ECE253/CSE208 Introduction to Information Theory

Lecture 15: Rate-Distortion Theory

Dr. Yu Zhang

ECE Department
University of California, Santa Cruz

- Chap 10 of *Elements of Information Theory (2nd Edition) by Thomas Cover & Joy Thomas*.

Lossy Compression

- Given a random source, we would like to compress it as much as possible.
- The source coding theorem says that the compression limit is the entropy if we insist on lossless compression; i.e., perfect reconstruction.
- However, in many applications we can tolerate distortion (the mismatch between source symbol and its reconstruction) to a certain extent.
- For example, JPEG/MPEG for images and MP3/AAC/Ogg Vorbis for audio signals.
- Given a source distribution and a distortion measure, what is the minimum rate description (number of bits per symbol) required to achieve a particular distortion?
- Or, what is the minimum expected distortion achievable at a particular rate?

Rate–distortion Theory

- Rate–distortion theory is the theoretical foundation of *lossy* data compression: It addresses the problem of the optimal tradeoff between data rate and distortion.
- We aim to determine the minimal number of bits per symbol which is communicated over a channel, so that the source can be approximately reconstructed at the receiver without exceeding an expected distortion D.

Distortion Measure

Definition (Distortion measure)

A single-letter distortion measure is a mapping

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+.$$

The value $d(x, \hat{x})$ denotes the distortion incurred when a source symbol and its reconstruction is x and $\hat{x}$, respectively.

Bounded distortion: $d_{\max} \triangleq \max_{x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x}) < \infty$.

Example (Distortion functions)

1. Hamming distortion: $d(x, \hat{x}) = \mathbb{1}_{\{x = \hat{x}\}}$.
2. Absolute-error distortion: $d(x, \hat{x}) = |x - \hat{x}|$.
3. Square-error distortion: $d(x, \hat{x}) = (x - \hat{x})^2$.

The distortion between sequences x^n and $\hat{x}^n$ is given by averaging per-letter distortion:

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

Rate-distortion Code

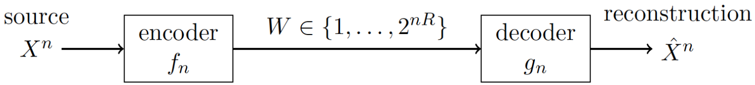


Figure: Rate distortion encoder and decoder: $\hat{X}^n$ is called as vector quantization, reproduction, reconstruction of X^n.

Definition

A $(2^{nR}, n)$-rate distortion code consists of an encoding function $f_n : \mathcal{X}^n \to \{1, 2, \dots, 2^{nR}\}$, and a decoding function $g_n : \{1, 2, \dots, 2^{nR}\} \to \hat{\mathcal{X}}^n$. The associated distortion is defined as

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

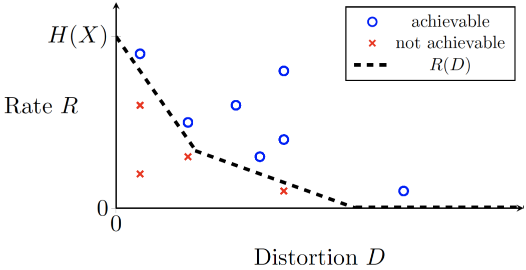
Rate-distortion region

Definition (Achievable rate-distortion pair)

A rate distortion pair (R, D) is said to be achievable if there exists a sequence of $(2^{nR}, n)$-rate distortion codes (f_n, g_n) with $\lim_{n\to\infty} \mathrm{E}[d(X^n, \hat{X}^n)] \leq D$.

Definition (Rate distortion region)

The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R, D).



Rate-distortion Function

Definition (Rate distortion function)

The rate distortion function $R(D)$ is the infimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.

Definition (Distortion rate function)

The distortion rate function $D(R)$ is the infimum of all distortions D such that (R, D) is in the rate distortion region of the source for a given rate R.

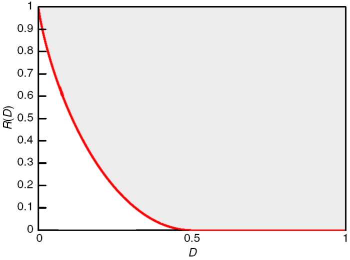


Figure: The rate-distortion function of a Bernoulli random variable with Hamming distortion

Information Rate-distortion Function

Definition

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

$$R^{(I)}(D) = \min_{p(\hat{x}|x):\mathrm{E}[d(X;\hat{X})] \leq D} I(X; \hat{X}),$$

where the minimization is over all conditional distributions $p(\hat{x}|x)$ for which the joint distribution $p(x, \hat{x}) = p(x)p(\hat{x}|x)$ satisfies the expected distortion constraint.

Theorem

The rate distortion function $R(D)$ is a non-increasing convex function of D.

Rate-distortion Function Examples

Example

The rate distortion function for a Bernoulli(p) source with Hamming distortion is

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

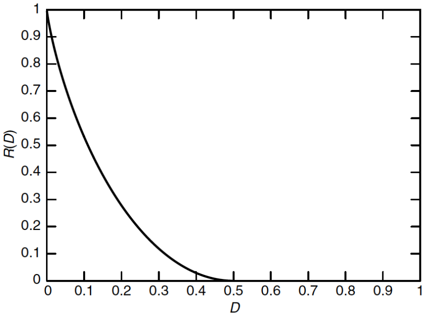


Figure: Rate distortion function for a Bernoulli(0.5) source.

Rate-distortion Function Examples (cont'd)

Example

The rate distortion function for a Gaussian $\mathcal{N}(0, \sigma^2)$ source with square-error distortion is

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

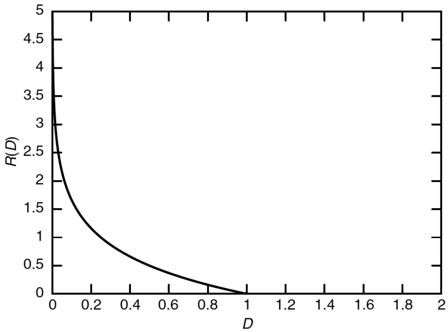


Figure: Rate distortion function for a Gaussian source.

Rate Distortion Coding Theorem

Theorem (Rate Distortion Coding Theorem)

The rate distortion function for an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x, \hat{x})$ is equal to the associated information rate distortion function. That is, $R(D) = R^{(I)}(D)$.

- Achievability: $R(D) > R^{(I)}(D) \implies$ the pair (R, D) is achievable. This can be proved by distortion ϵ-typicality.
- Converse: the pair (R, D) is achievable $\implies R(D) \geq R^{(I)}(D)$. This can be proved by DPI and convexity of $R(D)$.
- See detailed proof in Sections 10.5 and 10.6 of Cover's book.

Thank You!

Email: `<zhangy@ucsc.edu>`

Homepage: `https://people.ucsc.edu/~yzhan419/`