Rate Distortion Theory

- Lossy coding: decoded signal is an approximation of original
- Rate distortion theory is a theoretical discipline treating data compression from the viewpoint of information theory
- The results of rate distortion theory are obtained without consideration of a specific coding method
- The goal of rate distortion theory is to calculate the minimum transmission bit rate for a given distortion and source
- Below: example for a rate distortion function for a discrete iid source

\[ (H(S), D_{min} = 0) \rightarrow R \]

\[ (R = 0, D = D_{max}) \]

Rate Distortion Function
Outline

- Transmission System and Variables
- Rate Distortion Function for Discrete Random Variables
  - Definition: Source and Decoded Symbols, Coder/Decoder, Distortion, Rate
  - Rate Distortion Function and Properties
  - Rate Distortion Function for Binary IID Sources
- Rate Distortion Function for Amplitude-Continuous Random Variables
  - Definition: Source and Decoded Symbols, Coder/Decoder, Distortion, Rate
  - Rate Distortion Function
  - Shannon Lower Bound
  - Rate Distortion Function for Memoryless Gaussian Sources
  - Rate Distortion Function for Gaussian Sources with Memory
Transmission System and Variables

- Transmission system

![Diagram of transmission system]

- Derivation in two steps
  1. Define $S$, $S'$, coder/decoder, distortion $D$, and rate $R$
  2. Establish a functional relationship between $S$, $S'$, $D$, and $R$

- For two types of random variables
  1. Derivation for discrete random variables
  2. Derivation for amplitude-continuous random variables (Gaussian, Laplacian, etc.)
Operational Rate Distortion Function

- **Encoder:**
  - Irreversible encoder mapping \( \alpha : s \rightarrow i \)
  - Lossless mapping \( \gamma : i \rightarrow b \)

- **Decoder:**
  - Lossless mapping \( \gamma^{-1} : b \rightarrow i \)
  - Decoder mapping \( \beta : i \rightarrow s' \)
Source code

- **Source code**: $Q = (\alpha, \beta, \gamma)$
- $N$-dimensional block source code: $Q_N = \{\alpha_N, \beta_N, \gamma_N\}$
  - Blocks of $N$ consecutive input samples are independently coded
  - Each block of input samples $s^{(N)} = \{s_0, \cdots, s_{N-1}\}$ is mapped to a vector of $K$ quantization indexes
    \[
    \mathbf{i}^{(K)} = \alpha_N(s^{(N)})
    \] (1)
  - Resulting vector of indexes $\mathbf{i}$ is converted into bit sequence
    \[
    \mathbf{b}^{(\ell)} = \gamma_N(\mathbf{i}^{(K)}) = \gamma_N(\alpha_N(s^{(N)}))
    \] (2)
  - At decoder side, index vector is recovered
    \[
    \mathbf{i}^{(K)} = \gamma_N^{-1}(\mathbf{b}^{(\ell)}) = \gamma_N^{-1}(\gamma_N(\mathbf{i}^{(K)}))
    \] (3)
  - Index vector mapped to a block of reconstructed samples
    $s'^{(N)} = \{s'_0, \cdots, s'_{N-1}\}$
    \[
    s'^{(N)} = \beta_N(\mathbf{i}^{(K)}) = \beta_N(\alpha_N(s^{(N)}))
    \] (4)
Distortion 1

- This lecture: *additive distortion measures*

\[ d_1(s, s') \geq 0 \quad \text{with equality, if and only if} \quad s = s' \quad (5) \]

- Examples:

  Hamming distance: \[ d_1(s, s') = \begin{cases} 1, & \text{for } s \neq s' \\ 0, & \text{for } s = s' \end{cases} \quad (6) \]

  Mean squared error: \[ d_1(s, s') = (s - s')^2 \quad (7) \]

- Distortion for \( s^{(N)} = \{s_0, s_1, \ldots, s_{N-1}\} \) and \( s'^{(N)} = \{s'_0, s'_1, \ldots, s'_{N-1}\} \)

\[ d_N(s^{(N)}, s'^{(N)}) = \frac{1}{N} \sum_{i=0}^{N-1} d_1(s_i, s'_i) \quad (8) \]
Distortion II

- Stationary random process $S = \{S_n\}$ and $N$-dimensional block source code $Q_N = \{\alpha_N, \beta_N, \gamma_N\}$

\[
\delta(Q_N) = E \left\{ d_N(S^{(N)}, \beta_N(\alpha_N(S^{(N)}))) \right\} \\
= \int_{\mathcal{R}^N} f(s) d_N(s, \beta_N(\alpha_N(s))) \, ds
\]  

(9) \hspace{1cm} (10)

- Arbitrary random process $S = \{S_n\}$ and arbitrary code $Q$

\[
\delta(Q) = \lim_{N \to \infty} E \left\{ d_N(S^{(N)}, \beta_N(\alpha_N(S^{(N)}))) \right\}
\]  

(11)
Rate

- Average number of bits per input symbol (\(| \cdot |\) denotes the number of bits)

\[
r_N(s^{(N)}) = \frac{1}{N} |\gamma_N(\alpha_N(s^{(N)}))| \quad \text{with} \quad b^{(\ell)} = \gamma_N(\alpha_N(s^{(N)})) \quad (12)
\]

- Stationary random process \( S = \{S_n\} \) and \( N \)-dimensional block source code \( Q_N = \{\alpha_N, \beta_N, \gamma_N\} \)

\[
r(Q_N) = \frac{1}{N} E \left\{ |\gamma_N(\alpha_N(S^{(N)}))| \right\} \\
= \frac{1}{N} \int_{\mathcal{R}^N} f(s) |\gamma_N(\alpha_N(s))| \, ds \quad (13)
\]

- Arbitrary random process \( S = \{S_n\} \) and arbitrary code \( Q \)

\[
r(Q) = \lim_{N \to \infty} \frac{1}{N} E \left\{ |\gamma_N(\alpha_N(S^{(N)}))| \right\} \quad (15)
\]
Operational Rate Distortion Function

- For given source $S$: each code $Q$ is associated with a rate distortion point $(R, D)$
- Rate distortion point is achievable, if there exist a code $Q$ such that $r(Q) \leq R$ and $\delta(Q) \leq D$
- Operational rate distortion function $R(D)$ and its inverse, the operational distortion rate function $D(R)$

$$
R(D) = \inf_{Q: \delta(Q) \leq D} r(Q) \\
D(R) = \inf_{Q: r(Q) \leq R} \delta(Q)
$$

(16)

Diagram: Operational rate distortion function $R(D)$ with region of achievable rate distortion points $(R, D)$.
Operational rate distortion function specifies a fundamental performance bound for lossy source coding.

Difficulty to evaluate

\[ R(D) = \inf_{Q: \delta(Q) \leq D} r(Q) \]  

(17)

Information rate distortion function: introduced by Shannon in [Shannon, 1948, Shannon, 1959]

Obtain expression of rate distortion bound that involves the distribution of the source using mutual information.

