Ch. 8 Math Preliminaries for Lossy Coding

8.5 Rate-Distortion Theory

Introduction

- Theory provide insight into the trade between Rate & Distortion
- This theory is needed to answer:
  - What do typical R-D curves look like?
  - What factors impact the R-D trade-off?
  - For a given practical case, what is the best R-D curve that I can expect?
    - Tells designers when to stop trying to improve!
  - Etc.
- Our Goal Here: To express the R-D function in terms of Info Theory & see what it tells

Expression for Distortion

Need math form for Distortion… Recall:

$$D = E\{d(X,Y)\}$$

Expected Value taken w.r.t. joint PDF of X &Y

$$\Rightarrow \quad D = \int_{-\infty}^{\infty} \sum_{j=0}^{M-1} d(x, y_j) \underbrace{f_X(x) P(y_j \mid x)}_{} dx$$

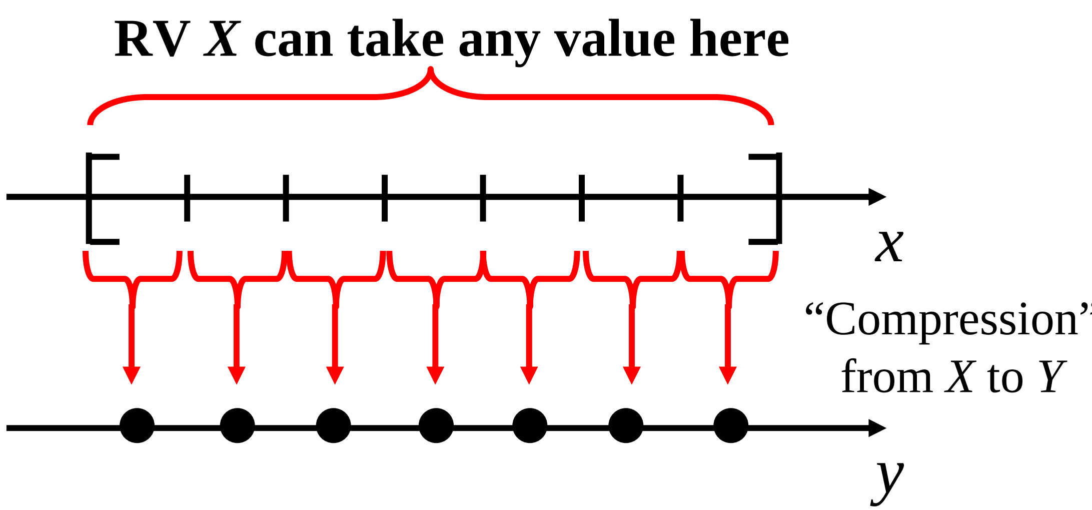
Describes joint prob of cont. *X* & disc. *Y*

<Differs from book>

Simple Example for Distortion

In compression we wish to minimize D by designing a mapping from values of X (continuous) into values of Y (discrete)

Simple Example:



Clearly our map doesn't impact $f_X(x)$…But it does specify $P(y_j|x)$!!

In fact, designing a compression algorithm is equivalent to specifying $P(y_j|x)$…. Including determining how ***many*** y_j values to use

Goal for Picking Compression Algorithm

Thus, D is a function of $P(y_j|x)$: $D = E\{d(X,Y)\} = D\left(P(y_j \mid x)\right)$

When we design a compression algorithm we want one that minimizes the distortion D

➔ Pick $P(y_j|x)$ to minimize $D(P(y_j|x))$

But wait!!!...

We need to worry about how large or small the Rate is!

