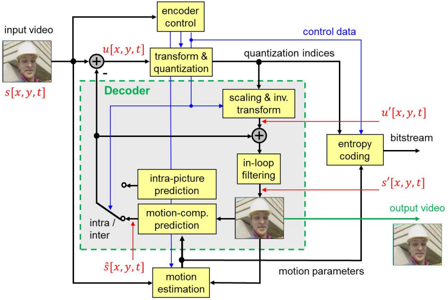
# Source Coding and Compression


## Heiko Schwarz

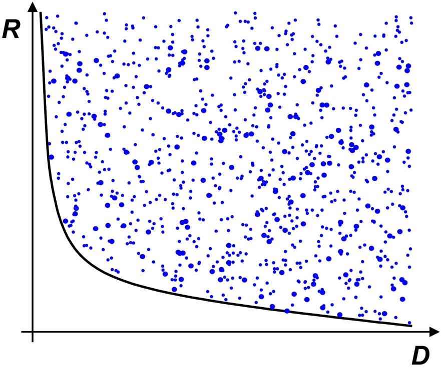





*Contact*:

Dr.-Ing. Heiko Schwarz
heiko.schwarz@hhi.fraunhofer.de

# Rate-Distortion Theory



# Outline

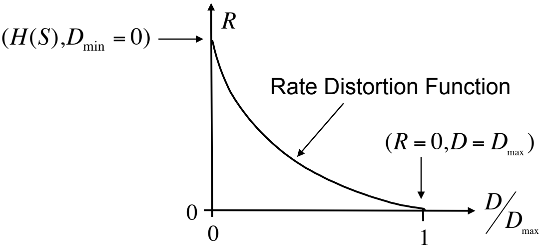
# Rate-Distortion Theory – Motivation

**Lossy coding**: Decoded signal is an approximation of original

**Rate-distortion theory**: Information theoretical bounds for lossy compression

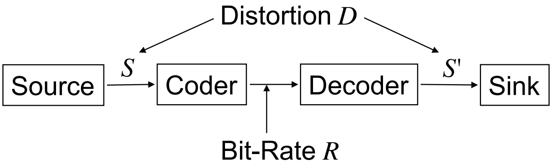
- Results are obtained without consideration of a specific coding method
- Goal of rate-distortion theory is to calculate the minimum transmission bit rate for a given distortion and source

Example for a rate-distortion function of a discrete iid source



# Transmission System and Variables

Transmission system

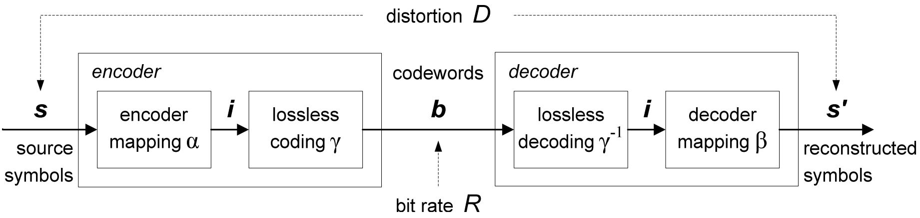



- Derivation in two steps
  - Define $S$, $S'$, coder/decoder, distortion $D$ and rate $R$
  - Establish a functional relationship between $S$, $S'$, $D$, and $R$

- For two types of random variables
  - Discrete random variables
  - Continuous-amplitude random variables (Gaussian, Laplacian, etc.)

# General Structure of Lossy Source Codecs

- **Encoder**:
  - Irreversible encoder mapping $\alpha : \boldsymbol{s} \to \boldsymbol{i}$
  - Lossless mapping $\gamma : \boldsymbol{i} \to \boldsymbol{b}$
- **Decoder**:
  - Lossless mapping $\gamma^{-1} : \boldsymbol{b} \to \boldsymbol{i}$
  - Decoder mapping $\beta : \boldsymbol{i} \to \boldsymbol{s}'$



# Source Codes

- A **source code** $Q = (\alpha, \beta, \gamma)$ is given by an encoder mapping $\alpha$, a decoder mapping $\beta$ and a lossless mapping $\gamma$
- Special case: $N$-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 $\boldsymbol{s}^{(N)} = \{s_0, \cdots, s_{N-1}\}$ is mapped to a vector of $K$ quantization indexes

$$\boldsymbol{i}^{(K)} = \alpha_N(\boldsymbol{s}^{(N)}) \tag{169}$$

  - Resulting vector of indexes $\boldsymbol{i}^{(N)}$ is converted into a bit sequence

$$\boldsymbol{b}^{(\ell)} = \gamma_N(\boldsymbol{i}^{(K)}) = \gamma_N(\alpha_N(\boldsymbol{s}^{(N)})) \tag{170}$$