Show that information rate distortion function is achievable.
Example: Discrete Binary Source

- Consider a discrete binary iid source with alphabet $A = \{0, 1\}$
  \[ p(0) = p(1) = 0.5 \]  
  (18)
- Assume distortion measure is Hamming distance
- Coding scheme: 1 out of $M$ symbols is not transmitted but guessed at receiver side
- Rate for this experimental codec is
  \[ R_E = \frac{M - 1}{M} = 1 - \frac{1}{M} \text{ bits/symbol} \]  
  (19)
- Distortion is given as
  \[ D = \frac{1}{M} \left[ \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 \right] = \frac{1}{2M} \]  
  (20)
- Operational rate distortion function for the above code:
  \[ R_E(D) = 1 - 2D \]  
  (21)
Can we do better than the example codec?

- $R_E(D)$ is the rate distortion performance for the experimental coding scheme
- Rate distortion function $R(D)$ for the same source

![Rate Distortion Function Graph]

$R_E(D)$
Mutual Information for Discrete IID Sources

- Entropy $H(S)$ is a measure of uncertainty about random variable $S$.
- Conditional entropy $H(S|S')$ is a measure on uncertainty about random variable $S$ after observing random variable $S'$.
- Mutual information
  
  $$I(S; S') = H(S) - H(S|S')$$  

  is a measure of the reduction of uncertainty about $S$ due to the observation of $S'$.
- Average amount of information that $S'$ carries about $S$.
- Mutual information for discrete random variables $S \in \mathcal{A}$ and $S' \in \mathcal{B}$
  
  $$I(S; S') = \sum_{s \in \mathcal{A}} \sum_{s' \in \mathcal{B}} p(s, s') \log_2 \frac{p(s|s')}{p(s)}$$
Mutual information rewritten using Bayes’ rule

\[
I(S; S') = \sum_{s \in \mathcal{A}} \sum_{s' \in \mathcal{B}} p(s, s') \log_2 \frac{p(s | s')}{p(s)}
\]

\[
= \sum_{s \in \mathcal{A}} \sum_{s' \in \mathcal{B}} p(s, s') \log_2 \frac{p(s, s')}{p(s)p(s')}
\]

\[
= \sum_{s \in \mathcal{A}} \sum_{s' \in \mathcal{B}} p(s, s') \log_2 \frac{p(s' | s)}{p(s')}
\]

\[
= H(S') - H(S' | S)
\] (24)

Since \( H(S' | S) \geq 0 \)

\[
H(S') \geq I(S; S')
\] (25)
Mutual Information for Discrete Sources and Rate

- Consider $N$-dimensional random vectors $\mathbf{S} = (S_0, S_1, \ldots, S_{N-1})^T \in \mathcal{A}^N$ and $\mathbf{S}' = (S'_0, S'_1, \ldots, S'_{N-1})^T \in \mathcal{B}^N$

  \[
  I_N(\mathbf{S}; \mathbf{S}') = H_N(\mathbf{S}') - H_N(\mathbf{S}' | \mathbf{S}) \geq 0
  \]

  \[
  \leq H_N(\mathbf{S}')
  \]  

  (26)

- Recall: entropy rate

  \[
  \bar{H}(\mathbf{S}') = \lim_{N \to \infty} \frac{H_N(\mathbf{S}')}{N}
  \]  

  (27)

- Rate vs. mutual information

  \[
  r(Q) \geq \lim_{N \to \infty} \frac{H_N(\mathbf{S}')}{N} \geq \lim_{N \to \infty} \frac{I_N(\mathbf{S}; \mathbf{S}')}{N}
  \]  

  (28)
Description of a Codec for Discrete Sources Using Conditional PMFs

- Statistical properties of a mapping \( s' = \beta(\alpha(s)) \) can be described by an \( N \)-th order conditional pmf \( g_N(s'|s) \).
- For \( N > 1 \), \( g_N(s'|s) \) are multivariate conditional pmfs.
- Pmfs \( g_N(s'|s) \) obtained by a deterministic mapping (codes) are a subset of the set of all conditional pmfs.

Example 1: Mapping \( s \rightarrow s' : s' = \lfloor s/\Delta \rfloor \cdot \Delta \)

\[
g_1(s'|s) = \begin{cases} 
1 & : s' = \lfloor s/\Delta \rfloor \cdot \Delta \\
0 & : \text{otherwise}
\end{cases}
\]
DESCRIPTION OF A CODEC USING CONDITIONAL PMFS II

Example 2: Mapping \( (s_n, s_{n+1}) \to (s'_n, s'_{n+1}) \)

\[
(s'_n, s'_{n+1}) = \begin{cases} 
(1, 1) & : \ s_n + s_{n+1} > 1 \\
(-1, -1) & : \ s_n + s_{n+1} < -1 \\
(0, 0) & : \ \text{otherwise}
\end{cases}
\]

\[
g_1(s'|s) = \begin{cases} 
x(s) & : \ s' = -1 \\
y(s) & : \ s' = 0 \\
z(s) & : \ s' = 1 \\
0 & : \ \text{otherwise}
\end{cases}
\]

with \( x(s) + y(s) + z(s) = 1 \)
Let \( g_N(s'|s) \) be the \( N \)-th order conditional pmf of \( s' \in \mathcal{B}^N \) given \( s \in \mathcal{A}^N \)

### Distortion

\[
\delta_N(g_N) = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s, s') \cdot d_N(s, s')
\]

\[
\quad = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s) \cdot g_N(s'|s) \cdot d_N(s, s') \tag{29}
\]

### Mutual information

\[
I_N(g_N) = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s, s') \cdot \log_2 \frac{p(s, s')} {p(s) \cdot p(s')}
\]

\[
\quad = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s) \cdot g_N(s'|s) \cdot \log_2 \frac{g_N(s'|s)} {p(s')} \tag{30}
\]

### Information rate distortion function

\[
R^{(I)}(D) = \lim_{N \to \infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{I_N(g_N)} {N} \tag{31}
\]
The class of conditional pmfs $g_Q^N(s'|s)$ representing a codec $Q = (\alpha, \beta, \gamma)$ is a subset of all conditional pmfs $g_N(s'|s)$

$$I_N(g_Q^N) \geq \inf_{g_N: \delta_N(g_N) \leq D} I_N(g_N)$$  \hspace{1cm} (32)

For a given maximum average distortion $D$, the information rate distortion function $R^{(I)}(D)$ is the lower bound for the transmission bit-rate

$$\forall Q : \delta(Q) \leq D \quad r(Q) \geq R^{(I)}(D)$$  \hspace{1cm} (33)

→ *Fundamental source coding theorem*

It can be shown [Cover and Thomas, 2006, p. 318] that for any $\varepsilon > 0$, there exists a code $Q$ with $\delta(Q) \leq D$ and $r(Q) \geq R^{(I)}(D) + \varepsilon$

Rate distortion function $R(D)$ is used for both, operational rate distortion function and information rate distortion function
Distortion Rate Function for Discrete Sources

- Distortion
  \[
  \delta_N(g_N) = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s) \cdot g_N(s' | s) \cdot d_N(s, s')
  \]  
  (34)

- Mutual information
  \[
  I_N(g_N) = \sum_{s \in \mathcal{A}^N} \sum_{s' \in \mathcal{B}^N} p(s) \cdot g_N(s' | s) \cdot \log_2 \frac{g_N(s' | s)}{p(s')}
  \]  
  (35)

