Information Theory and Channel Coding

Prof. Rodrigo C. de Lamare
CETUC, PUC-Rio, Brazil
delamare@cetuc.puc-rio.br

V. Rate distortion theory

- In this chapter, we study rate distortion theory and how to control the level of distortion or loss of information in encoding strategies.

- Typical situations that benefit from rate-distortion theory include those with constraints that force source coding to be lossy.

- For example, a communication channel might impose constraints on the transmission rate, which requires compression beyond the entropy rate.

- For speech and audio signals we often require quantization to obtain a representation in codewords with sufficiently short codelengths.

- In general, we have to deal with rate requirements that inevitably lead to lossy compression, which requires control of the level of distortion.

- In particular, we focus on source coding with a fidelity criterion and situations in which we must perform lossy signal compression.

- We consider a mathematical model of a source coding system and explore how it can benefit from lossy compression.

- We introduce the rate distortion function and develop an approach to computing the rate using constrained optimization of mutual information.

The applications that we are interested include:

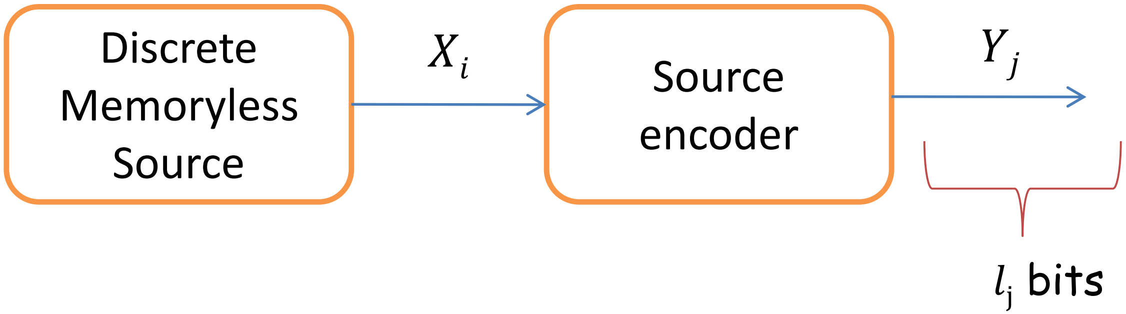
a) Source coding:

- Lossy compression using quantization
- Source codes that do not represent completely the source

b) Data transmission at a rate greater than the channel capacity, that is, when $R > C$

A. Mathematical model

- Let us consider a DMS defined by an M-ary alphabet and the random variable $x = \{X_i |\ i = 1,2, \ldots, M\}$ that produces symbols X_i .



- This alphabet is assumed to produce symbols s_i that are statistically independent with probabilities $p_i, i = 1,2, \ldots, M$.

- The symbols of the source are inputs to an encoder that produces Y_j that can be converted into codewords c_j with length l_j bits.

- The average code rate is described by

$$R \text{ bits /codeword}$$

- At the output of the encoder, the codewords could also be represented by an N-ary alphabet through the random variable $y = \{Y_j | j = 1,2, \dots, N\}$.

- By the source coding theorem, we have

  - Lossless coding: $R \geq H(x) \rightarrow$ perfect representation of the source

  - Lossy coding: $R < H(x) \rightarrow$ loss of information

B. Rate distortion function

- Consider the joint pdf $p_{xy}(X_i, Y_j)$ that describes the occurrence of symbol X_i at the input of the channel and its output representation Y_j related by

$$p_{xy}(X_i, Y_j) = p_y(Y_j|X_i)p_x(X_i),$$

where $p_y(Y_j|X_i)$ is the transition probability of the encoder.

- The distortion measure associated to the representation of X_i by Y_j is given by

$$d(X_i, Y_j),$$

where $d(X_i, Y_j)$ is also referred to as the distortion measure of a single symbol.

- Examples of distortion measures include
  - Hamming distortion:

$$d(X_i, Y_j) = \begin{cases} 0, & X_i = Y_j \\ 1, & X_i \neq Y_j \end{cases}$$

  - Squared error distortion:

$$d(X_i, Y_j) = (X_i - Y_j)^2$$

  - Mean-squared error distortion:

$$d(X_i, Y_j) = E\left[(X_i - Y_j)^2\right]$$

- The average distortion of all possible source symbols and of the encoding representation is described by

$$\bar{d} = \sum_{i=1}^{M} \sum_{j=1}^{N} p_x\,(X_i) p_y\left(Y_j \middle| X_i\right) d\left(X_i, Y_j\right),$$

where $\bar{d}$ is a continuous non negative function of $p_y\left(Y_j \middle| X_i\right)$, which are determined by the encoder/decoder pair.

