Evaluating the Rate-Distortion Function of Sources with Feed-Forward and the Capacity of Channels with Feedback.

Ramji Venkataramanan
S. Sandeep Pradhan

Dept. of EECS, University of Michigan
Outline

- Source coding with feed-forward.
- Rate-distortion function - how to evaluate?
- Example
  - Feedback capacity of a channel.
  - How to evaluate?
Outline

- Source coding with feed-forward.
- Rate-distortion function—how to evaluate?
- Example
- Feedback capacity of a channel.
- How to evaluate?
What is feed-forward?

The source field itself may be available in a delayed form at the decoder.

<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>Extra info</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X_1</td>
<td>X_2</td>
<td>X_3</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>

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

The source field itself may be available in a delayed form at the decoder.

<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>$W$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Extra info</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</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>

Here, block length= 5, delay is 6 time units.
Source Coding with Feed-Forward

- **Feed-forward** ⇒ Decoder knows some of the past source samples.

![Diagram of source coding with feed-forward](image)

Feed-forward with delay $k$, block length $N$.

- To reconstruct $X_n$, the decoder knows index $W$ and $(X_1, \ldots, X_{n-k})$.

- Applications in other areas too...
Source Coding with Feed-Forward

- **Feed-forward**: Decoder knows some of the past source samples.

![Diagram of Feed-forward system]

Feed-forward with delay $k$, block length $N$.

- To reconstruct $X_n$, the decoder knows index $W$ and $(X_1, \ldots, X_{n-k})$.
- Applications in other areas too...
Feed-Forward: A Formal Definition

- [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 $(i-k)$ source samples to reconstruct $i$th sample.

$$g_i : \{1,\ldots,2^{NR}\} \times \mathcal{X}^{i-k} \rightarrow \hat{\mathcal{X}}, \quad i = 1,\ldots,N.$$
Feed-Forward: A Formal Definition

- [Weissman et al 03], [Pradhan 04], [Martinian et al 04]

Source \( X \): Alphabet \( \mathcal{X} \), reconstruction alphabet \( \mathcal{\hat{X}} \)

Encoder: Rate \( R \), \( e : \mathcal{X}^N \rightarrow \{1, \ldots, 2^{NR}\} \)

Decoder: knows all the past \( (i-k) \) source samples to reconstruct \( i \)th sample.

\[
g_i : \{1, \ldots, 2^{NR}\} \times \mathcal{X}^{i-k} \rightarrow \mathcal{\hat{X}}, \quad i = 1, \ldots, N.
\]
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!
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 $X^{n-k}$.
- Number of bits required is reduced by $I(\hat{X}_n; X^{n-k} | \hat{X}^{n-k})$. 
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 $X^{n-k}$.

Number of bits required is reduced by $I(\hat{X}_n; X^{n-k} | \hat{X}^{n-k})$. 
Intuition contd...

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

Will denote it \( I_k(\hat{X}^N \rightarrow X^N) \)- ‘\( k \)–directed information’.

Massey’s Directed Information for \( k = 1 \).
Intuition contd...

![Diagram of encoder and decoder with feedback](image)

- No. of bits: \( I(\hat{X}^N; X^N) - \sum_{n=k+1}^{N} I(\hat{X}_n; X^{n-k} | \hat{X}^{n-1}) \)
- Will denote it \( I_k(\hat{X}^N \rightarrow X^N) - 'k-directed information'. \)
- Massey’s Directed Information for \( k = 1 \).
General source, general distortion measure

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

- Even when source is stationary and ergodic, with feed-forward, the optimal joint distribution may not be.
- Need to use information-spectrum methods [Han, Verdu]
General source, general distortion measure

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

- Even when source is stationary and ergodic, with feed-forward, the optimal joint distribution may not be.
- Need to use information-spectrum methods [Han, Verdu]
General source, general distortion measure

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

- Even when source is stationary and ergodic, with feed-forward, the optimal joint distribution may not be.
- Need to use information-spectrum methods [Han, Verdu]
Definitions

- \( P_X = \{ P_{X_1}, P_{X_2}, \ldots, P_{X_N}, \ldots \} \)
- \( P_{\hat{X}|X} = \{ P_{\hat{X}_1|X_1}, P_{\hat{X}_2|X_2}, \ldots, P_{\hat{X}_N|X_N}, \ldots \} \)