- Information distortion rate function is given via interchanging the roles of rate and distortion
  \[
  D(R) = \lim_{N \to \infty} \inf_{g_N : I_N(g_N)/N \leq R} \delta_N(g_N)
  \]  
  (36)

- For a given maximum average rate \( R \), the distortion rate function \( D(R) \) is the lower bound for the average distortion
  \[
  \forall Q : r(Q) \leq R \quad \delta(Q) \geq D(R).
  \]  
  (37)
Information rate distortion function was first derived by Shannon for iid sources [Shannon, 1948, Shannon, 1959]

For iid sources: joint pmf \( p(s) = \prod_{i=0}^{N-1} p(s_i) \)

\[
\delta_N(g_Q^N) = \delta_1(g^Q) \quad \text{and} \quad I_N(g_Q^N) = N \cdot I_1(g^Q) \quad (38)
\]

First order rate distortion function

\[
R_{1}^{(I)}(D) = \inf_{g: \delta_1(g) \leq D} I_1(g) \quad (39)
\]

First order distortion rate function

\[
D_1(R) = \inf_{g: I_1(g) \leq R} \delta_1(g) \quad (40)
\]
Properties of $R(D)$ for Discrete Sources and Additive Distortion Measures

- Example of $R(D)$ for a discrete iid source
- $R(D)$ is a non-increasing and convex function of $D$
- There exists a value $D_{\text{max}}$, so that

$$ \forall D \geq D_{\text{max}} \quad R(D) = 0 $$  \hspace{1cm} (41)

→ For MSE distortion measure: $D_{\text{max}}$ is equal to the variance $\sigma^2$ of the source
- Minimum rate required for lossless transmission of a discrete source is equal to the entropy rate

$$ D_{\text{min}} = 0 \quad R(0) = \bar{H}(S) $$  \hspace{1cm} (42)

→ Fundamental bound for lossless coding: special case of the fundamental bound for lossy coding
Example: Discrete Binary IID Source I

- Let the discrete binary iid source be given with probability \( p(0) = p \) and \( p(1) = 1 - p \), Hamming distance as distortion measure.
- Derive information rate distortion function \( R(D) \) by
  - Deriving a lower bound
  - Showing that the lower bound is achievable
- Maximum distortion is given by \( D(0) = \min(p, 1 - p) \), since we can assume that all symbols have the value with the higher probability.
- With \( \oplus \) denoting modulo 2 addition (XOR operator): derive a lower bound

\[
I(S; S') = H(S') - H(S|S') \\
= H_b(p) - H(S \oplus S'|S') \\
\geq H_b(p) - H_b(D) \tag{43}
\]

with \( H_b(x) = -x \log_2 x - (1 - x) \log_2 (1 - x) \) being the binary entropy function and \( D = S \oplus S' \) being the distortion (Hamming distance).
Example: Discrete Binary IID Source II

The derived lower bound can be achieved by the following choice

\[ p_{S'}(s') = \frac{1}{1 - 2D} \begin{cases} 
  p - D & : \ s' = 0 \\
  1 - p - D & : \ s' = 1 
\end{cases} \tag{44} \]

\[ g_{S|S'}(s|s') = \begin{cases} 
  1 - D & : \ s = s' \\
  D & : \ s \neq s 
\end{cases} \tag{45} \]

It is a valid choice, since

\[ p_S(s) = \sum_{s'=0}^{1} g_{S|S'}(s|s') \cdot p_{S'}(s') = \begin{cases} 
  p & : \ s = 0 \\
  1 - p & : \ s = 1 
\end{cases} \tag{46} \]

We obtain

\[ \delta_1(g) = \sum_{\forall s, s'} p_{S'}(s') \cdot g_{S|S'}(s|s') \cdot d_1(s, s') = D \tag{47} \]

\[ I_1(g) = H(S) - H(S|S') = H_b(p) - H_b(D) \tag{48} \]

Hence, the information rate distortion function \( R(D) \) is

\[ R(D) = \begin{cases} 
  H_b(p) - H_b(D) & 0 \leq D \leq \min(p, 1 - p) \\
  0 & D > \min(p, 1 - p) 
\end{cases} \tag{49} \]
Example: $R(D)$ for Different Discrete Binary IID Sources

- $R_E(D)$ is the rate distortion performance for the example coding scheme
- $R(D)$ for a binary source for several values of $p$
Operational Rate Distortion Function for Continuous Sources

- **Source symbols:** random process $S$ with joint pdf $f(s)$
- **Decoded symbols:** random process $S'$ with joint pmf $p(s')$

$\rightarrow$ For finite rate, decoding symbols must be discrete

- For stationary processes $S$ and block codes $Q_N = (\alpha_N, \beta_N, \gamma_N)$

$$\delta_N(Q_N) = \int_{\mathcal{R}^N} f(s) \, d_N(s, \beta_N(\alpha_N(s))) \, ds$$  \hspace{1cm} (50)

- Stationary processes $S$ and block codes $Q_N = (\alpha_N, \beta_N, \gamma_N)$

$$r_N(Q_N) = \frac{1}{N} \int_{\mathcal{R}^N} f(s) \, |\gamma_N(\alpha_N(s))| \, ds$$  \hspace{1cm} (51)

- Distortion and rate of the operational rate distortion function

$$\delta(Q) = \lim_{N \to \infty} \delta_N(Q_N) \quad \quad r(Q) = \lim_{N \to \infty} r_N(Q_N)$$  \hspace{1cm} (52)
Discretization of Continuous Random Variables

- Approximation $f_X^{(\Delta)}$ of pdf $f_X$

\[
\forall x : x_i \leq x < x_{i+1} \quad f_X^{(\Delta)}(x) = \frac{1}{\Delta} \int_{x_i}^{x_{i+1}} f_X(x') \, dx'
\]  
(53)

- Pmf $p_{X_\Delta}$ for random variable $X_\Delta$

\[
p_{X_\Delta}(x_i) = \int_{x_i}^{x_{i+1}} f_X(x') \, dx' = f_X^{(\Delta)}(x_i) \cdot \Delta
\]  
(54)

- Joint pmf of two discrete approximations $X_\Delta$ and $Y_\Delta$

\[
p_{X_\Delta Y_\Delta}(x_i, y_j) = f_{X Y}^{(\Delta)}(x_i, y_j) \cdot \Delta^2
\]  
(55)
Mutual Information for Continuous Random Variables

- Mutual information for discrete random variables $X_\Delta \in \mathcal{A}_{X_\Delta}$ and $Y_\Delta \in \mathcal{A}_{Y_\Delta}$

\[
I(X_\Delta; Y_\Delta) = \sum_{x_i \in \mathcal{A}_{X_\Delta}} \sum_{y_i \in \mathcal{A}_{Y_\Delta}} p_{X_\Delta Y_\Delta}(x_i, y_j) \log_2 \frac{p_{X_\Delta Y_\Delta}(x_i, y_j)}{p_{X_\Delta}(x_i) p_{Y_\Delta}(y_j)} \tag{56}
\]

\[
= \sum_{x_i \in \mathcal{A}_{X_\Delta}} \sum_{y_i \in \mathcal{A}_{Y_\Delta}} f_{XY}^{(\Delta)}(x_i, y_j) \cdot \log_2 \frac{f_{XY}^{(\Delta)}(x_i, y_j)}{f_X^{(\Delta)}(x_i) f_Y^{(\Delta)}(y_j)} \cdot \Delta^2
\]