- The transition probabilities $p_y\left(Y_j \middle| X_i\right)$ are said to be D –admissible if and only if $\bar{d} \leq D$, where D is a chosen distortion value.

- The set of all allocations of D –admissible conditional probabilities is given by

$$P_D = \{p_y(Y_i|X_i): \bar{d} \leq D\}$$

- For each set of transition probabilities $p_y(Y_j|X_i)$, we have the mutual information described by

$$I(x,y) = \sum_{i=1}^{M} \sum_{j=1}^{N} p_y(Y_j|X_i) p_x(X_i) \log_2\left(\frac{p_y(Y_j|X_i)}{p_y(Y_j)}\right) \text{ bits/codeword}$$

- The rate distortion function is defined by

$$R(D) = \min_{p_y(Y_i|X_i) \in \mathrm{P_D}} I(x,y) \text{ bits/codeword}$$

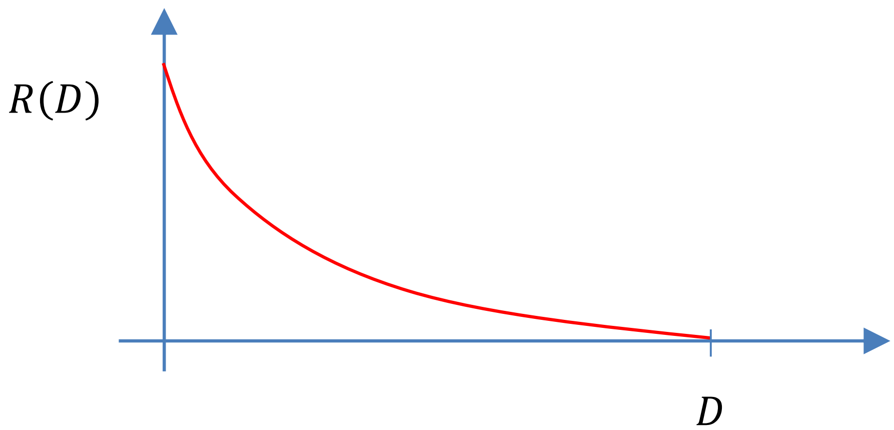
subject to the constraint

$$\sum_{j=1}^{N} p_y(Y_i|X_i) = 1, \ \text{for } i = 1,2, \dots, M$$

where $\bar{d} \le D$ for the computation of $R(D)$ and

P_D is the set to which $p_y(Y_i|X_i)$ belongs and ensures a distortion D.

- The previous optimization leads to the following illustration.

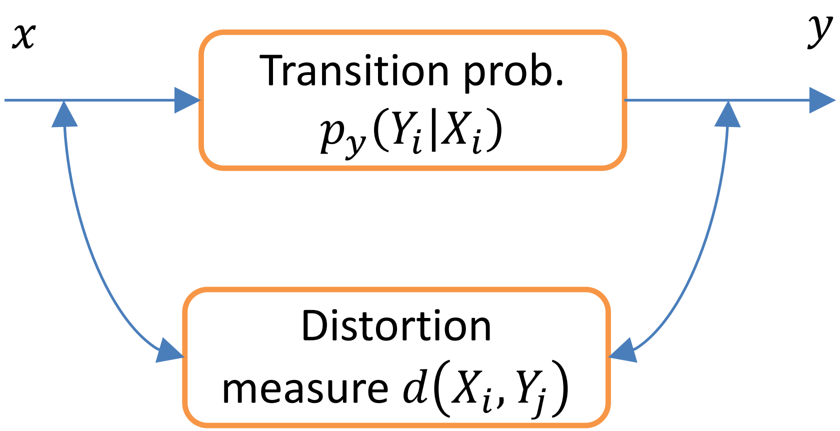


- When the distortion D is reduced $\rightarrow$ $R(D)$ increases.

- When the distortion D is increased $\rightarrow$ $R(D)$ decreases.

C. Computation of the rate distortion function

- In order to compute the rate distortion function, we consider the transition probabilities $p_y(Y_i|X_i)$ and proceed as follows:



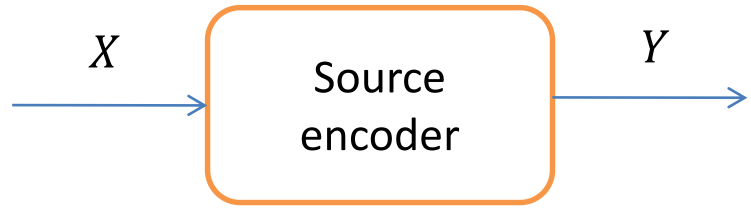
- Computation of $R(D)$:
  - Given $d(X_i, Y_j)$.
  - Compute $p_y(Y_i|X_i)$ that minimize $I(x, y)$ subject to constraints:

$$R(D) = \min_{p_y(Y_i|X_i) \in \mathrm{P_D}} I(x, y) \text{ subject to the constraint } \sum_{j=1}^{N} p_y(Y_i|X_i) = 1$$

Example 1

Consider a discrete memoryless source that outputs Gaussian random variables x with zero mean and variance σ^2.

