Lecture 4. Rate-distortion Theorem and Lossy Coding

Lin Zhang

Tsinghua-Berkeley Shenzhen Institute

Shenzhen, China, 2017

1. Motivation

2. Simple Examples

3. Lossy Source Coding

4. Summary

Outline

1. Motivation

2. Simple Examples

3. Lossy Source Coding

4. Summary

Coding revisited

- Lossless Coding: represent the information efficiently without any loss (distortion)?
- Channel Coding: Increase the redundancy of the sequence to combat the noise.
- These two efforts are all **entropy preserving**, a.k.a. no information is lost in Shannon notion.
- The question is, **is entropy preserving coding always necessary?**

Image Compression

Original

Image Compression

107k Byte

Image Compression

55k Byte

Image Compression

24k Byte

Image Compression

10k Byte

Image Compression

4k Byte

Video Compression

NTSC frame rate.

	High quality	Typical	Low bit rate
"Akiyo" Video conference Studio recording			
CPU Load (%) PSNR (dB) Bit Rate (kb/s)	9.1% 44.2 dB 305 kb/s	7.2% 38 dB 62 kb/s	6.7% 35.2 dB 33 kb/s
"Coastguard" Video Surveillance Outdoor Traveling camera			
CPU Load (%) PSNR (dB) Bit Rate (kb/s)	23 % 41.2 dB 2750 kb/s	17 % 33.5 dB 644 kb/s	14 % 30.8 dB 300 kb/s
"Foreman" Outdoor Hand-held camera			
CPU Load (%) PSNR (dB) Bit Rate (kb/s)	21 % 41.5 dB 1895 kb/s	15 % 34.9 dB 385 kb/s	14 % 32.3 dB 188 kb/s

Audio Codec

The Intuitions and Questions

- Is it necessary to complete encode the information?
  - **Not necessary** for a lot of information sources in nature.
  - **Not possible** for all continuous sources.
- The problem then becomes
  - How to **optimally** encode the information sources given **a finite bit rate**?
- What is OPTIMAL?
  - Smaller distortion is better?
  - How to define distortion?

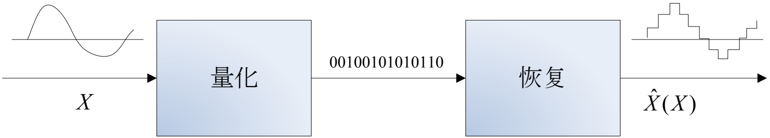
Outline

1. Motivation
2. Simple Examples
3. Lossy Source Coding
4. Summary

Quantization of Scalar Gaussian RVs

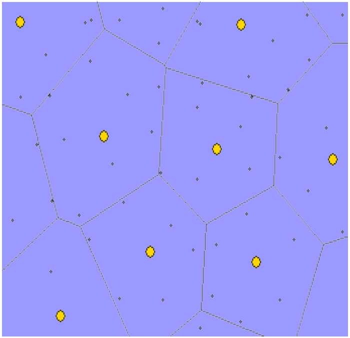
- Consider a random variable $X \sim N(0, \sigma^x)$.
- Use R bits to represent X.
- Distortion is measured by mean square error

$$E(X - \hat{X}(X))^2 = \int_{-\infty}^{\infty} (x - \hat{X}(x))dx \tag{1}$$



Quantization of Scalar Gaussian RVs (cntd)

- If $R = 1$, the solution is obvious.
- If $R > 1$, the solution is no longer straightforward.
  - There are all together 2^R reconstruction points to be selected.
  - S. P. Loyd proposed an iterative algorithm to converge to the optimal coding scheme.