- As $\Delta$ approaches zero

\[
I(X; Y) = \lim_{\Delta \to 0} I(X_\Delta; Y_\Delta) \tag{57}
\]

the piecewise constant pdf approximations $f_{XY}^{(\Delta)}$, $f_X^{(\Delta)}$, and $f_Y^{(\Delta)}$ approach pdfs $f_{XY}$, $f_X$, and $f_Y$

\[
I(X; Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \, dx \, dy \tag{58}
\]
Mutual Information for Random Vectors

- Given $S = (S_0, S_1, \ldots, S_{N-1})^T$ and $S' = (S'_0, S'_1, \ldots, S'_{N-1})^T$

$$I(X;Y) = \int_{R^N} \int_{R^N} f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} \, dx \, dy$$

$$= \int_{R^N} \int_{R^N} f_X(x)f_Y|X(x,y) \log_2 \frac{f_Y|X(x,y)}{f_Y(x)} \, dx \, dy \quad (59)$$

- Assume: $Y$ is a discrete random vector with alphabet $\mathcal{A}_Y^N$ and $\delta(\cdot)$ in these equations being the Dirac delta function

$$f_Y(y) = \sum_{a \in \mathcal{A}_Y^N} \delta(y - a) \, p_Y(a) \quad (60)$$

$$f_Y|X(y|x) = \sum_{a \in \mathcal{A}_Y^N} \delta(y - a) \, p_Y|X(a|x) \quad (61)$$

- Rewriting mutual information using above pmfs

$$I(X;Y) = \int_{R^N} f_X(x) \sum_{a \in \mathcal{A}_Y^N} p_Y|X(a|x) \log_2 \frac{p_Y|X(a|x)}{p_Y(a)} \, dx \quad (62)$$
Mutual Information for Random Vectors II

- Mutual information assuming $Y$ being a discrete process

$$I(X; Y) = \int_{\mathcal{R}^N} f_X(x) \sum_{a \in \mathcal{A}_Y^N} p_{Y|X}(a|x) \log_2 \frac{p_{Y|X}(a|x)}{p_Y(a)} \, dx$$  \hspace{1cm} (63)

- Can be re-written as

$$I(X; Y) = H(Y) - \int_{-\infty}^{\infty} f_X(x) H(Y|X=x) \, dx$$  \hspace{1cm} (64)

where $H(Y)$ is the entropy of the discrete random vector $Y$ and

$$H(Y|X=x) = - \sum_{a \in \mathcal{A}_Y^N} p_{Y|X}(a|x) \log_2 p_{Y|X}(a|x)$$  \hspace{1cm} (65)

is the conditional entropy of $Y$ given the event $\{X=x\}$

- Since the conditional entropy $H(Y|X=x)$ is always nonnegative, we have

$$I(X; Y) \leq H(Y)$$  \hspace{1cm} (66)
Description of a Codec for Continuous Processes Using Conditional PDFs

- Similar as for discrete sources, the statistical properties of a mapping \( s' = \beta(\alpha(s)) \) can be described using a conditional pdf \( g_N(s'|s) \).

- Example: Mapping \( s \to s' : s' = \lfloor s/\Delta \rfloor \cdot \Delta \).
  \[
g_1(s'|s) = \delta(s' - \lfloor s/\Delta \rfloor \cdot \Delta)
\]

- For \( N > 1 \), \( g_N(s'|s) \) are multivariate conditional pdfs.
- The pdfs \( g_N(s'|s) \) obtained by a deterministic mapping (codes) are a subset of the set of all conditional pdfs.
Let \( g_N(s'|s) \) be the \( N \)-th order conditional pdf of \( s' \in B^N \) given \( s \in \mathcal{R}^N \)

- **Distortion**

\[
\delta_N(g_N) = \int_{\mathcal{R}^N} \int_{\mathcal{R}^N} \frac{f_Ss'}{f_{s'}}(s, s') \cdot d_N(s, s') \, ds \, ds'
\]

\[
= \int_{\mathcal{R}^N} \int_{\mathcal{R}^N} f_S(s) \cdot g_N(s'|s) \cdot d_N(s, s') \, ds \, ds' \quad (67)
\]

- **Mutual information**

\[
I_N(g_N) = \int_{\mathcal{R}^N} \int_{\mathcal{R}^N} f_S(s) \cdot g_N(s'|s) \cdot \log_2 \frac{g_N(s'|s)}{f_{s'}(s')} \, ds \, ds' \quad (68)
\]

with

\[
f_{s'}(s') = \int_{\mathcal{R}^N} f_S(s) \cdot g_N(s'|s) \, ds. \quad (69)
\]

- **Information Rate Distortion Function**

\[
R^{(I)}(D) = \lim_{N \to \infty} \inf_{g_N: \delta_N(g_N) \leq D} \frac{I_N(g_N)}{N} \quad (70)
\]
Properties of $R^{(I)}(D)$ for Continuous Sources and Additive Distortion Measures

- Every code $Q$ that yields a distortion $\delta(Q)$ less than or equal to any given value $D$ for a source $S$ is associated with a rate $r(Q)$ that is greater than or equal to the information rate distortion function $R^{(I)}(D)$ for the source $S$

  \[ \forall Q : \delta(Q) \leq D \quad r(Q) \geq R^{(I)}(D) \quad (71) \]

- Example of $R(D)$ for an amplitude-continuous source

- Properties similar as for discrete sources, except $R(D)$ approaches infinity as $D$ approaches 0
Differential Entropy

- Mutual information of two continuous random vectors $X$ and $Y$
  \[
  I(X;Y) = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{XY}(x,y) \log_2 \frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} \, dx \, dy
  \]
  \[= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{XY}(x,y) \log_2 \frac{f_{X|Y}(x,y)}{f_X(x)} \, dx \, dy
  \]
  \[= -\int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{XY}(x,y) \log_2 f_{X}(x) \, dx \, dy
  \]
  \[+ \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{XY}(x,y) \log_2 f_{X|Y}(x,y) \, dx \, dy
  \]
  \[= h(X) - h(X|Y)
  \]

- Differential entropy of $X$
  \[
  h(X) = E \{- \log_2 f_X(X)\} = -\int_{\mathbb{R}^N} f_X(x) \log_2 f_X(x) \, dx
  \]

- Conditional differential entropy
  \[
  h(X|Y) = E \{- \log_2 f_{X|Y}(X|Y)\}
  \]
  \[= -\int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{XY}(x,y) \log_2 f_{X|Y}(x|y) \, dx \, dy.
  \]
Differential Entropy for an Uniform IID Source

- For an continuous iid source $S$, differential entropy is defined as
  \[
  h(S) = E \{ - \log_2 f(S) \} = - \int f(s) \log_2 f(s) \, ds \tag{75}
  \]

- $h(S)$ for uniform distribution $f(s) = 1/A$ for $-A/2 \leq s \leq A/2$
  \[
  h(S) = - \int_{-A/2}^{A/2} \frac{1}{A} \log_2 \frac{1}{A} \, ds = \log_2 \frac{A}{A} \int_{-A/2}^{A/2} \, ds = \log_2 A \tag{76}
  \]