  - At decoder side, index vector is recovered

$$\boldsymbol{i}^{(K)} = \gamma_N^{-1}(\boldsymbol{b}^{(\ell)}) = \gamma_N^{-1}(\gamma_N(\boldsymbol{i}^{(K)})) \tag{171}$$

  - Index vector is mapped to a block of reconstructed samples $\boldsymbol{s'}^{(N)} = \{s'_0, \cdots, s'_{N-1}\}$

$$\boldsymbol{s'}^{(N)} = \beta_N(\boldsymbol{i}^{(K)}) = \beta_N(\alpha_N(\boldsymbol{s}^{(N)})) \tag{172}$$

# Distortion

- **Distortion**: Measure of difference between a block of $N$ input samples $\boldsymbol{s}^{(N)} = \{s_0, s_1, \ldots, s_{N-1}\}$ and the corresponding block of reconstructed samples $\boldsymbol{s'}^{(N)} = \{s'_0, s'_1, \ldots, s'_{N-1}\}$,

$$d_N\left(\boldsymbol{s}^{(N)}, \boldsymbol{s'}^{(N)}\right)$$

- Typically: **Additive distortion measures**

$$d_N(\boldsymbol{s}^{(N)}, \boldsymbol{s'}^{(N)}) = \frac{1}{N}\sum_{i=0}^{N-1} d_1(s_i, s'_i) \qquad (173)$$

  with the single symbol distortion $d_1(s, s') \geq 0$ (equality, if and only if $s = s'$)

- In this lecture: Mean squared error $d_1(s, s') = (s - s')^2$

$$d_N\left(\boldsymbol{s}^{(N)}, \boldsymbol{s'}^{(N)}\right) = \frac{1}{N}\sum_{i=0}^{N-1} d_1(s_i, s'_i) = \frac{1}{N}\sum_{i=0}^{N-1} (s_i - s'_i)^2 \qquad (174)$$

# Average Distortion for Source Codes

- Average distortion for a stationary random process $\boldsymbol{S} = \{S_n\}$ and an $N$-dimensional block source code $Q_N = \{\alpha_N, \beta_N, \gamma_N\}$

$$\delta(Q_N) = E\Big\{d_N\big(\boldsymbol{S}^{(N)}, \beta_N(\alpha_N(\boldsymbol{S}^{(N)}))\big)\Big\} \quad (175)$$

$$= \int_{\mathcal{R}^N} f(\boldsymbol{s})\, d_N\big(\boldsymbol{s}, \beta_N(\alpha_N(\boldsymbol{s}))\big)\, \mathrm{d}\boldsymbol{s} \quad (176)$$

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

$$\delta(Q) = \lim_{N\to\infty} E\Big\{d_N\big(\boldsymbol{S}^{(N)}, \beta_N(\alpha_N(\boldsymbol{S}^{(N)}))\big)\Big\} \quad (177)$$

- For additive distortion measures (such as the MSE distortion)

$$\delta(Q) = \delta(S, S') = E\{d_1(S, S')\} = \int_s \int_{s'} f_{SS'}(s, s')\, d_1(s, s')\, \mathrm{d}s\, \mathrm{d}s' \quad (178)$$

# Average Rate for Source Codes

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

$$r_N(\boldsymbol{s}^{(N)}) = \frac{1}{N}\left|\gamma_N(\alpha_N(\boldsymbol{s}^{(N)}))\right| \qquad \text{with} \qquad \boldsymbol{b}^{(\ell)} = \gamma_N(\alpha_N(\boldsymbol{s}^{(N)})) \quad (179)$$

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

$$r(Q_N) = \frac{1}{N} E\Big\{\left|\gamma_N(\alpha_N(\boldsymbol{S}^{(N)}))\right|\Big\} \quad (180)$$

$$= \frac{1}{N}\int_{\mathcal{R}^N} f(\boldsymbol{s})\left|\gamma_N(\alpha_N(\boldsymbol{s}))\right| \mathrm{d}\boldsymbol{s} \quad (181)$$