We consider an encoder that quantizes X and produces $Y = \mathrm{Q}(\mathrm{X}) = \hat{X}$ through the mean squared error distortion measure given by



$$d(x, y) = E[(x - y)^2]$$

a) Compute the rate distortion function $R(D)$.

b) Determine the transition probability that achieves the lower bound of $R(D)$.

Solution:

a) We consider a Gaussian random variable x with zero mean and variance σ^2, i.e., $x \sim N(0, \sigma^2)$.

By extending the optimization that leads to the computation of the rate distortion function, we obtain

$$R(D) = \min_{p_y(Y_i|X_i) \in \mathrm{P_D}} I(x, y)$$

$$\text{subject to } p_y(Y|X): d(x, y) = E[(x-y)^2] \leq D$$

We first find a lower bound for the rate distortion function and then prove that this is achievable. $E[(x-y)^2] \leq D$, we observe that

$$
\begin{aligned}
I(x,y) &= \mathrm{h(x)} - \mathrm{h(x|y)} \\
&= \frac{1}{2}\log_2 2\pi e\sigma^2 - \mathrm{h(x-y|y)} = \frac{1}{2}\log_2 2\pi e\sigma^2 - \mathrm{h}(\mathrm{x} - \hat{x}|\hat{x}) \\
&\geq \frac{1}{2}\log_2 2\pi e\sigma^2 - \mathrm{h}(\mathrm{x} - \hat{x}) \\
&\geq \frac{1}{2}\log_2 2\pi e\sigma^2 - \mathrm{h}(N(0, E[(x-y)^2])) \\
&= \frac{1}{2}\log_2 2\pi e\sigma^2 - \frac{1}{2}\log_2 2\pi e E[(x-y)^2] \\
&\geq \frac{1}{2}\log_2 2\pi e\sigma^2 - \frac{1}{2}\log_2 2\pi e D \\
&= \frac{1}{2}\log_2 \frac{\sigma^2}{D}
\end{aligned}
$$

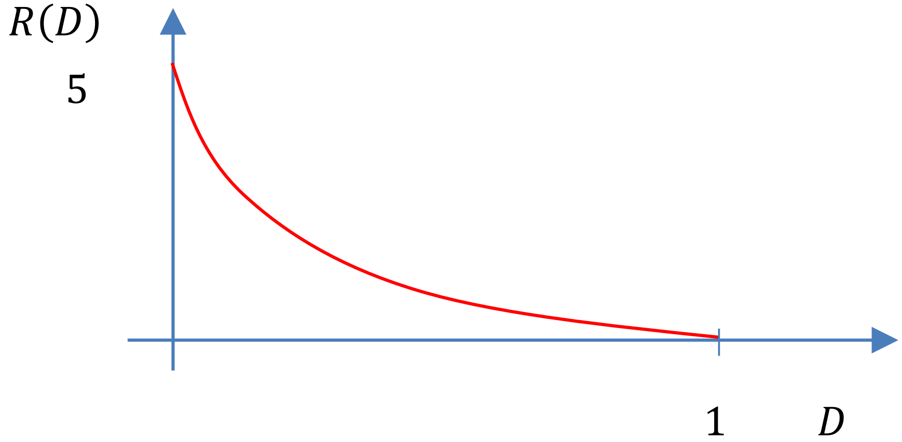
Conditioning reduces entropy

Gaussian distribution maximizes entropy

Therefore, we have

$$R(D) \geq \frac{1}{2}\log_2 \frac{\sigma^2}{D} \text{ bits / output}$$

The rate distortion function for a Gaussian source is illustrated by



b)

In order to find the transition probability $p_y(Y|X)$ that achieves the lower bound of item a), it is often more convenient to look at $p_y(X|Y)$which is sometimes called the *test channel*.

We construct $p_y(X|Y)$ to achieve equality in the bound. If $D \le \sigma^2$, we choose

$$x = y + w, \qquad \text{where } x \sim N(0, \sigma^2), w \sim N(0, D) \text{ and } y \sim ?$$

and y and w are independent.

We need to find the contribution to the mutual information of y that yields

$$I(x, y) = \frac{1}{2}\log_2 \frac{(\overbrace{?}^{\sigma^2 - D} + D)}{D} = \frac{1}{2}\log_2 \frac{\sigma^2}{D}$$

This requires the distribution of y to be Gaussian with

$$y \sim N(0, \sigma^2 - D)$$

The test channel can be illustrated by