- Differential entropy can become negative (in contrast to discrete entropy)

![Graphs showing differential entropy and log_2 A vs A]
Differential Entropy for an Gaussian IID Source

- Gaussian iid process

\[ f_S(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s^2}{2\sigma^2}} \]  

- Differential entropy

\[
h(S) = -\int f_S(s) \log_2 f_S(s) ds
\]

\[
= -\int f_S(s) \left[ -s^2 \frac{2\sigma^2}{2\sigma^2} \log_2 e - \log_2 \sqrt{2\pi\sigma^2} \right] ds
\]

\[
= \frac{E \{ S^2 \}}{\sigma^2} \frac{1}{2} \log_2 e + \frac{1}{2} \log_2 (2\pi\sigma^2)
\]

\[
= \frac{1}{2} \log_2 e + \frac{1}{2} \log_2 (2\pi\sigma^2)
\]

\[
= \frac{1}{2} \log_2 (2\pi e\sigma^2) \text{ bits}
\]
$N$-th Order Differential Entropy

- $N$-th order differential entropy
  \[
  h_N(S) = h(S^{(N)}) = h(S_0, \cdots, S_{N-1})
  \]

- Differential entropy rate
  \[
  \bar{h}(S) = \lim_{N \to \infty} \frac{h_N(S)}{N} = \lim_{N \to \infty} \frac{h(S_0, \cdots, S_{N-1})}{N}
  \]

- $N$-th order pdf of a stationary Gaussian process
  \[
  f_G(s) = \frac{1}{(2\pi)^{N/2} |C_N|^{1/2}} e^{-\frac{1}{2} (s - \mu_N)^T C_N^{-1} (s - \mu_N)}
  \]

- Differential entropy of $N$-th order stationary Gaussian process
  \[
  h_N^{(G)}(S) = -\int_{\mathcal{R}^N} f_G(s) \log_2 f_G(s) \, ds
  \]
  \[
  = \frac{1}{2} \log_2 ((2\pi)^N |C_N|)
  \]
  \[
  + \frac{1}{2} \log_2 e \int_{\mathcal{R}^N} f_G(s) (s - \mu_N)^T C_N^{-1} (s - \mu_N) \, ds
  \]
Differential entropy of $N$-th order stationary Gaussian process

- $N$-th order random process: pdf $f(s)$, mean $\mu_N$, covariance matrix $C_N$

\[
\int_{\mathbb{R}^N} f(s) (s - \mu_N)^T C_N^{-1} (s - \mu_N) \, ds
\]

\[
= E \left\{ (S - \mu_N)^T C_N^{-1} (S - \mu_N) \right\}
\]

\[
= E \left\{ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (S_i - \mu_i)(C^{-1})_{i,j} (S_j - \mu_j) \right\}
\]

\[
= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E \left\{ (S_i - \mu_i)(S_j - \mu_j) \right\} (C^{-1})_{i,j}
\]

\[
= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{j,i} (C^{-1})_{i,j}
\]

\[
= \sum_{i=0}^{N-1} (CC^{-1})_{j,j}
\]

\[
= N
\]
Differential entropy of $N$-th order stationary Gaussian process

- Differential entropy of $N$-th order stationary Gaussian process

\[ h_N^{(G)}(S) = - \int_{\mathcal{R}^N} f_G(s) \log_2 f_G(s) \, ds \]
\[ = \frac{1}{2} \log_2 \left( (2\pi)^N |C_N| \right) + \frac{\log_2 e}{2} \cdot N \]
\[ = \frac{1}{2} \log_2 \left( (2\pi e)^N |C_N| \right) \quad (84) \]

- $N$-th order differential entropy of any stationary non-Gaussian process is less than the $N$-th order differential entropy of a stationary Gaussian process (with co-variance matrices being equal for both processes)

\[ h_N(S) = - \int_{\mathcal{R}^N} f(s) \log_2 f(s) \, ds \]
\[ \leq - \int_{\mathcal{R}^N} f(s) \log_2 f_G(s) \, ds \]
\[ = \frac{1}{2} \log_2 \left( (2\pi)^N |C_N| \right) + \frac{\log_2 e}{2} \int_{\mathcal{R}^N} f(s)(s - \mu_N)^T C_N^{-1}(s - \mu_N) \, ds \]
\[ = h_N^{(G)}(S) \quad (85) \]
Eigendecomposition of the Covariance Matrix

- Determinant $|C_N|$: product of the eigenvalues $\xi_i$ of the matrix $C_N$,

$$C_N = A_N \Xi_N A_N^T \quad \rightarrow \quad |C_N| = |A_N| \cdot |\Xi_N| \cdot |A_N^T| = \prod_{i=0}^{N-1} \xi_i^{(N)} \quad (86)$$

- $A_N$: matrix whose columns are build by the $N$ unit-norm eigenvectors

$$A_N = \left( \mathbf{v}_0^{(N)}, \mathbf{v}_1^{(N)}, \cdots, \mathbf{v}_{N-1}^{(N)} \right) \quad (87)$$

- $\Xi_N$: diagonal matrix that contains the $N$ eigenvalues of $C_N$ on its main diagonal

$$\Xi_N = \begin{pmatrix}
\xi_0^{(N)} & 0 & \cdots & 0 \\
0 & \xi_1^{(N)} & \cdots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & 0 & \xi_{N-1}^{(N)}
\end{pmatrix} \quad (88)$$
Maximum Differential Entropy

- Inequality of arithmetic and geometric means,

\[
\left( \prod_{i=0}^{N-1} x_i \right)^{\frac{1}{N}} \leq \frac{1}{N} \sum_{i=0}^{N-1} x_i,
\]

with equality if \( x_0 = x_1 = \ldots = x_{N-1} \) (geometric mean is maximized)

- Application to product of eigenvalues

\[
|C_N| = \prod_{i=0}^{N-1} \xi_i \leq \left( \frac{1}{N} \sum_{i=0}^{N-1} \xi_i \right)^N = \sigma^{2N}
\]

\( \Rightarrow \) Equality holds if and only if all eigenvalues of \( C_N \) are the same, i.e, if and only if the random process is iid

- \( N \)-th order differential entropy of a stationary process \( S \) with a variance \( \sigma^2 \) is upper bounded by differential entropy of Gaussian iid process with the same variance

\[
h_{N}(S) \leq \frac{N}{2} \log_2 \left( 2\pi e\sigma^2 \right)
\]
Shannon Lower Bound

- $R(D)$ is nicely defined, but its computation is actually not so simple
- Shannon lower bound provides lower bound for $R(D)$

$$
R^{(I)}(D) = \lim_{N \to \infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{I_N(S; S')}{N}
= \lim_{N \to \infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{h_N(S) - h_N(S|S')}{N}
= \lim_{N \to \infty} \frac{h_N(S)}{N} - \lim_{N \to \infty} \sup_{g_N : \delta_N(g_N) \leq D} \frac{h_N(S|S')}{N}
= \overline{h}(S) - \lim_{N \to \infty} \sup_{g_N : \delta_N(g_N) \leq D} \frac{h_N(S - S'|S')}{N} \quad \text{(92)}
$$

