}_k(\hat{X} \rightarrow X)$ s.t

$$\limsup d_n(X^n, \hat{X}^n) \leq D$$

in prob

Multi-letter optimization - difficult!
Pick a conditional distribution $P_{\hat{X}|X} = \{P_{\hat{X}_n|X_n}\}$.

For what sequence of distortion measures $d_n$ does $P_{\hat{X}|X}$ achieve the infimum in the rate-distortion formula?

Approach—similar in spirit to [Csiszar and Korner]
A stationary, ergodic source $X$ characterized by $P_X = \{P_X^n\}_{n=1}^{\infty}$ with feed-forward delay $k$. $P_{\hat{X}|X} = \{P_{X^n|X^n}\}_{n=1}^{\infty}$ is a conditional distribution such that the joint distribution is stationary and ergodic. Then $P_{\hat{X}|X}$ achieves the rate-distortion function if for all sufficiently large $n$, the distortion measure satisfies

$$d_n(x^n, \hat{x}^n) = -c \cdot \frac{1}{n} \log \frac{P_{X^n,\hat{X}^n}(x^n, \hat{x}^n)}{P_{\hat{X}^n|X^n}(\hat{x}^n|x^n)} + d_0(x^n),$$

where

$$P_{\hat{X}^n|X^n}(\hat{x}^n|x^n) = \prod_{i=1}^{n} P_{\hat{X}_i|X^{i-k},\hat{X}^{i-1}}(\hat{x}_i|x^{i-k}, \hat{x}^{i-1})$$

and $c$ is any positive number and $d_0(\cdot)$ is an arbitrary function.
Value of the stock: Markov source \( \{X_n\} \)

Decision of investor on day \( n \): \( \hat{X}_n \) (0 or 1)
Stock example revisited

- Value of the stock: Markov source \( \{X_n\} \)

- Decision of investor on day \( n \): \( \hat{X}_n \) (0 or 1)
- $R_{ff}(D)$: Minimum rate the investor needs to predict drops with distortion $D$.

- Try first-order Markov joint distribution
- Distortion can be cast in the required form!
Proposition

The minimum rate in (bits/sample) is

\[
\sum_{i=1}^{k-1} \pi_i \left[ h(p_i, q_i, 1 - p_i - q_i) - h(\epsilon, 1 - \epsilon) \right] \\
\quad + \pi_k \left( h(q_k, 1 - q_k) - h(\epsilon, 1 - \epsilon) \right)
\]

where

- \( h() \) is the entropy function
- \( \left[ \pi_0, \pi_1, \ldots, \pi_k \right] \) is the stationary distribution of the stock
- \( \epsilon = \frac{D}{1 - \pi_0} \)
Computing Rate-distortion function with FF

1. ‘Predict’ a conditional distribution
2. Check if distortion function can be put into required form.

Next . . . A multi-terminal problem
Computing Rate-distortion function with FF

1. ‘Predict’ a conditional distribution
2. Check if distortion function can be put into required form.

Next ... A multi-terminal problem
- Source $X$: compressed into packets
- Packets may be dropped
- Compress $X$ into two different packets
Multiple Descriptions

- Source X: compressed into packets
- Packets may be dropped
- Compress X into two different packets
Rate $R_1$ yields reconstruction $\hat{X}_1$ with distortion $D_1$
Rate $R_2$ yields reconstruction $\hat{X}_2$ with distortion $D_2$

- Want better quality $D_0$ if both packets received
- $\hat{X}_1$ and $\hat{X}_2$ need to refine each other!
**GOAL**

Given i.i.d source $P_X$:

Find all achievable $(R_1, R_2, D_1, D_2, D_0)$

- Still an open problem
- Studied by [Cover, El Gamal], [Ahlswede], [Zhang, Berger], . . .
- Best known rate-region: [Zhang, Berger '87]
Rate

$R_1$ bits/sample

$R_2$ bits/sample

$R_1 + R_2$ bits/sample

Distortion

$D_1$

$D_2$

$D_0$

GOAL

Given i.i.d source $P_X$:

Find all achievable $(R_1, R_2, D_1, D_2, D_0)$

- Still an open problem
- Studied by [Cover, El Gamal], [Ahlswede], [Zhang, Berger], . . .
- Best known rate-region: [Zhang, Berger '87]
Source propagates to destination with delay

To reconstruct $\hat{X}_n$, decoder has packet and $X^{n-k}$
Alice has an i.i.d binary source $\sim$ Bernoulli(1/2)

Bob and Carol: distortion $d$ using $R_B, R_C$ bits/sample

Dave gets Bob’s bits and Carol’s bits- needs to reconstruct perfectly!

Characterize

$$r_{\text{sum}}(d) \triangleq \text{Smallest sum-rate } R_B + R_C \text{ for distortion } (d, d, 0)$$
Same model as before, one extra feature...  