➔ Constrained minimization of $D(P(y_j|x))$ to find the theoretical lower bound on the R-D curve

Info Theory View of Rate-Distortion

Recall that we said the Avg. Mutual Info $I(X;Y)$ was the theoretical minimum rate needed to convey the amount of info about X that is in some specified Y…

Info-Theory R-D Goal #1: For given rate value R find the $P(y_j|x)$ that minimizes the avg distortion $E\{d(X,Y)\}$ under the rate constraint $I(X;Y) \leq R$

Info-Theory R-D Goal #2: For given distortion value D find the $P(y_j|x)$ that minimizes the avg rate $I(X;Y)$ under the distortion constraint $E\{d(X,Y)\} \leq D$

These two complementary goals actually result in the same Info Theory R-D function…

We'll focus on #2:

$$R(D) = \min_{P(y_j|x):\, E\{d(X,Y)\} \leq D} I(X;Y)$$

Notice that the minimization is over $P(y_j|x)$ and involves $E\{d(X,Y)\}$ and $I(X;Y)$… We need these two things as functions of $P(y_j|x)$…

We've already seen that for the first one:

$$E\{d(X,Y)\} = D\big(P(y_j \mid x)\big)$$

The second one is a bit harder to see: $I(X;Y) = h(X) - h(X \mid Y)$

$$\text{with } h(X \mid Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) \log_2\left[f_{X|Y}(x \mid y)\right] dxdy$$

Depends on the reverse of the conditioning we need!!

But… using Bayes' Rule we see

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{\int f_{Y|X}(y \mid x) f_X(x) dy}$$

Has the order of the conditioning we need… but is a PDF

Recall: A Prob. Function can be written as a PDF that has only deltas in it

Example: R-D Function for Gaussian Source

Approach:
- Find lower bound for $I(X;Y)$ given the desired distortion level D
- Show this bound "can be achieved"

Let X be Gaussian w/ Zero Mean & Variance σ^2

Distortion Constraint: E{(X-Y)2} ≤ D

First consider case where $D < \sigma^2$

Don't Use formal Math to get this – reason it out by definition

$$I(X;Y) = h(X) - h(X|Y)$$

$$= h(X) - \underbrace{h(X-Y|Y)}_{\le h(X-Y)}$$

$$\ge \underbrace{h(X) - h(X-Y)} \qquad (\bigstar)$$

Now want to minimize this lower bound while meeting the distortion constraint

➔ That means we need to maximize $h(X - Y)$

For notational ease let $h(Z) \triangleq h(X-Y)$

We know that the Diff Entropy $h(Z)$ is maximized if Z is Gaussian… so <u>assume</u> that.

Further… if Z has variance of D, then $E\{(X-Y)^2\} = D$

… and we meet the distortion goal!

Since Z is Gaussian we know that $h(Z) = \frac{1}{2}\log_2\left[2\pi e D\right]$

Since X is Gaussian we know that $h(X) = \frac{1}{2}\log_2\left[2\pi e \sigma^2\right]$

Then (★) gives $I(X;Y) \geq \frac{1}{2}\log_2\left[2\pi e \sigma^2\right] - \frac{1}{2}\log_2\left[2\pi e D\right]$

⇒ $$I(X;Y) \geq \frac{1}{2}\log_2\left[\frac{\sigma^2}{D}\right]$$ (For $D < \sigma^2$)

Now consider case where $D \geq \sigma^2$

Note that: $E\{(X-0)^2\} = \sigma^2 \leq D$

➔ Setting $Y = 0$ satisfies the Dist. Goal

➔ Don't have to even send anything and you still meet the Dist. Goal!!!