- \( a_1, a_2, \ldots : \) random sequence

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

- \( P_X = \{P_{X_1}, P_{X_2}, \ldots, P_{X_N}, \ldots\} \)
- \( P_{\hat{X}|X} = \{P_{\hat{X}_1|X_1}, P_{\hat{X}_2|X_2}, \ldots, P_{\hat{X}_N|X_N}, \ldots\} \)

- \( a_1, a_2, \ldots : \) random sequence

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

We will need

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

\[
i_k(\hat{x}^n \rightarrow x^n) = \frac{1}{n} \log \frac{P(x^n, \hat{x}^n)}{P(x^n) \cdot \prod_{i=1}^{n} P(\hat{x}_i | \hat{x}_i^{-1}, x^{i-k})}
\]
Rate-Distortion Theorem for a general source

[IT Trans. June 07]

Theorem

\[ R_{ff}(D) = \min \bar{I}_k(\hat{X} \rightarrow X), \]

where

\[ \min \text{ is over } P_{\hat{X}|X} \text{ such that } \limsup_{n \to \infty} d_n(x^n, \hat{x}^n) \leq D \]
Source Coding Optimization

- Source $X$ with distribution $P_X$.

- Multi-letter optimization - difficult!
Given source $\mathbf{P}_X = \{P_{X^n}\}$

Pick a conditional distribution $\mathbf{P}_{\hat{X}|X} = \{P_{\hat{X}_n|X^n}\}$

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

$\mathbf{P}_{\hat{X}|X}$ has to minimize $\overline{I}_k(\hat{X} \rightarrow X)$ over the set

$$Q(D) = \{W_{\hat{X}|X} : \limsup_{\text{in prob } PW} d_n(X^n, \hat{X}^n) \leq D\}.$$
Source Coding Optimization

- Given source $P_X = \{P_{X^n}\}$
- 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?
- $P_{\hat{X}|X}$ has to minimize $\bar{I}_k(\hat{X} \rightarrow X)$ over the set

$$Q(D) = \{W_{\hat{X}|X} : \limsup_{in \text{prob } PW} d_n(X^n, \hat{X}^n) \leq D\}.$$  

- Approach- similar in spirit to [Csiszar and Korner], [Gastpar et al], [Pradhan et al]
Structure of Distortion Function

Theorem

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

\[
d_n(x^n, \hat{x}^n) = -c \cdot \frac{1}{n} \log \frac{P_{X^n, \hat{X}^n}(x^n, \hat{x}^n)}{\bar{P}_{\hat{X}^n | X^n}(\hat{x}^n | x^n)} + d_0(x^n),
\]

where

\[
\bar{P}_{\hat{X}^n | X^n}(\hat{x}^n | x^n) = \prod_{i=1}^{n} P_{\hat{X}_i | X^{i-k}, \hat{X}^{i-1}}(\hat{x}_i | x^{i-k}, \hat{x}^{i-1})
\]

and \( c \) is any positive number and \( d_0(\cdot) \) is an arbitrary function.
Stock Market Example

- Behavior of a particular stock over an $N$-day period.
- Value of the stock modeled as a $k+1$-state Markov chain.

Investor has this stock over an $N$–day period, needs to be forewarned whenever the value drops.
- There is an insider with a priori knowledge about the behavior of the stock.
- Can give information to the investor at a cost $c$/bit of info.
Stock Market Example

- Behavior of a particular stock over an $N$-day period.
- Value of the stock- modeled as a $k + 1$-state Markov chain.

Investor has this stock over an $N$–day period, needs to be forewarned whenever the value drops.

There is an insider with *a priori* knowledge about the behavior of the stock.

Can give information to the investor at a cost $c$/bit of info.
Stock Price Model

- Value of the stock: Markov source \( \{X_n\} \)
- Decision of investor on day \( n \): \( \hat{X}_n \)
- \( \hat{X}_n = 1 \Rightarrow \text{price is going to drop from day } n - 1 \text{ to } n, \hat{X}_n = 0 \text{ means otherwise.} \)
- Hamming distortion:
  - Distortion 1 when investor fails to predict drop, or falsely predicts.
- Before day \( n \), investor knows all the previous values of the stock \( X^{n-1} \), has to make the decision \( \hat{X}_n \) - feed-forward!
Stock Price Model

