Chapter 10

Rate Distortion Theory

Peng-Hua Wang

Graduate Inst. of Comm. Engineering

National Taipei University

Chapter Outline

Chap. 10 Rate Distortion Theory

10.1 Quantization

10.2 Definitions

10.3 Calculation of the Rate Distortion Function

10.4 Converse to the Rate Distortion Theorem

10.5 Achievability of the Rate Distortion Function

10.6 Strongly Typical Sequences and Rate Distortion

10.7 Characterization of the Rate Distortion Function

10.8 Computation of Channel Capacity and the Rate Distortion Function

10.1 Quantization

Introduction

- Finite representation of a continuous r.v.
  - can't be perfect
  - How well can we do ? $\Rightarrow$ need to define "goodness" or distortion measurement means the distance between a r.v. and its representation
  - This is in fact the lossy compression

Quantization

- Let X be a r.v., $\hat{X} = \hat{X}(X)$ be its representation.
- If we quantize X into R bits, means we use 2^R distinct values to represent X.
- Problem. Find optimal set $\hat{X}$, called the representation points or code points, and the region associated with each value in $\hat{X}$ such that certain error measurement is minimized.

Example. $X \sim \mathcal{N}(0, \sigma^2)$, $R = 1$, $\min E[(X - \hat{X})^2]$.

$$\hat{X} = \begin{cases} \sqrt{\frac{2}{\pi}}\sigma, & X \geq 0 \\ -\sqrt{\frac{2}{\pi}}\sigma, & X < 0 \end{cases}$$