The Voronoi Constellation

Outline

1. Motivation

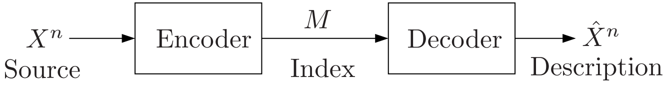
2. Simple Examples

3. Lossy Source Coding

4. Summary

Lossy Source Coding

- A DMS X is encoded (described) at rate R by the encoder. The decoder receives the description index over a noiseless link and generates a reconstruction (estimate) $\hat{X}$ of the source with a prescribed distortion D. What is the optimal tradeoff between the communication rate R and distortion between X and the estimate $\hat{X}$



Measurement of Distortion

- The distortion criterion is defined as follows. Let $\hat{\mathcal{X}}$ be a *reproduction* alphabet and define a *distortion measure* as a mapping

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty) \tag{2}$$

- It measures the cost of representing the symbol x by the symbol $\hat{x}$ The average per-letter distortion between x^n and $\hat{x}^n$ is defined as

$$d(x^n, \hat{x}^n) := \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \tag{3}$$

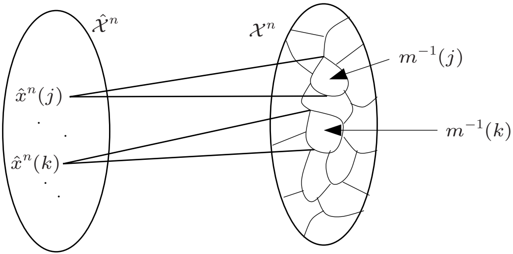
- Example: Hamming distortion (loss): Assume $\mathcal{X} = \hat{\mathcal{X}}$. The Hamming distortion is the indicator for an error, i.e.,

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x}, \\ 1 & \text{if } x \neq \hat{x} \end{cases} \tag{4}$$

$d(x^n, \hat{x}^n)$ is the fraction of symbols in error (bit error rate for the binary alphabet)

Formal Definition of Lossy Source Coding

- Formally, a $(2^{nR}, n)$ rate-distortion code consists of:
  1. An encoder that assigns to each sequence $x^n \in \mathcal{X}^n$ an index $m(x^n) \in [1 : 2^{nR})$, and
  2. A decoder that assigns to each index $m \in [1 : 2^{nR})$ an estimate $\hat{x}^n(m) \in \mathcal{X}^n$.



The set $\mathcal{C} = \left\{\hat{x}^n(1), ..., \hat{x}^n(2^{\lfloor nR \rfloor})\right\}$ constitutes the *codebook*, and the sets $m^{-1}(1), \ldots, m^{-1}(2^{\lfloor nR \rfloor}) \in \mathcal{X}^n$ are the *associated assignment regions*

Rate-Distortion Pair and Rate-Distortion Function

- The distortion associated with the $(2^{nR}, n)$ code is

$$E(d(X^n, \hat{X}^n)) = \sum_{x^n} p(x^n) d(x^n, \hat{x}^n(m(x^n))) \tag{5}$$

- A rate-distortion pair (R, D) is said to be **achievable** if there exists a sequence of $(2^{nR}, n)$ rate-distortion codes with

$$\limsup_{n \to \infty} E(d(X^n, \hat{X}^n)) \le D \tag{6}$$

- The **rate-distortion function** $R(D)$ is the infimum of rates R such that (R, D) is achievable.

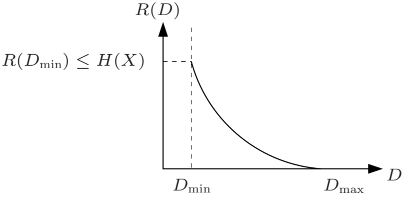
Lossy Source Coding Theorem

- **Shannon's Lossy Source Coding Theorem** : The rate-distortion function for a DMS $(X, p(x))$ and a distortion measure $d(x, \hat{x})$ is

$$R(D) = \min_{p(\hat{x}|x):E(d(x,\hat{x}))\leq D} I(X;\hat{X}) \tag{7}$$

  for $D \geq D_{min} := E(min_{\hat{x}} d(X, \hat{x}))$

- $R(D)$ is nonincreasing and convex (and thus continuous) in $D \in [D_{min}, D_{max}]$, where $D_{max} := min_{\hat{x}} E(d(X, \hat{x}))$



Without loss of generality we assume throughout that $D_{min} = 0$, i.e., for every $x \in \mathcal{X}$, there exists an $\hat{x} \in \hat{\mathcal{X}}$ such that $d(x, \hat{x}) = 0$.