- Shannon Lower Bound

$$
R_L(D) = \overline{h}(S) - \lim_{N \to \infty} \sup_{g_N : \delta_N(g_N) \leq D} \frac{h_N(S - S')}{N} \quad \text{(93)}
$$

- Conditioning may reduce differential entropy

$$
R(D) \geq R_L(D) \quad \text{(94)}
$$
Shannon Lower Bound for MSE Distortion

- Gaussian iid process has maximum differential entropy

\[
h_N(Z) = \frac{N}{2} \log_2 \left( 2\pi e \cdot \sigma_Z^2 \right)
\]  \hspace{1cm} (95)

- Setting \( Z = S - S' \) and \( D = \sigma_Z^2 \)

\[
h_N(S - S') = \frac{N}{2} \log_2 \left( 2\pi e \cdot D \right)
\]  \hspace{1cm} (96)

- **Shannon Lower Bound for MSE distortion**

\[
R_L(D) = \bar{h}(S) - \frac{1}{2} \log_2 \left( 2\pi e D \right)
\]

\[
D_L(R) = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(S)} \cdot 2^{-2R}
\]  \hspace{1cm} (97)

- **Shannon Lower Bound for MSE distortion and iid sources**

\[
R_L(D) = h(S') - \frac{1}{2} \log_2 \left( 2\pi e D \right)
\]

\[
D_L(R) = \frac{1}{2\pi e} \cdot 2^{2h(S)} \cdot 2^{-2R}
\]  \hspace{1cm} (98)
Shannon Lower Bound Various IID Sources

- $R_L(D)/D_L(R)$ is only different in $h(S)$ for various distributions

  **Uniform pdf:**
  \[
  h(S) = \frac{1}{2} \log_2(12\sigma^2) \quad \rightarrow \quad D_L(R) = \frac{6}{\pi e} \cdot \sigma^2 \cdot 2^{-2R} \approx 0.7
  \]  
  (99)

  **Laplacian pdf:**
  \[
  h(S) = \frac{1}{2} \log_2(2e^2\sigma^2) \quad \rightarrow \quad D_L(R) = \frac{e}{\pi} \cdot \sigma^2 \cdot 2^{-2R} \approx 0.865
  \]  
  (100)

  **Gaussian pdf:**
  \[
  h(S) = \frac{1}{2} \log_2(2\pi e\sigma^2) \quad \rightarrow \quad D_L(R) = \sigma^2 \cdot 2^{-2R}
  \]  
  (101)
Asymptotic Tightness of the Shannon Lower Bound

- Shannon lower bound approaches distortion rate function for small distortions or high rates
  \[
  \lim_{D \to 0} R(D) - R_L(D) = 0. 
  \] (102)

- Comparison of \( D(R) \) with \( D_L(R) \) for the Laplacian iid source
Rate Distortion Function for Amplitude-Continuous Random Variables

\( R_L(D) \) for Various IID Sources

- Shannon lower bound \( R_L(D) \) for
  - Uniform iid process: red
  - Laplace iid process: green
  - Gauss iid process: blue
- Left: MSE vs. Rate, right: SNR vs. Rate

\[
\text{SNR} = 10 \log_{10} \frac{\sigma^2}{\text{MSE}}
\]  

(103)
**Rate Distortion Theory**

**Rate Distortion Function for Amplitude-Continuous Random Variables**

\[ R_L(D) \] for **Gaussian Sources with Memory**

- **Differential entropy for Gaussian sources**

\[
h_N^{(G)}(S) = \frac{1}{2} \log_2 \left( (2\pi e)^N |C_N| \right)
\]  

(104)

- **Shannon lower bound for MSE distortion**

\[
R_L(D) = \lim_{N \to \infty} \frac{h_N^{(G)}(S)}{N} - \frac{1}{2} \log_2(2\pi e D)
\]

\[
= \lim_{N \to \infty} \log_2 \left( (2\pi e)^N |C_N| \right) - \frac{1}{2} \log_2(2\pi e D)
\]

\[
= \frac{1}{2} \log_2(2\pi e) + \lim_{N \to \infty} \frac{\log_2(|C_N|)}{2N} - \frac{1}{2} \log_2(2\pi e D)
\]

\[
= \lim_{N \to \infty} \frac{\log_2 |C_N|}{2N} - \frac{1}{2} \log_2 D
\]

\[
= \lim_{N \to \infty} \frac{1}{2N} \sum_{i=0}^{N-1} \log_2 \xi_i^{(N)} - \frac{1}{2} \log_2 D
\]  

(105)
Grenander and Szegö’s theorem

- Assume zero-mean process: $C_N = R_N$
- Given the conditions
  - $R_N$ is a sequence of Hermitian Toeplitz matrices with elements $\phi_k$ on the $k$-th diagonal
  - The infimum $\Phi_{\text{inf}} = \inf_\omega \Phi(\omega)$ and supremum $\Phi_{\text{sup}} = \sup_\omega \Phi(\omega)$ of the Fourier series are finite
    \[
    \Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi_k e^{-j\omega k} \tag{106}
    \]
  - The function $G$ is continuous in the interval $[\Phi_{\text{inf}}, \Phi_{\text{sup}}]$
- The following expression holds
  \[
  \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} G(\xi_i^{(N)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Phi(\omega)) \, d\omega \tag{107}
  \]
  where $\xi_i^{(N)}$, for $i = 0, 1, \ldots, N - 1$, denote the eigenvalues of the $N$-th matrix $R_N$
Power Spectral Density of a Gauss-Markov Process

- Zero-mean Gauss-Markov process with $|\rho| < 1$

\[ S_n = Z_n + \rho \cdot S_{n-1} \quad (108) \]

- Auto-correlation function

\[ \phi[k] = \sigma^2 \cdot |\rho|^k \quad (109) \]

- Using the relationship

\[ \sum_{k=1}^{\infty} a^k e^{-j k x} = \frac{a}{e^{-j x} - a} \quad (110) \]

we obtain

\[ \Phi_{SS}(\omega) = \sum_{k=-\infty}^{\infty} \phi[k] \cdot e^{-j \omega k} \]

\[ = \sum_{k=-\infty}^{\infty} \sigma^2 \cdot |\rho|^k \cdot e^{-j \omega k} \]

\[ = \sigma^2 \cdot \left( 1 + \frac{\rho}{e^{-j \omega} - \rho} + \frac{\rho}{e^{j \omega} - \rho} \right) \]

\[ = \sigma^2 \cdot \frac{1 - \rho^2}{1 - 2 \rho \cos \omega + \rho^2} \quad (111) \]
$R_L(D)$ for Gaussian-Markov Processes

- Shannon lower bound for a zero-mean Gauss-Markov process with $|\rho| < 1$

\[
R_L(D) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \frac{\Phi_{SS}(\omega)}{D} \, d\omega
\]

\[
= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \frac{\sigma^2 (1 - \rho^2)}{D} \, d\omega
\]

\[
= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 (1 - 2\rho \cos \omega + \rho^2) \, d\omega
\]

\[
= 0
\]

\[
= \frac{1}{2} \log_2 \frac{\sigma^2 (1 - \rho^2)}{D}
\]