Quantization

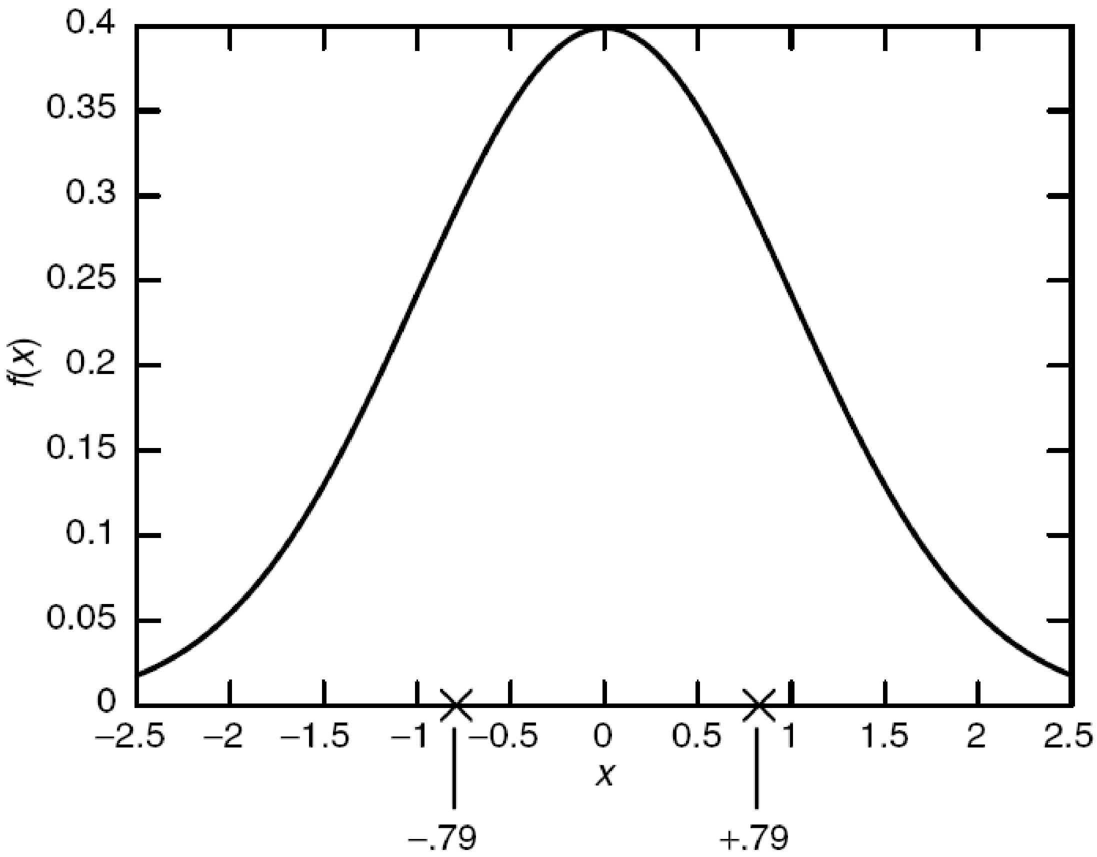

FIGURE 10.1. One-bit quantization of Gaussian random variable.

10.2 Definitions

Definitions

Definition 1 (Distortion Function)

$$d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^+$$

which means a function d *with* $d(x, \hat{x}) \geq 0$ *for* $x \in \mathcal{X}$, $\hat{x} \in \hat{\mathcal{X}}$.

- A distortion function $d(x, \hat{x})$ is **bounded** if $\max d(x, \hat{x}) < \infty$
- The distortion between sequences x and $\hat{x}$ is defined by

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

- Example.
  - **Hamming distance.** $d(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and $d(x, \hat{x}) = 0$ if $x = \hat{x}$.
  - **squared-error distortion.** $d(x, \hat{x}) = (x - \hat{x})^2$.

Definitions

Definition 2 (Rate Distortion Codes) *A* $2^{nR}, n$*-rate distortion code consists of an encoding function*

$$f_n : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$$

and a decoding function

$$g_n : \{1, 2, \ldots, 2^{nR}\} \to \mathcal{X}^n.$$

The distortion D *associated with the code is the average distortion over all codewords*

$$D = E[d(X^n, g_n(f_n(X^n)))] = \sum_{x^n} p(x^n) d(x^n, g_n(f_n(x^n)))$$

We may call $\hat{X}^n$ the vector quantization, reproduction, reconstruction, source code, or estimation of X.

Definitions

Definition 3 (Achievable) *A rate distortion pair (R, D) is said to be achievable if there exists a sequence of $(2^{nR}, n)$-rate distortion code (f_n, g_n) with*

$$\lim_{n\to\infty} E[d(X^n, g_n(f_n(X^n)))] \le D.$$

- A **rate distortion region** for a source is the closure of the set of achievable rate distortion pairs (R, D).
- The **rate distortion function** $R(D)$ for a source is the infimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.
- The **rate distortion function** $R(D)$ for a source is the infimum of all distortion D such that (R, D) is in the rate distortion region of the source for a given rate R.

Definitions

Definition 4 (Information Rate Distortion Function) *The information rate distortion function* $R^{(I)}(D)$ *for a source* X *with distortion measure* $d(x, \hat{x})$ *is defined as*

$$R^{(I)}(D) = \min_{p(\hat{x}|x): \sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) \le D} I(X; \hat{X})$$

Theorem 1 (Rate Distortion Function) *The rate distortion function for an i.i.d. source* X *with distribution* $p(x)$ *a nd bounded distortion function* $d(x, \hat{x})$ *is equal to the associated information rate distortion function. Thus,* $R(D) = R^{(I)}(D)$ *is the minimum achievable rate at distortion* D*.*

10.3 Calculation of the Rate Distortion Function

Binary Source

Theorem 2 *The rate distortion function for a Bernoulli(p) source with Hamming distortion is given by*

$$R(D) = \begin{cases} H(p) - H(D), & 0 \le D \le \min\{p, 1-p\} \\ 0, & D > \min\{p, 1-p\} \end{cases}$$

- If $D \ge p$, we can achieve $R(D) = 0$ (one code to represent two values) by letting $\hat{X} = 0$ since the distortion is

$$\begin{aligned} & p(x=1, \hat{x}=0) \times d(x=0, \hat{x}=1) \\ =& p(x=1) \underbrace{p(\hat{x}=0|x=1)}_{=1} \times 1 = p \end{aligned}$$