R-D Functions Examples

- The rate-distortion function for a Bern(p) source $X, p \in [0, 1/2]$, with Hamming distortion (loss) is

$$R(D) = \begin{cases} H(p) - H(D) & \text{for } 0 \leq D < p, \\ 0 & \text{for } D \geq p \end{cases} \tag{8}$$

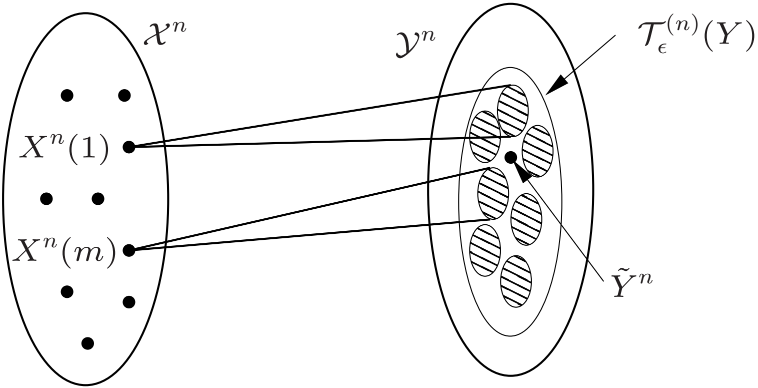
- The rate-distortion function for a Gaussian source $X \sim N(0, \sigma^2)$ with mean-square distortion (loss) is

$$R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & \text{for } 0 \leq D < \sigma^2, \\ 0 & \text{for } D \geq \sigma^2 \end{cases} \tag{9}$$

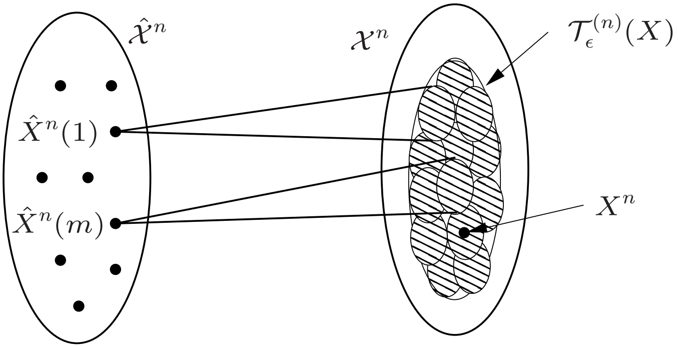
Proof of Lossy Source Coding Theorem

- Achievability
  - Random code generation: $p(\hat{x}|x)$
  - Encoding: Joint typicality encoding
  - Decoding: Simple mapping
  - Analysis of distortion
- Converse

Packing Lemma



Covering Lemma



Outline

1. Motivation
2. Simple Examples
3. Lossy Source Coding
4. Summary

Summary

- Lossless Coding is neither Necessary nor Possible
- Rate-Distortion Tradeoff
- Rate-Distortion Function and its Properties
- Lossy Source Coding Theorem
  - Achievability
  - Converse
- Revisit: Packing lemma and Covering Lemma

Reference

- *Lin Zhang*, "Lecture notes on Fundamentals of applied information theory", *', in Chinese.*
- *Claude E. Shannon*, "A Mathematical Theory of Communication", *Bell System Technical Journal, 1948.*
- *Cover T M, Thomas J A.*, "Elements of information theory", *John Wiley and Sons, 2012.*
- *Xuelong Zhu*, "Fundamentals of applied information theory", *Tsinghua Univ. Press, 2001, in Chinese.*