Multimedia Communications

Mathematical Preliminaries for Lossy compression
Lossy compression

- In a lossless compression, the reconstructed signal is identical to the original sequence.
- Only a limited amount of compression can be obtained with lossless compression.
- There is a floor (entropy of the source) below which we cannot drive the size of the compressed sequence.
- In some applications, consequences of loss of information may prohibit us from loss of information (bank records, medical images).
Lossy compression

- If the resources are limited and we do not require absolute integrity, we can improve the amount of compression by accepting certain degree of loss during compression.
- Lossless compression: rate is the only performance measure.
- In lossy compression rate by itself is not enough.
- A measure of difference between original and reconstructed data.
- Our goal: incur minimum amount of distortion while compressing to the lowest possible rate.
- There is a tradeoff between minimizing rate and keeping distortion small.
Lossy compression

- X can be modeled as a random variable
- Y is also a random variable
Lossy compression

- Exp: suppose source output consists of \{0, 1, 2, ..., 15\}. Source encoder quantizes the output into one of \{0, 1, 2, ..., 7\}. Source decoder inverse-quantizes the output of source encoder to \{0, 2, ..., 14\}
- \(X = \{0, 1, 2, ..., 15\}\): 16 elements, \(Y = \{0, 2, 4, ..., 14\}\): 8 elements
- \(Y\) is a random variable:
  - \(P(Y=0) = P(Y=0|X=0)P(X=0) + P(Y=0|X=1)P(X=1)\)
- Source encoder-decoder can be described by \(P(Y|X)\)

\[
Q = \begin{bmatrix}
    p(y_1 | x_1) & p(y_1 | x_2) & \cdots & p(y_1 | x_N) \\
    p(y_2 | x_1) & p(y_2 | x_2) \\
    \vdots & \vdots & \ddots & \vdots \\
    p(y_M | x_1) & p(y_M | x_2) & \cdots & p(y_M | x_N)
\end{bmatrix}
\]

\[
\sum_{i=1}^{M} p(y_i | x_j) = 1
\]
Joint Entropy

• Let X and Y be two information sources with alphabets $A_N$ and $B_M$ and joint probability distribution $P : A_N \times B_M \rightarrow [0,1]$.

• The self-information of the event $(X=x, Y=y)$ is $i(x,y) = -\log P(x,y)$. The joint entropy of the two sources is

$$H(X, Y) = E[i(x, y)] = - \sum_{x \in A_N, y \in B_M} P(x, y) \log P(x, y)$$

• $H(X, Y)$ is the average amount of information carried by a pair of values of the two information sources.

• $H(X, Y) \leq H(X) + H(Y)$ and the equality holds when the two sources are independent.
Conditional Entropy

- The conditional self-information of the event \(X=x|Y=y\) is \(i(x|y)=-\log P(x|y)\). The conditional entropy of the two sources is

\[
H(X|Y) = E[i(x|y)] = -\sum_{x,y} P(x, y) \log P(x|y) = -\sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(y)}
\]

- \(H(X|Y)\) is the average amount of information carried by \(X\) when the value of \(Y\) is known.
- \(H(X|Y) + H(Y) = H(Y|X) + H(X) = H(X, Y)\)
- \(H(X|Y) \leq H(X)\), equality holds for independent sources.
Mutual Information

- Probability of source symbols $p(x)$: a priori probability
- After reconstructing $y$ the probability of input symbol being $x$ becomes $p(x|y)$: a posteriori probability
- Difference between information (uncertainty) before and after reconstruction of $y$ measures the gain in information due to $y$:
  $$i(x;y) = i(x) - i(x|y) = -\log(p(x)) + \log(p(x|y)) = \log(p(x|y)/p(x)) = \log(p(x,y)/p(x)p(y))$$
- $i(x,y)$ is called mutual information of the $x$ and $y$.
- Average mutual information of the two sources is the information that $Y$ carries about $X$:
  $$I(X;Y) = E[i(x; y)] = - \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)}$$
Mutual Information

- \( I(X;Y) = I(Y;X) \)
- \( I(X;Y) \geq 0 \), equality holds when the sources are independent.
- \( I(X;Y) \leq H(X) \) and \( I(X;Y) \leq H(Y) \)
- \( I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) \)
Distortion

• Distortion criteria:

1. Subjective criteria: measured by the effect the distortion has on the receiver. Not possible to model analytically.

2. Objective mathematical measures: use mathematical formula to measure distortion. Analytical or numerical optimizations are possible.