Binary Source

Proof. Let $Z = 0$ if $X = \hat{X}$ and $Z = 1$ if $X \neq \hat{X}$. Denote $p(Z = 1) = t$. The distortion

$$\begin{aligned} E[d(X, \hat{X}] &= p(X = 0, \hat{X} = 1) \times 1 + p(X = 1, \hat{X} = 0) \times 1 \\ &= p(Z = 1) = t \leq D \end{aligned}$$

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) = H(p) - H(Z|\hat{X}) \\ &\geq H(p) - H(Z) = H(p) - H(t) \geq H(p) - H(D) \end{aligned}$$

Equality holds when $H(X|\hat{X}) = H(D)$.

Binary Source

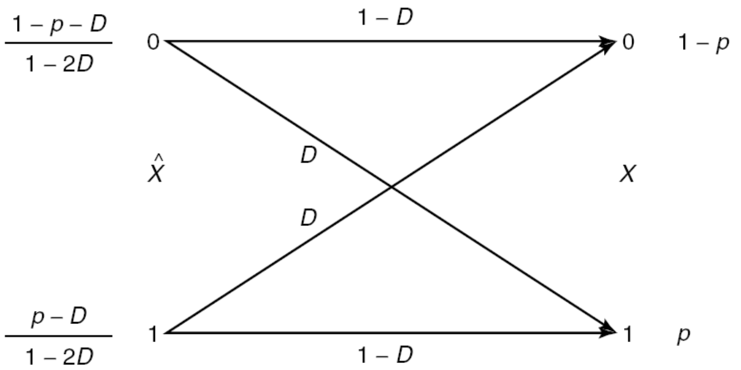

FIGURE 10.3. Joint distribution for binary source.

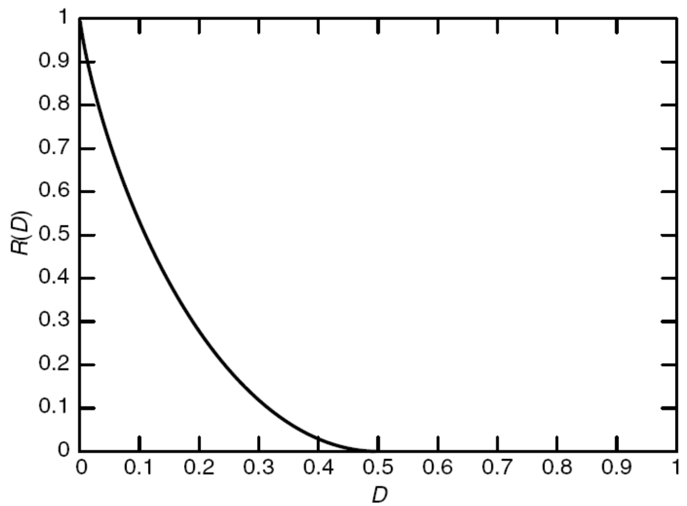

FIGURE 10.4. Rate distortion function for a Bernoulli $(\frac{1}{2})$ source.

Gaussian Source

Theorem 3 *The rate distortion function for a $\mathcal{N}(0,\sigma^2)$ source with squared error distortion is*

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \le D \le \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

- If $D \ge \sigma^2$, we can achieve $R(D) = 0$ (one code to represent ALL values) by letting $\hat{X} = 0$ since the distortion is

$$\int (x-\hat{x})^2 \phi(x) dx = \int x^2 \phi(x) dx = \sigma^2$$

where $\phi(x)$ is the pdf of $\mathcal{N}(0,\sigma^2)$.