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

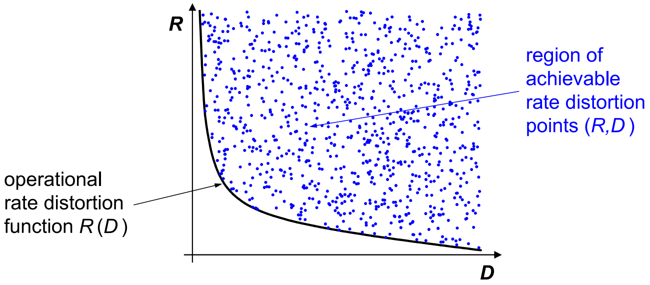
$$r(Q) = \lim_{N\to\infty} \frac{1}{N} E\Big\{\left|\gamma_N(\alpha_N(\boldsymbol{S}^{(N)}))\right|\Big\} \quad (182)$$

# Operational Rate-Distortion Function

For given source $\boldsymbol{S}$:

- Each code $Q$ is associated with a rate distortion point $(R, D) = (r(Q), \delta(Q))$
- A rate distortion point is **achievable**, if there exist a code $Q$ such that $r(Q) \leq R$ and $\delta(Q) \leq D$
- The **operational rate-distortion function** $R(D)$ and its inverse, the **operational distortion-rate function** $D(R)$ are defined by

$$\boxed{R(D) = \inf_{Q:\, \delta(Q) \leq D} r(Q)} \qquad \boxed{D(R) = \inf_{Q:\, r(Q) \leq R} \delta(Q)} \tag{183}$$



# Motivation for Information Rate-Distortion Function

## Operational rate-distortion function

- Defined by

$$R(D) = \inf_{Q:\,\delta(Q)\leq D} r(Q) \tag{184}$$

- Specifies a fundamental performance bound for lossy source coding
- Difficulty to evaluate (minimization over all possible codes)

## Information rate-distortion function

- Introduced by SHANNON in [Shannon 1948; Shannon1959]
- Obtain expression of rate-distortion bound that involves the distribution of the source using **mutual information**
- Show that information rate-distortion function is achievable

# Mutual Information for Discrete Random Variables

- **Mutual information** between two discrete random variables $A$ and $B$ is defined by

$$\boxed{I(A;B) = H(A) - H(A|B)} \tag{185}$$

- Entropy $H(A)$ is a measure of uncertainty about random variable $A$
- Conditional entropy $H(A|B)$ is a measure of uncertainty about random variable $A$ after observing random variable $B$
- Mutual information is a measure for the reduction of uncertainty about $A$ due to the observation of $B$

$\Longrightarrow$ **Average amount of information that $A$ carries about $B$**

- Mutual information for discrete random variables $A \in \mathcal{A}$ and $B \in \mathcal{B}$

$$I(A;B) = H(A) - H(A|B) = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} p(a,b) \log_2 \frac{p(a|b)}{p(a)} \tag{186}$$

# Mutual Information for Discrete Random Variables

- Mutual information rewritten using Bayes' rule

$$\begin{aligned} I(A;B) &= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p(a,b)\log_2\frac{p(a|b)}{p(a)} \\ &= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p(a,b)\log_2\frac{p(a,b)}{p(a)\,p(b)} \\ &= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p(a,b)\log_2\frac{p(b|a)}{p(b)} \\ &= H(B) - H(B|A) \end{aligned} \tag{187}$$

- Mutual information between two random variables $A$ and $B$ represents the average amount of information that
  - the random variable $A$ carries about the random variable $B$, or
  - the random variable $B$ carries about the random variable $A$

# Mutual Information for Discrete Random Vectors

- Mutual information between two random variables $A$ and $B$

$$\begin{aligned} I(A;B) &= H(A) - H(A|B) \\ &= H(B) - H(B|A) \\ &= \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} p(a,b) \log_2 \frac{p(a,b)}{p(a)\,p(b)} \end{aligned} \qquad (188)$$

- Extension to $N$-dimensional random vectors $\boldsymbol{A} = (A_0, A_1, \cdots, A_{N-1})^{\mathrm{T}}$ and $\boldsymbol{B} = (B_0, B_1, \cdots, B_{N-1})^{\mathrm{T}}$

$$\begin{aligned} I_N(\boldsymbol{A};\boldsymbol{B}) &= H_N(\boldsymbol{A}) - H_N(\boldsymbol{A}|\boldsymbol{B}) \\ &= H_N(\boldsymbol{B}) - H_N(\boldsymbol{B}|\boldsymbol{A}) \\ &= \sum_{\boldsymbol{a}\in\mathcal{A}^N} \sum_{\boldsymbol{b}\in\mathcal{B}^N} p(\boldsymbol{a},\boldsymbol{b}) \log_2 \frac{p(\boldsymbol{a},\boldsymbol{b})}{p(\boldsymbol{a})\,p(\boldsymbol{b})} \end{aligned} \qquad (189)$$