3. Objective measures in conjunction with models of the receiver (human auditory system, human visual system).
   - The process of human perception is very difficult to model and the models obtained are very complex
Mathematical Distortion Measures

• Average distortion

\[ d_1(x, y) = \frac{1}{N} \sum_{i=1}^{N} | x_i - y_i | \]

• \( l = 1 \rightarrow \) mean absolute error (MAE) \( d_l(x, y) = | x - y | \)

• \( l = 2 \rightarrow \) mean square error (MSE)

• \( l = \infty \rightarrow \) maximum error \( d_\infty(x, y) = \max_i | x_i - y_i | \)

• Sometimes the MSE is measured in dB:

\[ \text{SNR (dB)} = 10 \log_{10} \frac{\sigma_x^2}{\sigma_d^2} \quad \text{PSNR (dB)} = 10 \log_{10} \frac{x_{\max}^2}{\sigma_d^2} \]

• In a probabilistic modeling framework:

\[ D(X, Y) = \sum_{x, y} P(x, y) d(x, y) = \sum_{x, y} P(x) P(y \mid x) d(x, y) \]
Rate-Distortion Function: Definition

- Main trade-off: number of bits vs. signal distortion
- $R(D)$ specifies the lowest rate $R$ at which the output of a source can be encoded while keeping the distortion $\leq D$.

- continuous
- monotonic decreasing
- convex
- $R(0) = H(X)$
- $R(D) = 0$ for $D \geq D_{\text{max}}$
Rate-Distortion Function: Definition

- A coding scheme can be viewed as

\[ X, A_N \xrightarrow{P(x)} P(y|x) \xrightarrow{Y, B_M} P(y) \]

where \( Y \) is the reconstructed signal, \( A_N \) and \( P(x) \) describe the source and the conditional pdf \( P(y|x) \) and \( B_M \) describe the coding scheme.

- We can express the rate-distortion function of a source in terms of these functions.
Rate-Distortion Function: Definition

• The average distortion of a coding scheme is:

\[ D(X, Y) = \sum_{x, y} P(x, y)d(x, y) = \sum_{x, y} P(x)P(y \mid x)d(x, y) \]

• To find \( R(D) \), we choose from the set of coding schemes that have a distortion smaller than \( D \):

\[ \Gamma = \left\{ P(y \mid x); \ D(X, Y) \leq D \right\} \]

the one that achieves the smallest rate \( R \).

• \( R(D) \) is defined as

\[
R(D) = \min_{p(y \mid x) \in \Gamma} I(X, Y) = - \sum_{x, y} P(x, y) \log \frac{P(x \mid y)}{P(x)}
\]
Rate-Distortion Function: Computation

- How do we compute R(D)?
  - Analytically: find lower bound, show that it can be achieved.
  - Method of Lagrangian multipliers.
  - Computational approach: convex programming.

- In general, a difficult task.

- Can be found analytically only in a few simple cases.
Rate-Distortion Function: Example

- Consider a binary source with $P(0) = p, \ P(1) = 1-p, \ p<1/2$.
- Let the distortion function be $d(x,y) = x \oplus y$ (xor)
- Find the rate distortion function for this source.
- We follow two approaches: Lagrangian multipliers and an analytical method
Rate-Distortion Function: Example