Gaussian Source

Proof.

$$
\begin{aligned}
I(X;\hat{X}) &= h(X) - h(X|\hat{X}) = h(X) - h(X-\hat{X}|\hat{X}) \\
&\geq h(X) - h(X-\hat{X}) \geq h(X) - h(\mathcal{N}(0, E[(X-\hat{X})^2])) \\
&= \frac{1}{2}\log(2\pi e\sigma^2) - \frac{1}{2}\log(2\pi e E[(X-\hat{X})^2]) \\
&\geq \frac{1}{2}\log(2\pi e\sigma^2) - \frac{1}{2}\log(2\pi e D) \\
&= \frac{1}{2}\log\frac{\sigma^2}{D}
\end{aligned}
$$

Equality holds when $Z = X - \hat{X}$ has a normal distribution of zero mean and variance D.

Gaussian Source

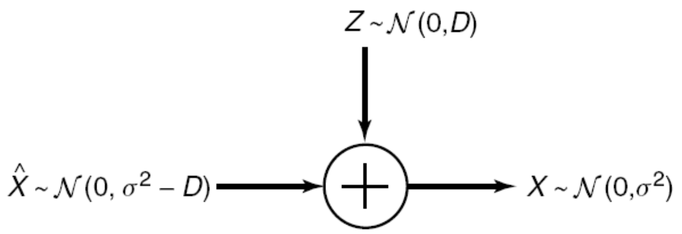

FIGURE 10.5. Joint distribution for Gaussian source.

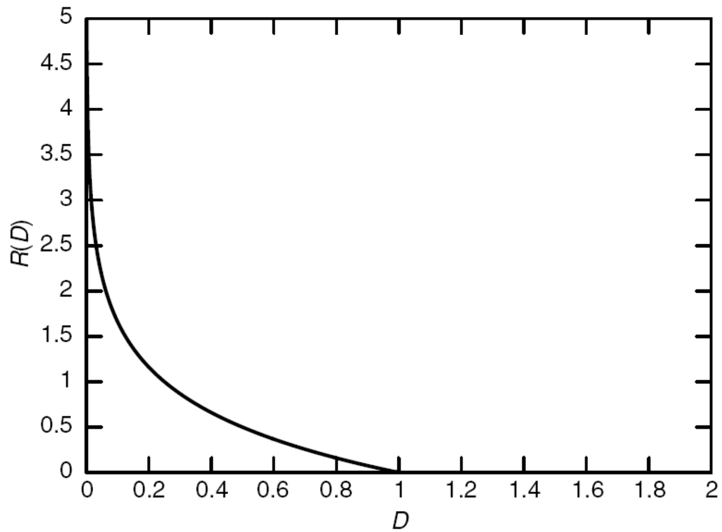

FIGURE 10.6. Rate distortion function for a Gaussian source.