- After Carol reconstructs each sample, Alice reveals the value to her.  

*Feed-forward*

- Before reconstructing each sample, Carol knows past source samples

Characterize with feed-forward

\[ r_{\text{sum}}(d) \triangleq \text{Smallest sum-rate } R_B + R_C \text{ for } (d, d, 0) \]
Same model as before, one extra feature...  

- After Carol reconstructs each sample, Alice reveals the value to her. 

*Feed-forward*

- Before reconstructing each sample, Carol knows *past* source samples

**Characterize with feed-forward**

\[ r_{\text{sum}}(d) \triangleq \text{Smallest sum-rate } R_B + R_C \text{ for } (d, d, 0) \]
Encoder mappings: $e_m : \mathcal{X}^N \rightarrow \{1, \ldots, 2^{NR_m}\}, \quad m = 1, 2.$

Mappings for decoders 1 and 0:

$g_1 : \{1, \ldots, 2^{NR_1}\} \rightarrow \hat{\mathcal{X}}_1^N$

$g_0 : \{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\} \rightarrow \hat{\mathcal{X}}_0^N$

A sequence of mappings for decoder 2:

$g_{2n} : \{1, \ldots, 2^{NR_2}\} \times \mathcal{X}^{n-k} \rightarrow \hat{\mathcal{X}}_2, \quad n = 1, \ldots, N.$
General feed-forward model

- Encoder mappings: $e_m : \mathcal{X}^N \rightarrow \{1, \ldots, 2^{NR_m}\}, \; m = 1, 2$.
- Mappings for decoders 1 and 0:
  \[
  g_1 : \{1, \ldots, 2^{NR_1}\} \rightarrow \mathcal{\hat{X}}_1^N
  
  g_0 : \{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\} \rightarrow \mathcal{\hat{X}}_0^N
  \]
- A sequence of mappings for decoder 2:
  \[
  g_{2n} : \{1, \ldots, 2^{NR_2}\} \times \mathcal{X}^{n-k} \rightarrow \mathcal{\hat{X}}_2, \; n = 1, \ldots, N.
  \]
Zhang-Berger '87

\[ P_{U, \hat{X}_1, \hat{X}_2, \hat{X}_0 \mid X} \text{ such that} \]

\[ Ed_m(X; \hat{X}_m) \leq D_m, \quad m = 0, 1, 2 \]

\[ R_1 > I(X; \hat{X}_1 U) \quad R_2 > I(X; \hat{X}_2 U) \]

\[ R_1 + R_2 > I(X; \hat{X}_1 U) + I(X; \hat{X}_2 U) + I(X; \hat{X}_0 \mid \hat{X}_1 \hat{X}_2 U) + I(\hat{X}_1; \hat{X}_2 \mid XU) \]
Correlation in MD

Stringent Decoder 0 distortion $\Rightarrow$ Need correlation in $\hat{X}_1, \hat{X}_2$

‘Cloud’ Center

- $U$: Cloud center of $X$ sent to all decoders
- Rate $I(U; X)$ each for decoder 1 and 2

Penalty Term

- Can’t have $\hat{X}_1 \sim P(\hat{X}_1|XU)$ and $\hat{X}_2 \sim P(\hat{X}_2|XU)$ indep’ly
- Need to be jointly distributed: $\sim P(\hat{X}_1, \hat{X}_2|XU)$
- $I(\hat{X}_1; \hat{X}_2|XU)$: Penalty in sum-rate

FF: decoder 2 knows past samples with some delay
- Can help build correlation!
Stringent Decoder 0 distortion $\Rightarrow$ Need correlation in $\hat{X}_1, \hat{X}_2$

### ‘Cloud’ Center
- $U$: Cloud center of $X$ sent to all decoders
- Rate $I(U; X)$ each for decoder 1 and 2

### Penalty Term
- Can’t have $\hat{X}_1 \sim P(\hat{X}_1|XU)$ and $\hat{X}_2 \sim P(\hat{X}_2|XU)$ indep’ly
- Need to be *jointly* distributed: $\sim P(\hat{X}_1, \hat{X}_2|XU)$
- $I(\hat{X}_1; \hat{X}_2|XU)$: Penalty in sum-rate

- FF: decoder 2 knows past samples with some delay
- Can help build correlation!