# Properties of Mutual Information for Discrete RV

- Mutual information between discrete random vectors $\boldsymbol{A}$ and $\boldsymbol{B}$

$$I_N(\boldsymbol{A};\boldsymbol{B}) = H_N(\boldsymbol{A}) - H_N(\boldsymbol{A}|\boldsymbol{B}) \tag{190}$$

$$= H_N(\boldsymbol{B}) - H_N(\boldsymbol{B}|\boldsymbol{A}) \tag{191}$$

- Since the conditional entropies are non-negative

$$I_N(\boldsymbol{A};\boldsymbol{B}) \leq H_N(\boldsymbol{A}) \tag{192}$$

$$I_N(\boldsymbol{A};\boldsymbol{B}) \leq H_N(\boldsymbol{B}) \tag{193}$$

- For independent random vectors $\boldsymbol{A}$ and $\boldsymbol{B}$

$$I_N(\boldsymbol{A};\boldsymbol{B}) = 0 \tag{194}$$

- If the random vector $\boldsymbol{B}$ is a deterministic function of the random vector $\boldsymbol{A}$,

$$\boldsymbol{B} = f(\boldsymbol{A}) \qquad \Longrightarrow \qquad I_N(\boldsymbol{A};\boldsymbol{B}) = H_N(\boldsymbol{B}) \tag{195}$$

# Mutual Information for Coding of Discrete Sources

- Consider mutual information $I_N(\boldsymbol{S};\boldsymbol{S'})$ between a vector of $N$ successive input samples $\boldsymbol{S}$ and the corresponding vector of $N$ reconstructed samples $\boldsymbol{S'}$

$$\begin{aligned} I_N(\boldsymbol{S};\boldsymbol{S'}) &= H_N(\boldsymbol{S'}) - H_N(\boldsymbol{S'}|\boldsymbol{S}) \\ &\leq H_N(\boldsymbol{S'}) \end{aligned} \tag{196}$$

  where equality is achieved if and only if the vector $\boldsymbol{S'}$ of reconstructed samples is a deterministic function of the input vector $\boldsymbol{S}$

- Recall: Fundamental bound for lossless coding

$$r(Q) \geq \bar{H}(\boldsymbol{S'}) = \lim_{N\to\infty} \frac{H_N(\boldsymbol{S'})}{N} \tag{197}$$

- Rate of for code $Q$

$$\boxed{r(Q) \geq \lim_{N\to\infty} \frac{H_N(\boldsymbol{S'})}{N} \geq \lim_{N\to\infty} \frac{I_N(\boldsymbol{S'};\boldsymbol{S})}{N}} \tag{198}$$

# Mutual Information for Continuous Random Variables

Remember: Discrete random variables

- Mutual information for discrete random variables $A$ and $B$

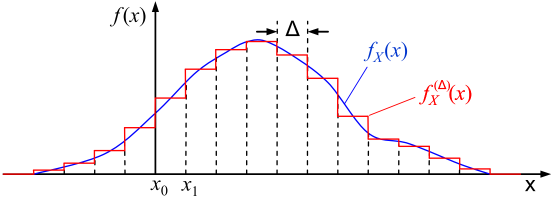
$$I(A;B) = H(A) - H(A|B) \tag{199}$$
$$= H(B) - H(B|A) \tag{200}$$

- For continuous random variables, the discrete entropies are not defined (they approach infinity)

Definition of mutual information for continuous random variables

- Quantize pdfs with a quantization step size $\Delta$
- Calculate mutual information for resulting discrete random variables
- Consider limit for quantization step size $\Delta$ approaching zero

# Discretization of Continuous Random Variables




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

$$\forall x : x_i \leq x < x_{i+1} \qquad f_X^{(\Delta)}(x) = \frac{1}{\Delta} \int_{x_i}^{x_{i+1}} f_X(x')\, \mathrm{d}x' \tag{201}$$

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

$$\boxed{p_{X_\Delta}(x_i) = \int_{x_i}^{x_{i+1}} f_X(x')\, \mathrm{d}x' = f_X^{(\Delta)}(x_i) \cdot \Delta} \tag{202}$$

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

$$\boxed{p_{X_\Delta Y_\Delta}(x_i, y_j) = f_{XY}^{(\Delta)}(x_i, y_j) \cdot \Delta^2} \tag{203}$$