Sphere Packing for Channel Coding

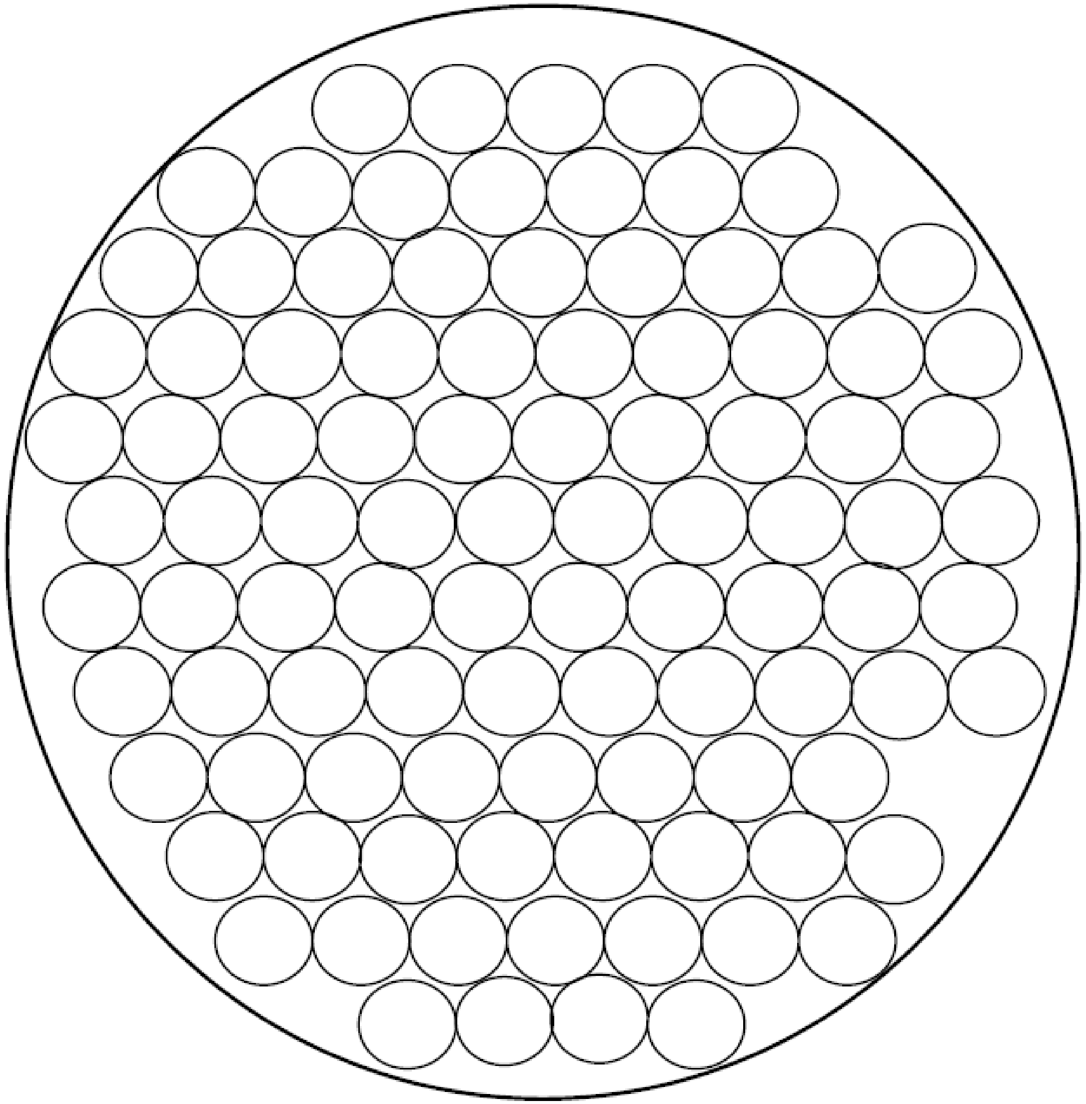

FIGURE 9.2. Sphere packing for the Gaussian channel.

Sphere Packing for Channel Coding

For each sent codeword, the received codeword is contained in a sphere of radius $\sqrt{nN}$. The received vectors have energy no grater than $n(P+N)$, so they lie in a sphere of radius $\sqrt{n(P+N)}$. How many codeword can we use without intersection in the decoding sphere?

$$M = \frac{A_n\left(\sqrt{n(P+N)}\right)^n}{A_n(\sqrt{nN})^n} = \left(1+\frac{P}{N}\right)^{n/2}$$

where A the constant for calculating the volume of n-dimensional sphere. For example, $A_2 = \pi$, $A_3 = \frac{4}{3}\pi$. Therefore, the capacity is