---

39 / 47
Stringent Decoder 0 distortion ⇒ Need correlation in $\hat{X}_1, \hat{X}_2$

### ‘Cloud’ Center
- $U$: Cloud center of $X$ sent to all decoders
- Rate $I(U; X)$ each for decoder 1 and 2

### Penalty Term
- Can’t have $\hat{X}_1 \sim P(\hat{X}_1|XU)$ and $\hat{X}_2 \sim P(\hat{X}_2|XU)$ indep’ly
- Need to be jointly distributed: $\sim P(\hat{X}_1, \hat{X}_2|XU)$
- $I(\hat{X}_1; \hat{X}_2|XU)$: Penalty in sum-rate

- FF: decoder 2 knows past samples with some delay
- Can help build correlation!
Coding Strategy

- Consider $B$ long blocks of source samples
  
  $X_1 \ldots X_N \quad X_{N+1} \ldots X_{2N} \quad \ldots \ldots \quad X_{NB}$
  
- While encoding one block, give ‘preview’ of next block

![Diagram showing the coding strategy with blocks and preview mechanism.](image-url)
Coding Strategy

Restricted encoding for user 1- within cell \( j \)

After reconstructing block \( b \):
- User 1 gets 'preview' of block \( b + 1 \)
- User 2 knows it too- due to FF!

Block-Markov, superposition; [Cover,Leung], [Willems] for MAC
After reconstructing block $b$:
- User 1 gets 'preview' of block $b + 1$
- User 2 knows it too- due to FF!

Block-Markov, superposition; [Cover,Leung], [Willems] for MAC
Theorem

\( P_{U, \hat{X}_1, \hat{X}_2, \hat{X}_0 | X} \) such that

\[ Ed_m(X ; \hat{X}_m) \leq D_m, \quad m = 0, 1, 2 \]

\[ R_1 > I(X ; \hat{X}_1 U) \]

\[ R_2 > I(X ; \hat{X}_2 | U) + \max\{0, R_1 - I(X \hat{X}_2 ; \hat{X}_1 | U)\} \]

\[ R_1 + R_2 > I(X ; \hat{X}_1 U) + I(X ; \hat{X}_2 | U) + I(X ; \hat{X}_0 | \hat{X}_1 \hat{X}_2 U) + I(\hat{X}_1 ; \hat{X}_2 | XU) + \max\{0, R_1 - I(X \hat{X}_2 ; \hat{X}_1 | U)\} \]
Improvement over [Zhang-Berger]

- Fix $R_1 = I(X; \hat{X}_1 U) + \epsilon$
- If $\max\{0, R_1 - I(X \hat{X}_2; \hat{X}_1 | U)\} = 0$:
  - Savings in $R_2 = I(U; X)$ bits/sample
  - FF conveys ‘cloud center’ $U$ for free
- If $\max\{0, R_1 - I(X \hat{X}_2; \hat{X}_1 | U)\} \neq 0$:
  - Savings in $R_2 = I(\hat{X}_1; \hat{X}_2 | XU)$ bits/sample
  - FF eliminates penalty term
Fix $R_1 = I(X; \hat{X}_1 U) + \epsilon$

If $\max\{0, R_1 - I(X \hat{X}_2; \hat{X}_1 | U)\} = 0$:
- Savings in $R_2 = I(U; X)$ bits/sample
- FF conveys ‘cloud center’ $U$ for free

If $\max\{0, R_1 - I(X \hat{X}_2; \hat{X}_1 | U)\} \neq 0$:
- Savings in $R_2 = I(\hat{X}_1; \hat{X}_2 | XU)$ bits/sample
- FF eliminates penalty term
Example: No feed-forward

\[ r_{sum}(d) \geq 2 - h \left( \frac{4d + 1 - \sqrt{12d^2 - 4d + 1}}{2} \right) \]

Without FF [Zhang Berger '87]
(a): Lower bound without feed-forward [Zhang, Berger ’87]
(b): Achievable sum-rate with FF- better than optimal w/o FF
(c): Rate-dist lower bound with FF- $R_B + R_C > 2(1 - h(d))$
(a): Lower bound without feed-forward [Zhang, Berger ’87]
(b): Achievable sum-rate with FF- better than optimal w/o FF
(c): Rate-dist lower bound with FF- $R_B + R_C > 2(1 - h(d))$
(a): Lower bound without feed-forward [Zhang, Berger '87]
(b): Achievable sum-rate with FF- better than optimal w/o FF
(c): Rate-dist lower bound with FF- $R_B + R_C > 2(1 - h(d))$
Summary

- Feed-forward: helps build correlation
- Single-letter achievable region for MD with FF
- FF helps even for an i.i.d source

- How to use feed-forward to all decoders?
- FF for Gaussian multiple descriptions [Pradhan IT ’07]
Summary

- Source Coding with feed-forward
  - Directed Information
  - Rate-Distortion Function
  - How to evaluate optimization
  - Multiple Descriptions with FF

- Channel Coding with feedback

Some questions...

- Feedback/FF in multi-terminal setting
- Noisy feedback/FF
- Applications of directed-info like quantities
### Summary
- Source Coding with feed-forward
  - Directed Information
  - Rate-Distortion Function
  - How to evaluate optimization
  - Multiple Descriptions with FF
- Channel Coding with feedback

### Some questions...
- Feedback/FF in multi-terminal setting
- Noisy feedback/FF
- Applications of directed-info like quantities