(112)

where we used

\[
\int_{0}^{\pi} \ln(a^2 - 2ab \cos x + b^2) \, dx = 2\pi \ln a
\]

(113)

- Shannon lower bound for information distortion rate function

\[
D_L(R) = (1 - \rho^2) \sigma^2 2^{-2R}
\]

(114)
Goal: derive information rate distortion function for Gaussian iid sources

For distortions $D < \sigma^2$, choose

$$ f_{S'}(s') = \frac{1}{\sqrt{2\pi(\sigma^2 - D)}} e^{-\frac{(s' - \mu)^2}{2(\sigma^2 - D)}} $$

(115)

$$ f_{Z|S'}(z|s') = \frac{1}{\sqrt{2\pi D}} e^{-\frac{z^2}{2D}} = f_Z(z) $$

(116)

with

$$ S = S' + Z $$

(117)

The random variables $S'$ and $Z$ are independent, since $f_{Z|S'}(z|s') = f_Z(z)$. Hence, the pdf $f_S$ is given by the convolution of $f_{S'}$ and $f_Z$

Pdfs $f_{S'}(s')$ and $f_Z(z)$ are possible choices, since convolution of two Gaussians yields a Gaussian

$$ f_S(s) = f_{S'}(s') \ast f_Z(z) $$

(118)

with mean and variance adding up

$$ \mu = \mu + 0 \quad \text{and} \quad \sigma^2 = (\sigma^2 - D) + D $$

(119)
\( R(D) \) for Gaussian IID Sources II

- Distortion given by variance of difference process \( Z_n = S_n - S_n' \)
  \[
  \delta_1(g_1) = E \{(S_n - S_n')^2\} = E \{Z_n^2\} = D. \tag{120}
  \]

- Mutual information
  \[
  I_1(g_1) = h(S_n) - h(S_n | S_n') \tag{121}
  = h(S_n) - h(S_n - S_n' | S_n') \tag{122}
  = h(S_n) - h(Z_n | S_n') \tag{123}
  = h(S_n) - h(Z_n) \tag{124}
  = \frac{1}{2} \log_2(2\pi e\sigma^2) - \frac{1}{2} \log_2(2\pi eD) \tag{125}
  \]

  \[
  R(D) = \frac{1}{2} \log_2 \frac{\sigma^2}{D} \tag{126}
  \]

- The information rate distortion function coincides with the Shannon lower bound for Gaussian iid sources
We assume a memoryless Gaussian source with a variance $\sigma^2$

For the memoryless Gaussian source, the Shannon lower bound coincides with the rate distortion function

The information rate distortion function is given as

$$R(D) = \begin{cases} \frac{1}{2} \log_2 \frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

(127)

The information distortion rate function is given as

$$D(R) = \sigma^2 \cdot 2^{-2R}$$

(128)

The signal-to-noise ratio (SNR) is given as

$$\text{SNR}(R) = 10 \cdot \log_{10} \frac{\sigma^2}{D(R)} = 10 \cdot \log_{10} 2^{2R} \approx 6R \text{ dB}$$

(129)
\( R(D) \) for a Gaussian Source with Memory

- Assume a stationary Gaussian random process with zero mean and pdf

\[
f_{S}^{(G)}(s) = \frac{1}{(2\pi)^{N/2} |C_N|^{1/2}} e^{-\frac{1}{2}(s-\mu_N)^T C_N^{-1} (s-\mu_N)}
\]  

(130)

- Eigendecomposition of covariance matrix \( C_N \),

\[
C_N = A_N \Xi_N A_N^T
\]  

(131)

- \( A_N \): matrix whose columns are build by the \( N \) unit-norm eigenvectors

\[
A_N = \left( v_0^{(N)}, v_1^{(N)}, \cdots, v_{N-1}^{(N)} \right)
\]  

(132)

- \( \Xi_N \): diagonal matrix that contains the \( N \) eigenvalues of \( C_N \) on its main diagonal

\[
\Xi_N = \begin{pmatrix}
\xi_0^{(N)} & 0 & \cdots & 0 \\
0 & \xi_1^{(N)} & \cdots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & 0 & \xi_{N-1}^{(N)}
\end{pmatrix}
\]  

(133)
Signal Space Rotation

- Given stationary Gaussian source \( \{S_n\} \): construct source \( \{U_n\} \) by decomposing \( \{S_n\} \) into vectors \( S \) of size \( N \) and applying the transform

\[
U = A_N^{-1} (S - \mu_N) = A_T^N (S - \mu_N) \tag{134}
\]

- Linear transformation of a Gaussian random vector results in another Gaussian random vector
- The chosen transform yields independent random variables \( U_i \)

\[
f_U(u) = \frac{1}{(2\pi)^{N/2} |\Xi_N|^{1/2}} e^{-\frac{1}{2} u^T \Xi_N^{-1} u} = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi \xi_i^{(N)}}} e^{-\frac{u_i^2}{2 \xi_i^{(N)}}} \tag{135}
\]

- Mean

\[
E\{U\} = A_T^N (E\{S\} - \mu_N) = A_T^N (\mu_N - \mu_N) = 0 \tag{136}
\]

- Covariance

\[
E\{UU^T\} = A_T^N E\{(S - \mu_N)(S - \mu_N)^T\} A_N \]
\[
= A_T^N C_N A_N = \Xi_N \tag{137}
\]
Distortion and Mutual Information

- Inverse transform after compression identical to forward transform
  \[ S' = A_N U' + \mu_N, \]  
  \[ (138) \]

- With
  \[ (U' - U) = A_N^T (S' - S) \leftrightarrow (S' - S) = A_N (U' - U) \]  
  \[ (139) \]

- MSE distortion between any realization \( s \) of \( S \) and its reconstruction \( s' \)
  \[ d_N(s; s') = \frac{1}{N} \sum_{i=0}^{N-1} (s_i - s'_i)^2 = \frac{1}{N} (s - s')^T (s - s') \]
  \[ = \frac{1}{N} (u - u')^T A_N^T A_N (u - u') = \frac{1}{N} (u - u')^T (u - u') \]
  \[ = \frac{1}{N} \sum_{i=0}^{N-1} (u_i - u'_i)^2 = d_N(u; u') \]  
  \[ (140) \]

- Since \( A_N A_N^T = I_N \) (identity matrix)
  \[ I_N(S; S') = I_N(U; U') \]  
  \[ (141) \]
Distortion Rate Function

- **Mutual information and average distortion considering independence of the components** $U_i$

  \[
  I_N(g_N^Q) = \sum_{i=0}^{N-1} I_1(g_i^Q) \quad \text{and} \quad \delta_N(g_N^Q) = \frac{1}{N} \sum_{i=0}^{N-1} \delta_1(g_i^Q)
  \]  
  \(\text{(142)}\)

- **$N$-th order distortion rate function** $D_N(R)$

  \[
  D_N(R) = \frac{1}{N} \sum_{i=0}^{N-1} D_i(R_i) \quad \text{with} \quad R = \frac{1}{N} \sum_{i=0}^{N-1} R_i
  \]  
  \(\text{(143)}\)