$$\frac{1}{n}\log M = \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

Sphere Packing for Rate Distortion

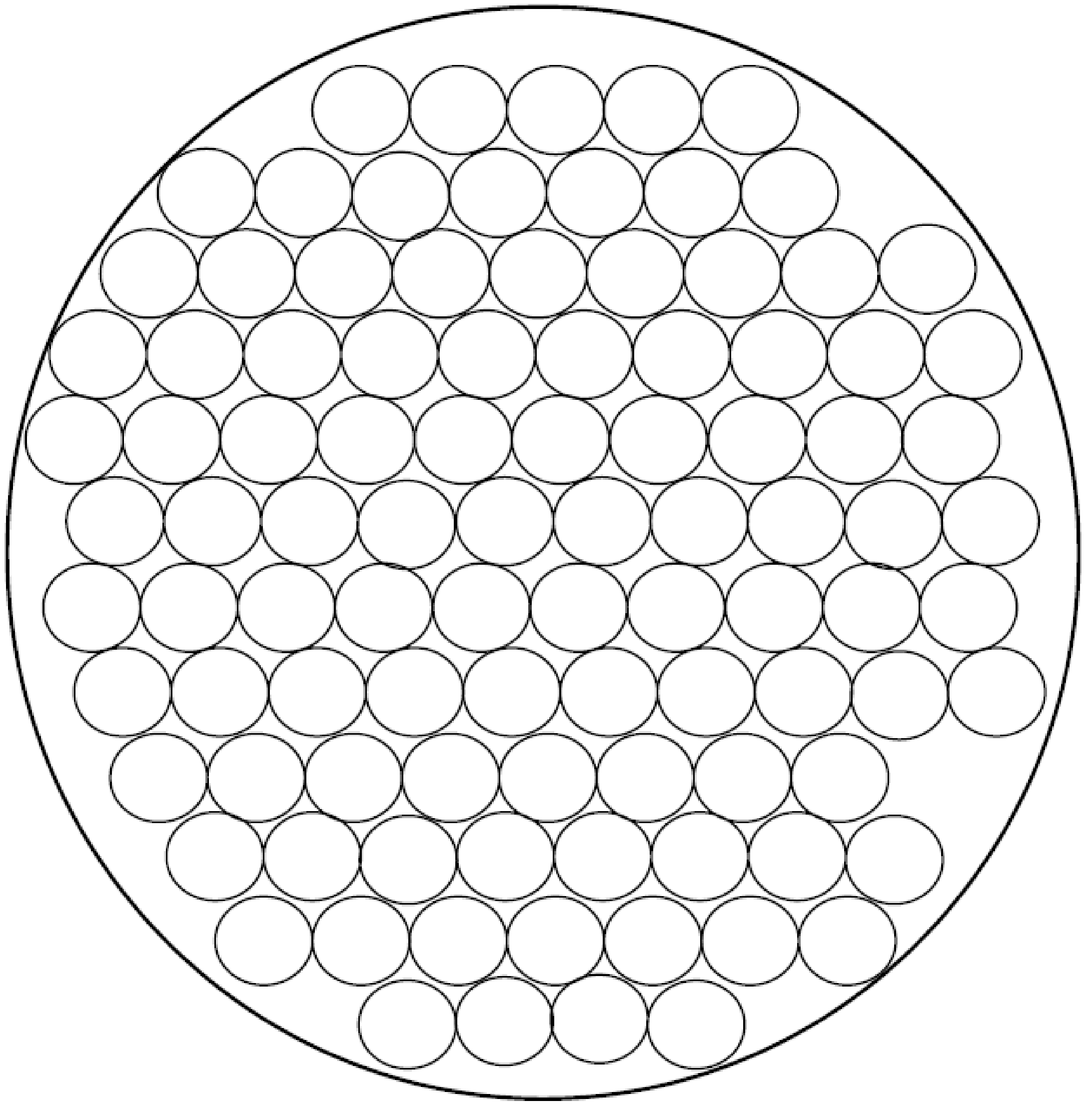

FIGURE 9.2. Sphere packing for the Gaussian channel.

Sphere Packing for Rate Distortion

Consider a Gaussian source of variance σ^2. A $(2^{nR}, n)$ rate distortion code for this source with distortion D is a set of $M = 2^{nR}$ sequences in $\mathcal{R}^n$. All these sequences lie within a sphere of radius $\sqrt{n\sigma^2}$. The source sequences are within a distance $\sqrt{D}$ of some codewords. How many codeword can we use without intersection in the decoding sphere?

$$M = \frac{A_n\left(\sqrt{n\sigma^2}\right)^n}{A_n(\sqrt{nD})^n} = \left(\frac{\sigma^2}{D}\right)^{n/2}$$

Therefore, the rate is

$$\frac{1}{n}\log M = \frac{1}{2}\log\frac{\sigma^2}{D}.$$