# Mutual Information for Continuous Random Variables

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

$$I(X_\Delta; Y_\Delta) = \sum_{x_i \in \mathcal{A}_{X_\Delta}} \sum_{y_i \in \mathcal{A}_{Y_\Delta}} p_{X_\Delta Y_\Delta}(x_i, y_j) \log_2 \frac{p_{X_\Delta Y_\Delta}(x_i, y_i)}{p_{X_\Delta}(x_i)\, p_{Y_\Delta}(y_j)} \quad (204)$$

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

- The mutual information $I(X;Y)$ for the continuous random variables $X$ and $Y$ is obtained for $\Delta$ approaching zero,

$$I(X;Y) = \lim_{\Delta \to 0} I(X_\Delta; Y_\Delta) \quad (205)$$

- For $\Delta \to 0$, the piecewise constant pdf approximations $f_{XY}^{(\Delta)}$, $f_X^{(\Delta)}$, and $f_Y^{(\Delta)}$ approach the pdfs $f_{XY}$, $f_X$, and $f_Y$, and we obtain

$$\boxed{I(X;Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x)\, f_Y(y)} \,\mathrm{d}x\,\mathrm{d}y} \quad (206)$$

# Mutual Information for Continuous Random Vectors

- Mutual information for continuous random variables $X$ and $Y$

$$I(X;Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) \, \log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y \tag{207}$$

- Consider extension to $N$-dimensional random vectors $\boldsymbol{X} = (X_0, X_1, \cdots, X_{N-1})^{\mathrm{T}}$ and $\boldsymbol{Y} = (Y_0, Y_1, \cdots, Y_{N-1})^{\mathrm{T}}$

$$I_N(\boldsymbol{X};\boldsymbol{Y}) = \int_{\mathcal{R}^N} \int_{\mathcal{R}^N} f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) \, \log_2 \frac{f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{X}}(\boldsymbol{x})\, f_{\boldsymbol{Y}}(\boldsymbol{y})} \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \tag{208}$$

- Using $f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x},\boldsymbol{y})$, we can also write

$$I_N(\boldsymbol{X};\boldsymbol{Y}) = \int_{\mathcal{R}^N} \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x},\boldsymbol{y}) \, \log_2 \frac{f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \tag{209}$$

# Mutual Information between Discrete and Continuous RV

- Let $\boldsymbol{Y}$ be a discrete random vector with alphabet $\mathcal{A}_Y^N$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} \delta(\boldsymbol{y}-\boldsymbol{a})\, p_{\boldsymbol{Y}}(\boldsymbol{a}) \tag{210}$$

$$f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} \delta(\boldsymbol{y}-\boldsymbol{a})\, p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \tag{211}$$

- Rewriting mutual information using above pmfs yields

$$\begin{aligned} I_N(\boldsymbol{X};\boldsymbol{Y}) &= \int_{\mathcal{R}^N}\int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x},\boldsymbol{y}) \log_2 \frac{f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\ &= \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x})}{p_{\boldsymbol{Y}}(\boldsymbol{a})}\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \Big( \log_2 p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) - \log_2 p_{\boldsymbol{Y}}(\boldsymbol{a}) \Big)\,\mathrm{d}\boldsymbol{x} \end{aligned} \tag{212}$$

# Mutual Information between Discrete and Continuous RV

- Continue reformulation of mutual information $I_N(\boldsymbol{X};\boldsymbol{Y})$

$$
\begin{aligned}
I_N(\boldsymbol{X};\boldsymbol{Y}) &= \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \Big( \log_2 p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) - \log_2 p_{\boldsymbol{Y}}(\boldsymbol{a}) \Big)\, \mathrm{d}\boldsymbol{x} \\
&= -\sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} \left( \int_{\mathcal{R}^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \right) \log_2 p_{\boldsymbol{Y}}(\boldsymbol{a}) \\
&\quad + \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) \left( \sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \log_2 p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \right) \mathrm{d}\boldsymbol{x} \\
&= -\sum_{\boldsymbol{a}\in\mathcal{A}_Y^N} p_{\boldsymbol{Y}}(\boldsymbol{a}) \log_2 p_{\boldsymbol{Y}}(\boldsymbol{a}) - \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x})\, H_N(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\
&= H_N(\boldsymbol{Y}) - \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x})\, H_N(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \qquad (213)
\end{aligned}
$$

# Mutual Information between Discrete and Continuous RV

- Mutual information between a discrete random vector $\boldsymbol{Y}$ and a continuous random vector $\boldsymbol{X}$

$$I_N(\boldsymbol{X};\boldsymbol{Y}) = H_N(\boldsymbol{Y}) - \int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x})\, H_N(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \qquad (214)$$

  where $H_N(\boldsymbol{Y})$ is the entropy of the discrete random vector $\boldsymbol{Y}$ and

$$H_N(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}) = -\sum_{\boldsymbol{a}\in\boldsymbol{\mathcal{A}}_Y^N} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \log_2 p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{a}|\boldsymbol{x}) \qquad (215)$$