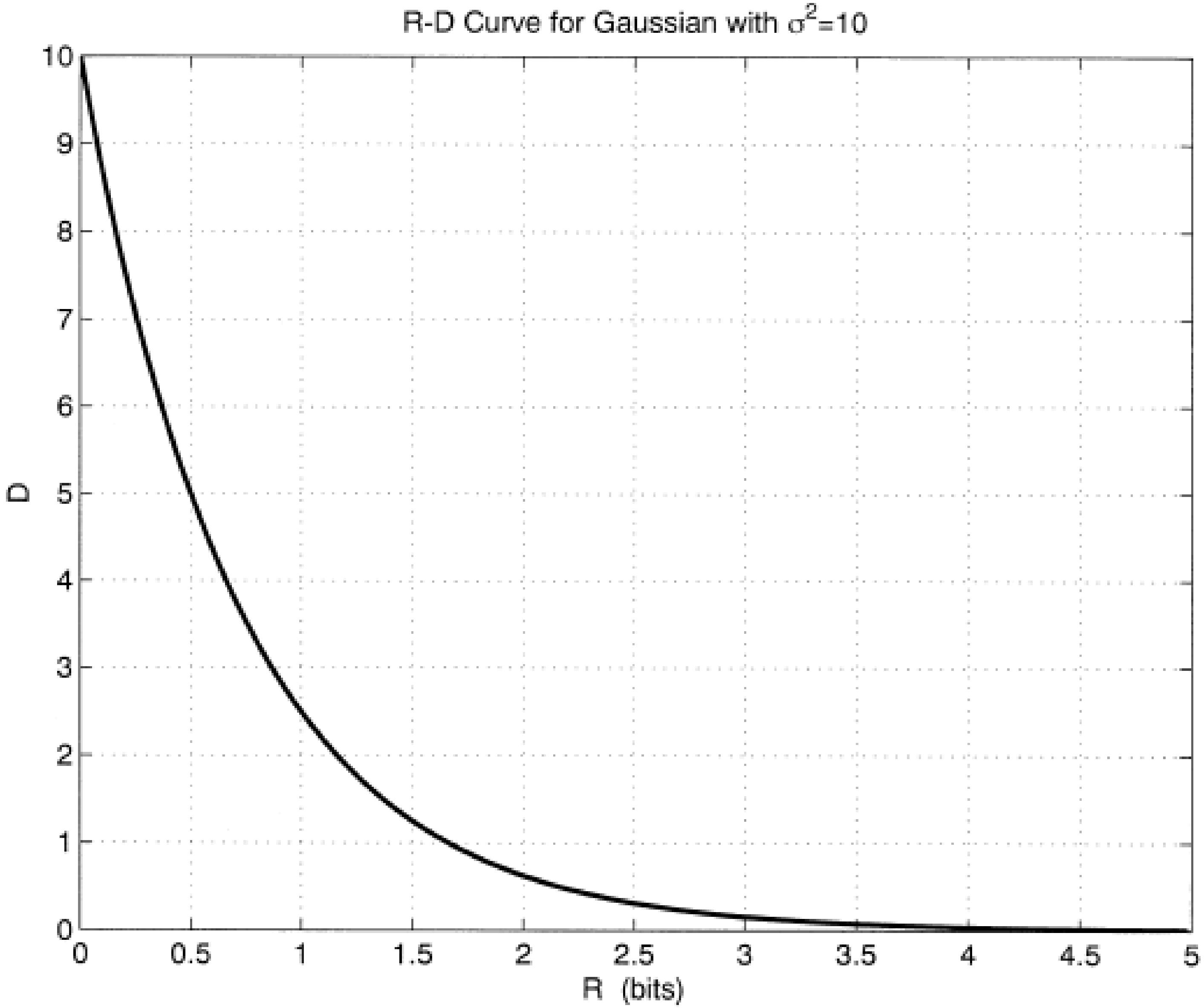
➔ $I(X;Y) = I(X;0) = 0$

Combining these two cases (and noting that it is possible to actually achieve this bound – see the textbook) we have

$$R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D}, & \text{for } D < \sigma^2 \\ 0, & \text{for } D \geq \sigma^2 \end{cases}$$

➔

$$D(R) = \sigma^2 2^{-2R}, \quad R \geq 0$$


R-D Curve for Gaussian with σ^2=10
D
R (bits)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
10
9
8
7
6
5
4
3
2
1
0

<u>In Practice</u>

The above results guide practical ideas and give bounds for comparing actual Rates & Distortions

These ideas also motivate “Operational Rate-Distortion”

In the above <u>Info Theory View</u>:

D was a probabilistic average over the ensemble

R was a lower limit

In <u>Operational R-D View</u>:

D is “what we actually achieve on this particular signal”

R is “the best rate we can get using a specified algorithm

< More On This Later>