- **Source encoder-decoder:**

\[
Q = \begin{bmatrix}
p(y_0 | x_0) & p(y_0 | x_1) \\
p(y_1 | x_0) & p(y_1 | x_1)
\end{bmatrix} = \begin{bmatrix}
a & 1-b \\
1-a & b
\end{bmatrix}
\]

\[
I(X;Y) = -\sum_{x,y} P(y | x)P(x) \log \frac{P(y | x)P(x)}{P(x)P(y)}
\]

\[
D(X,Y) = \sum_{x,y} P(x)P(y | x)d(x,y)
\]

\[
R(D) = \min I(x, y)
\]

\[D(X,Y) < D\]

\[
J(a,b,\lambda) = I(x,y) - \lambda(D(X,Y) - D)
\]

\[
R(D) = -p \log p - (1-p) \log(1-p) + D \log D + (1-D) \log(1-D)
\]
Rate-Distortion Function: Example

- $P(x=0) = p$, $P(x=1) = 1-p$, $p<1/2$.
- $d(x,y) = x \oplus y$ (xor)
- $R(0) = H(X) = -p \log p - (1-p) \log (1-p)$
- $H_b(p)$ is defined as $H_b(p) = -p \log p - (1-p) \log (1-p)$
- $H_b(p) = H_b(1-p)$
- $D(X,Y) = 1 \cdot P(x=0,y = 1) + 1 \cdot P(x=1,y = 0) = P(x \oplus y = 1)$
- $D_{\text{max}}$ is obtained when we send no bits, so we always decode $Y=1$ ($D = p$) or $Y=0$ ($D = 1-p$). The best of these two is $Y=1$, so $D_{\text{max}} = p$. 

Copyright S. Shirani
Rate-Distortion Function: Example

- Then \( \Gamma = \{ P(y \mid x); \ D(X, Y) \leq D \} = \{ P(y \mid x); \ P(x \oplus y = 1) \leq D \} \)
- The mutual information is \( I(X, Y) = H(X) - H(X \mid Y) \)
- Since, when \( Y \) is known, we can obtain \( X \oplus Y \) from \( X \) and vice versa, \( H(X|Y) = H(X \oplus Y|Y) \) and then
  \[
  I(X, Y) = H(X) - H(X \oplus Y \mid Y).
  \]
- A lower bound on \( I(X,Y) \) is then
  \[
  I(X, Y) = H(X) - H(X \mid Y)
  = H(X) - H(X \oplus Y \mid Y)
  \geq H(X) - H(X \oplus Y)
  = H_b(p) - H_b(P(x \oplus y) = 1).
  \]
Rate-Distortion Function: Example

• But \( P(x \oplus y = 1) \leq D \), so \( I(X, Y) \geq H_b(p) - H_b(D) \).

• Let us now show that the lower bound is actually achievable:

Let \( P(X=0|Y=1) = P(X=1|Y=0) = D \). Then

- \( P(X \oplus Y = 1) = P(X=1,Y=0) + P(X=0,Y=1) \)
  \[ = P(Y=0)P(X=1|Y=0) + P(Y=1)P(X=0|Y=1) \]
  \[ = P(Y=0) \cdot D + P(Y=1) \cdot D = D \]

- \( P(X \oplus Y = 0) = 1-D \)

- \( I(X,Y) = H(X) - H(X \oplus Y) = H_b(p) - H_b(D), \) q.e.d.
Rate-Distortion Function: Example

• Putting it together,

\[ R(D) = \begin{cases} H_b(p) - H_b(D), & D < p \\ 0, & \text{otherwise} \end{cases} \]
Continuous Amplitude Sources

- Continuous-amplitude memoryless sources are modeled as continuous RVs described by their pdf.
- Continuous RVs have infinite absolute entropy. We define the differential entropy as
  \[ h(X) = E[- \log p_X(x)] = - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx \]
- The entropy of a Gaussian random variable \( h(X) = 1/2 \log(2\pi e \sigma^2) \)
- For any source with variance \( \sigma^2 \), \( h(X) \leq h_G \).
Continuous Amplitude Sources

• The rate-distortion function of a continuous source is defined similarly with the discrete source:

\[ I(X, Y) = -\int \int p(x, y) \log \frac{p(x | y)}{p(x)} \, dx \, dy \]

\[ d(X, Y) = \int \int p(x)p(y | x)d(x, y) \, dx \, dy \]

\[ R(D) = \inf_{p(y|x) \in \Gamma} I(X, Y) \]

\[ \Gamma = \{p(y | x); \ d(X, Y) \leq D\} \]
Rate distortion

• The rate-distortion function for a zero mean Gaussian with distortion function \( d(x,y) = (x-y)^2 \) is

\[
R(D) = \begin{cases} 
\frac{1}{2} \log \frac{\sigma^2}{D}, & D < \sigma^2 \\
0, & \text{otherwise}
\end{cases}
\]