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

- Since the conditional entropy $H_N(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})$ is always nonnegative, we have

$$\boxed{I_N(\boldsymbol{X};\boldsymbol{Y}) \leq H_N(\boldsymbol{Y})} \qquad (216)$$

  with equality if and only if $\boldsymbol{Y}$ is a deterministic function of $\boldsymbol{X}$

- If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, we have $I_N(\boldsymbol{X};\boldsymbol{Y}) = 0$

# Mutual Information for Coding of Continuous Sources

- Consider mutual information $I_N(\boldsymbol{S};\boldsymbol{S'})$ between a vector of $N$ successive input samples $\boldsymbol{S}$ and the corresponding vector of $N$ reconstructed samples $\boldsymbol{S'}$
- Since vectors of reconstructed samples are discrete, we can write

$$I_N(\boldsymbol{S};\boldsymbol{S'}) \;=\; H_N(\boldsymbol{S'}) - \int_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, H_N(\boldsymbol{S'}|\boldsymbol{S}=\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} \;\leq\; H_N(\boldsymbol{S'}) \qquad (217)$$

  where equality is achieved if and only if the vector $\boldsymbol{S'}$ of reconstructed samples is a deterministic function of the input vector $\boldsymbol{S}$

- Using the fundamental bound for lossless coding, we have for the average rate of a source code $Q$,

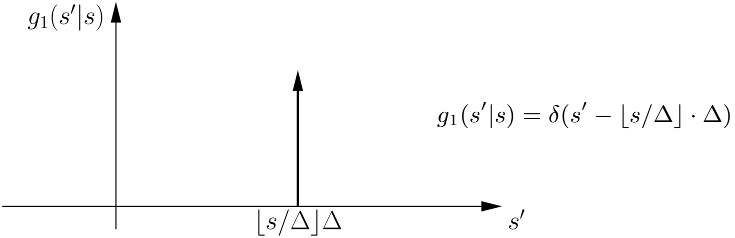
$$\boxed{r(Q) \geq \lim_{N\to\infty} \frac{H_N(\boldsymbol{S'})}{N} \geq \lim_{N\to\infty} \frac{I_N(\boldsymbol{S'};\boldsymbol{S})}{N}} \qquad (218)$$

$\Longrightarrow$ Same expression as for coding of discrete sources

# Description of a Source Code using a Conditional Pdf

- Statistical properties of a mapping $\boldsymbol{s'} = \beta(\alpha(\boldsymbol{s}))$ can be described by an $N$-th order conditional pdf $g_N(\boldsymbol{s'}|\boldsymbol{s})$

- **Example 1**: Mapping $s \rightarrow s' : \; s' = \lfloor s/\Delta \rfloor \cdot \Delta$



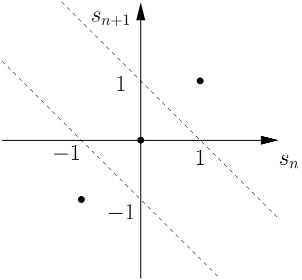

- For $N > 1$, $g_N(\boldsymbol{s'}|\boldsymbol{s})$ are multivariate conditional pdfs

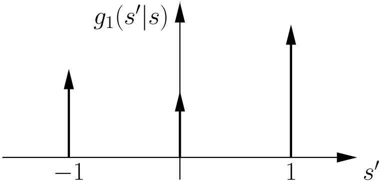
- The pdfs $g_N(\boldsymbol{s'}|\boldsymbol{s})$ obtained by a deterministic mapping (codes) are a subset of the set of all conditional pmfs

# Description of a Source Code using a Conditional Pdf

- **Example 2**: Mapping $(s_n, s_{n+1}) \rightarrow (s'_n, s'_{n+1})$

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




$$g_1(s'|s) = x \cdot \delta(s'+1) + y \cdot \delta(s') + z \cdot \delta(s'-1)$$

with $x + y + z = 1$

# Distortion for a Source Code using Conditional Pdf

- Let $g_N^Q(\boldsymbol{s'}|\boldsymbol{s})$ be the $N$-th order conditional pdf of a source code $Q$ with $\boldsymbol{s'} \in \mathcal{R}^N$ and $\boldsymbol{s} \in \mathcal{R}^N$
- $N$-th order distortion