- Value of the stock: Markov source \( \{X_n\} \)
- Decision of investor on day \( n \): \( \hat{X}_n \)
  \( \hat{X}_n = 1 \Rightarrow \) price is going to drop from day \( n - 1 \) to \( n \), \( \hat{X}_n = 0 \) means otherwise.
- Hamming distortion:
  Distortion 1 when investor fails to predict drop, or falsely predicts.
- Before day \( n \), investor knows all the previous values of the stock \( X^{n-1} \), has to make the decision \( \hat{X}_n \) - feed-forward!
Stock Market Example

- $R_{ff}(D)$: Minimum information (in bits/sample) the investor needs to predict drops in value with distortion $D$.
- Try first-order Markov conditional distribution.

**Proposition**

For the stock-market problem described above,

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

where $h()$ is the entropy function, $[\pi_0, \pi_1, \ldots, \pi_k]$ is the stationary distribution of the Markov chain and $\epsilon = \frac{D}{1-\pi_0}$. 
Stock Market Example

- $R_{ff}(D)$: Minimum information (in bits/sample) the investor needs to predict drops in value with distortion $D$.
- Try first-order Markov conditional distribution.

**Proposition**

*For the stock-market problem described above,*

$$R_{ff}(D) = \sum_{i=1}^{k-1} \pi_i \left[ h(p_i, q_i, 1 - p_i - q_i) - h(\epsilon, 1 - \epsilon) \right]$$

$$+ \pi_k \left( h(q_k, 1 - q_k) - h(\epsilon, 1 - \epsilon) \right),$$

*where $h()$ is the entropy function, $[\pi_0, \pi_1, \cdots, \pi_k]$ is the stationary distribution of the Markov chain and $\epsilon = \frac{D}{1 - \pi_0}$.***
Channel with Feedback

Channel: \( \{ P^{ch}(Y_n|X^n, Y^{n-1}) \} \).

Input Distribution \( P^{k}_{X|Y} \): \( \{ P(X_n|X^{n-1}, Y^{n-k}) \} \)

Massey, Kramer, Tatikonda,...
Channel Capacity

[Tatikonda, Mitter]

- **Capacity**: \( \max I(X \rightarrow Y) \)
  
  \[ \max \text{ over } P_X^k: \{P(X_n | X^{n-1}, Y^{n-k})\} \]

Given input distribution \( P_X^k \), for what sequence of cost measures does \( P_X^k \) achieve the maximum in the capacity formula?
Channel Capacity

[Tatikonda, Mitter]

- **Capacity**: \( \max \ I(X \rightarrow Y) \)

\[
\max \text{ over } P_X^k: \{ P(X_n|X^{n-1}, Y^{n-k}) \}
\]

Given input distribution \( P_X^k \), for what sequence of cost measures does \( P_X^k \) achieve the maximum in the capacity formula?
Cost function for feedback channels

Theorem

Suppose we are given a channel $P_{Y|X}^{ch}$ with $k-$delay feedback and an input distribution $P_{X|Y}^k$ such that the joint process $P_{X,Y}$ is stationary, ergodic. Then the input distribution $P_{X|Y}^k$ achieves the $k-$delay feedback capacity of the channel if for all sufficiently large $n$, the cost measure satisfies

$$c_n(x^n, y^n) = \lambda \cdot \frac{1}{n} \log \frac{\tilde{P}_{Y^n|x^n}^{ch}(y^n|x^n)}{P_{Y^n}(y^n)} + d_0,$$

where $\lambda$ is any positive number and $d_0$ is an arbitrary constant.
Cost function for feedback channels

Theorem

Suppose we are given a channel $P_{Y|X}^c$ with $k$-delay feedback and an input distribution $P_{X|Y}^k$ such that the joint process $P_{X,Y}$ is stationary, ergodic. Then the input distribution $P_{X|Y}^k$ achieves the $k$-delay feedback capacity of the channel if for all sufficiently large $n$, the cost measure satisfies

$$c_n(x^n, y^n) = \lambda \cdot \frac{1}{n} \log \frac{\bar{P}_{Y^n|X^n}^c(x^n|y^n)}{P_{Y^n}(y^n)} + d_0,$$

where $\lambda$ is any positive number and $d_0$ is an arbitrary constant.
Rate-distortion function with FF, channel capacity with FB:

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