• Rate-distortion function for the Gaussian source is larger than the rate distortion function for any other source with a continuous distribution and the same variance

• The following are the lower bounds for a random variable \( X \):
  • squared error 
    \[
    R(D) = h(X) - \frac{1}{2} \log(2eD)
    \]
  • Absolute value error 
    \[
    R(D) = h(X) - \log(2eD)
    \]
Noisy Source Coding Theorem

• Let $R(D)$ be the rate-distortion function of a stationary source. Then, for any $D>0$ and $\varepsilon>0$:
  
  – **POSITIVE THEOREM:** a source code of block length $n$ (sufficiently large) exists that encodes source vectors $\mathbf{X} = \{X_1, \ldots, X_n\}$ at a rate $R < R(D) + \varepsilon$ with $d(X,Y) \leq D$.
  
  – **NEGATIVE THEOREM:** a source code does not exist that encodes the source at a rate $R < R(D)$ with $d(X,Y) \leq D$. 
Models for Continuous Sources

- **Uniform distribution:**
  
  \[ p_X(x) = \begin{cases} 
  \frac{1}{b-a}, & x \in [a, b] \\
  0, & \text{otherwise} 
  \end{cases} \]
  
  - The entropy \( h(X) = \log(b-a) \).

- **Gaussian distribution:**

  \[ p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \]
  
  - The entropy \( h(X) = \frac{1}{2} \log(2\pi e \sigma^2) = h_G \)
  
  - For any source with variance \( \sigma^2 \), \( h(X) \leq h_G \).
Models for Continuous Sources

- **Exponential distribution:**
  - single-sided distribution
  \[ p_X(x) = \lambda \exp[-\lambda x] \quad x \geq 0 \]

- **Laplacian distribution:**
  - good for modeling peaked distributions
  \[ p_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left[-\frac{\sqrt{2}|x|}{\sigma}\right] \]

- **Gamma distribution:**
  - even more peaked
  \[ p_X(x) = \frac{3^{1/4}}{\sqrt{8\pi\sigma}} \frac{1}{|x|} \exp\left[-\frac{\sqrt{3}|x|}{2\sigma}\right] \]
Models for Continuous Sources

- **Generalized Gaussian distribution:**
  
  \[ p_X(x) = a \exp[-b(x - \mu)^c] \]
  
  \[ a = \frac{bc}{2\Gamma(1/c)} \]
  
  \[ b = \frac{1}{\sigma \sqrt[2]{\Gamma(3/c)}} \]

- \( c = 2 \) Gaussian
- \( c = 1 \) Laplacian
- \( c = 0.7 \) good model for subbands
Models for Sources with Memory

- Linear system models:

\[ x_n = \sum_{i=1}^{N} a_i x_{n-i} + \sum_{j=1}^{M} b_j \epsilon_{n-j} + \epsilon_n, \]

where \( \epsilon_n \) is a Gaussian white noise sequence with variance \( \sigma_\epsilon^2 \).

- LTI system with N poles and M zeros.
- Auto-regressive moving average (ARMA) model.
- For \( M = 0 \) we obtain the “all-pole” or “AR” or “Markov” model. Good model for speech production.
Models for Sources with Memory

• AR(N) is an N-th order Markov model
  \[ x_n = \sum_{i=1}^{N} a_i x_{n-i} + \varepsilon_n \]

• For the first-order Markov source with an autocorrelation function of \( \phi_k = |r|^k \) the rate-distortion function is
  \[
  R(D) = \begin{cases} 
  \frac{1}{2} \log \frac{1 - r^2}{D}, & D \leq \frac{1 - r}{1 + r} \\
  \text{(complex)}, & \text{otherwise}
  \end{cases}
  \]