$$\begin{aligned} \delta_N(g_N) &= E\{d_N(\boldsymbol{S},\boldsymbol{S'})\} \\ &= \int_{\mathcal{R}^N}\int_{\mathcal{R}^N} f_{\boldsymbol{S}\boldsymbol{S'}}(\boldsymbol{s},\boldsymbol{s'}) \cdot d_N(\boldsymbol{s},\boldsymbol{s'})\,\mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{s'} \\ &= \int_{\mathcal{R}^N}\int_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s'}|\boldsymbol{s}) \cdot d_N(\boldsymbol{s},\boldsymbol{s'})\,\mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{s'} \end{aligned} \tag{219}$$

- Recall: General expression for distortion $\delta(Q)$ of a source code $Q$

$$\delta(Q) = \lim_{N\to\infty} E\{d_N(\boldsymbol{S},\boldsymbol{S'})\} \tag{220}$$

- Distortion for a source code $Q$ can be written as

$$\delta(Q) = \lim_{N\to\infty} \delta_N(g_N^Q) \tag{221}$$

# Mutual Information for a Source Code using Conditional Pdf

- $N$-th order mutual information

$$
\begin{aligned}
I_N(g_N) &= E\left\{\log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{S},\boldsymbol{S}')}{f_{\boldsymbol{S}}(\boldsymbol{S}) f_{\boldsymbol{S}'}(\boldsymbol{S}')}\right\} \\
&= \int\limits_{\mathcal{R}^N}\int\limits_{\mathcal{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s},\boldsymbol{s}') \cdot \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{S},\boldsymbol{S}')}{f_{\boldsymbol{S}}(\boldsymbol{S}) f_{\boldsymbol{S}'}(\boldsymbol{S}')} \,\mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{s}' \\
&= \int\limits_{\mathcal{R}^N}\int\limits_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N(\boldsymbol{s}'|\boldsymbol{s}) \cdot \log_2 \frac{g_N(\boldsymbol{s}'|\boldsymbol{s})}{f_{\boldsymbol{S}'}(\boldsymbol{s}')} \,\mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{s}' \qquad (222)
\end{aligned}
$$

with

$$
f_{\boldsymbol{S}'}(\boldsymbol{s}') = \int_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N(\boldsymbol{s}'|\boldsymbol{s}) \,\mathrm{d}\boldsymbol{s}. \qquad (223)
$$

- For a given source $\boldsymbol{S}$, both the $N$-th order distortion $\delta_N$ and the $N$-th order mutual information $I_N$ are completely determined by the $N$-th order conditional pdf $g_N^Q(\boldsymbol{s}'|\boldsymbol{s})$

# Information Rate-Distortion Function

- Consider any source code $Q$ with a distortion $\delta(Q) \leq D$
- Associated rate is denoted by $r(Q)$
- Output $\boldsymbol{S'}$ of source codec is a discrete random process
- Remember: Fundamental theorem for lossless coding

$$r(Q) \geq \bar{H}(\boldsymbol{S'}) = \lim_{N \to \infty} \frac{H_N(\boldsymbol{S'})}{N} \tag{224}$$

- Using mutual information, we can write

$$r(Q) \geq \lim_{N \to \infty} \frac{H_N(\boldsymbol{S'})}{N} \geq \lim_{N \to \infty} \frac{I_N(\boldsymbol{S};\boldsymbol{S'})}{N} = \lim_{N \to \infty} \frac{I_N(g_N^Q)}{N} \tag{225}$$

- Deterministic mapping $g_N^Q$ as given by a source code $Q$ is a special case of a random mapping $g_N$

$$I_N(g_N^Q) \geq \inf_{g_N : \delta_N(g_N) \leq D} I_N(g_N) \tag{226}$$

# Information Rate-Distortion Function

- Hence, we have

$$r(Q) \geq \lim_{N\to\infty} \frac{I_N(g_N^Q)}{N} \geq \lim_{N\to\infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{I_N(g_N)}{N} \qquad (227)$$

- **Information rate-distortion function**

$$\boxed{R^{(I)}(D) = \lim_{N\to\infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{I_N(g_N)}{N}} \qquad (228)$$

- **Fundamental source coding theorem**

$$\boxed{\forall Q : \delta(Q) \leq D, \quad r(Q) \geq R^{(I)}(D)} \qquad (229)$$

$\Longrightarrow$ For a given maximum distortion $D$, the rate $r(Q)$ for each source code $Q$ that yields a distortion $\delta(Q) \leq D$ is greater than or equal to the information rate-distortion function $R^{(I)}(D)$