- $D_i(R_i)$: **first-order distortion rate function for Gaussian iid processes for component** $U_i$

  \[
  D_i(R_i) = \sigma_i^2 2^{-2R_i} = \xi_i^{(N)} 2^{-2R_i}
  \]  
  \(\text{(144)}\)

  with $\xi_i^{(N)}$ being the eigenvalues of $C_N$
Optimal Bit Allocation

- Task

\[
\min_{R_0, R_1, \ldots, R_{N-1}} D_N(R) = \frac{1}{N} \sum_{i=0}^{N-1} \xi^{(N)}_i 2^{-2R_i} \quad \text{such that} \quad R \geq \frac{1}{N} \sum_{i=0}^{N-1} R_i
\]

- Comparison on different types of mean computations

\[
D_N(R) = \frac{1}{N} \sum_{i=0}^{N-1} \xi^{(N)}_i 2^{-2R_i} \geq \left( \prod_{i=0}^{N-1} \xi^{(N)}_i 2^{-2R_i} \right)^\frac{1}{N} = \left( \prod_{i=0}^{N-1} \xi^{(N)}_i \right)^\frac{1}{N} \cdot 2^{-2R}
\]

with \( \prod_{i=0}^{N-1} 2^{-2R_i} = 2^{-2R_0} \cdot 2^{-2R_1} \ldots 2^{-2R_{N-1}} = 2^{-\sum_{i=0}^{N-1} 2R_i} = 2^{-2RN} \)

- Expression on the right-hand side of above inequality is constant:
  equality achieved when all terms \( \xi^{(N)}_i 2^{-2R_i} = \tilde{\xi}(N) 2^{-2R} \)

\[
R_i = R + \frac{1}{2} \log_2 \frac{\xi^{(N)}_i}{\tilde{\xi}(N)} = \frac{1}{2} \log_2 \frac{\xi^{(N)}_i}{\tilde{\xi}(N) 2^{-2R}} \quad \text{with} \quad \tilde{\xi}(N) = \left( \prod_{i=0}^{N-1} \xi^{(N)}_i \right)^\frac{1}{N}
\]
Parametric Formulation

- So far, we have ignored that $R_i$ cannot be less than 0

$$R_i = \frac{1}{2} \log_2 \frac{\xi_i^{(N)}}{\tilde{\xi}^{(N)} 2^{-2R}} \geq 0 \rightarrow R_i = 0 \text{ if } \xi_i^{(N)} \leq \tilde{\xi}^{(N)} 2^{-2R} \quad (145)$$

- Introducing the parameter $\theta \geq 0$

$$R_i = \begin{cases} \frac{1}{2} \log_2 \frac{\xi_i^{(N)}}{\theta}, & 0 \leq \theta \leq \xi_i^{(N)} \\ 0, & \theta \geq \xi_i^{(N)} \end{cases} \quad \text{and} \quad D_i = \begin{cases} \theta, & 0 \leq \theta \leq \xi_i^{(N)} \\ \xi_i^{(N)}, & \theta \geq \xi_i^{(N)} \end{cases}$$

- or

$$R_i(\theta) = \max \left( 0, \frac{1}{2} \log_2 \frac{\xi_i^{(N)}}{\theta} \right) \quad \text{and} \quad D_i(\theta) = \min \left( \xi_i^{(N)}, \theta \right) \quad (146)$$

- Parametric expressions

$$D_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} D_i = \frac{1}{N} \sum_{i=0}^{N-1} \min \left( \xi_i^{(N)}, \theta \right) \quad (147)$$

$$R_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} R_i = \frac{1}{N} \sum_{i=0}^{N-1} \max \left( 0, \frac{1}{2} \log_2 \frac{\xi_i^{(N)}}{\theta} \right) \quad (148)$$
Rate Distortion Theory

Rate Distortion Function for Amplitude-Continuous Random Variables

\( R^{(I)}(D) \) for \( N \to \infty \)

- Recall Grenander and Szegö's theorem for infinite Toeplitz matrices
  \[
  \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} G(\xi_{i}^{(N)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Phi(\omega))d\omega \tag{149}
  \]

- Parametric expressions
  \[
  \lim_{N \to \infty} D_{N}(\theta) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} \min \left( \xi_{i}^{(N)}, \theta \right) \tag{150}
  \]
  \[
  \lim_{N \to \infty} R_{N}(\theta) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} \max \left( 0, \frac{1}{2} \log_{2} \frac{\xi_{i}^{(N)}}{\theta} \right) \tag{151}
  \]

- With the spectral density function \( \Phi_{ss}(\omega) \)
  \[
  D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \{ \Phi_{ss}(\omega), \theta \} d\omega
  \]
  \[
  R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \frac{1}{2} \log_{2} \frac{\Phi_{ss}(\omega)}{\theta} \right\} d\omega \tag{152}
  \]
Illustration of Minimization Approach

- It can be interpreted that at each frequency, the variance of the process as given by the spectral density function $\Phi_{ss}(\omega)$ is compared to the parameter $\theta$, which represents the mean squared error which is the considered parameter.
- When $\Phi_{ss}(\omega)$ is found to be larger than $\theta$, the rate $\frac{1}{2} \log_2 \frac{\Phi_{ss}(\omega)}{\theta}$ is assigned, otherwise zero rate is assigned to that frequency component.
Calculate $R(D)$ for Gauss-Markov process with $|\rho| < 1$ and variance $\sigma^2$

$$S_n = Z_n + \rho \cdot S_{n-1}$$ (153)

Auto-correlation function and spectral density function are given as

$$\phi[k] = \sigma^2 |\rho|^k \quad \Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi[k] e^{-jk\omega} = \frac{\sigma^2(1 - \rho^2)}{1 - 2\rho \cos \omega + \rho^2}$$ (154)

All rate terms in the parametric formulation of $R(D)$ are positive when

$$\frac{D}{\sigma^2} \leq \frac{1 - \rho}{1 + \rho}$$ (155)

and $R(D)$ can be written

$$R(D) = \frac{1}{2} \log_2 \frac{\sigma^2(1 - \rho^2)}{D}$$ (156)
Corresponding distortion rate function for \( R \geq \log_2(1 + \rho) \) is given by

\[
D(R) = (1 - \rho^2) \cdot \sigma^2 \cdot 2^{-2R}
\]  

(157)

- Includes result for Gaussian iid sources (\( \rho = 0 \))
- Below: \( R(D) \) for Gauss-Markov sources for different values of \( \rho \)
Summary on Rate Distortion Theory

- Rate distortion theory: minimum bit-rate for a given distortion
- Operational rate distortion function difficult to compute - derive information rate distortion function instead and show that they coincide
- $R(D)$ for discrete sources is a convex and continuous function in the range $R = (0, H(S)), D = (D_{\text{max}}, 0)$
- $R(D)$ for amplitude-continuous sources: similar to $R(D)$ for discrete sources, except $R(D) \to \infty$ for $D \to 0$
- Shannon lower bound $R_L(D)$ can often be computed as analytical lower bound for $R(D)$
- $R(D)$ for Gaussian iid sources and MSE: 6 dB/bit
- Any other source than the Gaussian iid source with equal variance requires less bits given the same mean squared error
- $R(D)$ for Gaussian source with memory and MSE:
  - Encode spectral components independently
  - Introduce white noise
  - Suppress small spectral components