# Information vs Operational Rate-Distortion Function

- We have shown that information rate-distortion function $R^{(I)}(D)$ represents a lower bound for all source codes $Q$

$\Longrightarrow$ Lower bound for operational rate-distortion function

- It can also be shown that $R^{(I)}(D)$ is asymptotically achievable
  - For any $\epsilon > 0$, there exists a code $Q$ with

$$\delta(Q) \leq D \qquad \text{and}$$
$$r(Q) \leq R^{(I)}(D) + \epsilon$$

(see proof in [Cover and Thomas])

$\Longrightarrow$ **Information rate-distortion function is equal to operational rate-distortion function**

- Use the term **rate-distortion function $R(D)$** for both in the following

# (Information) Distortion-Rate Function

- Fundamental source coding theorem

$$\forall Q : \delta(Q) \leq D, \quad r(Q) \geq R(D) \tag{230}$$

with (information) rate-distortion function

$$R(D) = \lim_{N \to \infty} \inf_{g_N : \delta_N(g_N) \leq D} \frac{I_N(g_N)}{N} \tag{231}$$

- Alternative formulation by interchanging roles of rate and distortion

$$\forall Q : r(Q) \leq R, \quad \delta(Q) \geq D^{(I)}(R) \tag{232}$$

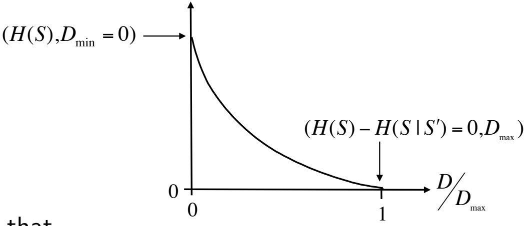
with (information) **distortion-rate function**

$$D(R) = \lim_{N \to \infty} \inf_{g_N : I_N(g_N)/N \leq R} \delta_N(g_N) \tag{233}$$

- Distortion-rate function $D(R)$ is inverse of rate-distortion function $R(D)$

# $R(D)$ 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 $D_{\max}$, so that

$$\forall D \geq D_{\max} \qquad R(D) = 0 \tag{234}$$

$\Longrightarrow$ For MSE distortion measure: $D_{\max}$ is equal to the variance $\sigma^2$ of the source

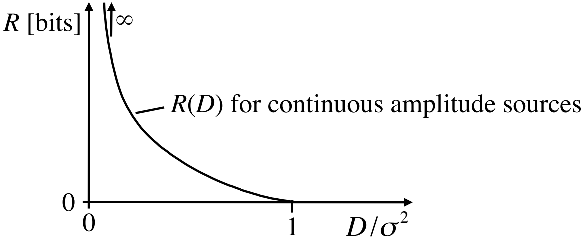
- Minimum rate required for lossless transmission of a discrete source is equal to the entropy rate

$$D_{\min} = 0 \qquad R(0) = \bar{H}(\boldsymbol{S}) \tag{235}$$

$\Longrightarrow$ Fundamental bound for lossless coding:
Special case of the fundamental bound for lossy coding

# $R(D)$ for Continuous Sources and Additive Distortion Meas.

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




- $R(D)$ is a **non-increasing** and **convex** function of $D$
- There exists a value $D_{\max}$, so that

$$\forall D \geq D_{\max} \qquad R(D) = 0 \tag{236}$$

$\Longrightarrow$ For MSE distortion measure: $D_{\max}$ is equal to the variance $\sigma^2$ of the source

- $R(D)$ approaches infinity as $D$ approaches zero

# Rate-Distortion Function for IID Sources

- $N$-th order distortion $\delta_N(g_N)$ for additive distortion measures

$$\delta_N(g_N) \;=\; E\{d_N(\boldsymbol{S},\boldsymbol{S'})\} = E\left\{\frac{1}{N}\sum_{i=0}^{N-1} d_1(S_i,S'_i)\right\} = E\{d_1(S,S')\}$$
$$= \int_{-\infty}^{\infty} f_S(s)\cdot g_1(s'|s)\cdot d_1(s,s')\,\mathrm{d}s = \delta_1(g) \qquad (237)$$

- $N$-th order mutual information for iid sources
(Note: If the source $\boldsymbol{S}$ is iid, the reconstruction $\boldsymbol{S'}$ is also iid)

$$I_N(g_N) \;=\; E\left\{\log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S'}}(\boldsymbol{S},\boldsymbol{S'})}{f_{\boldsymbol{S}}(\boldsymbol{S})\,f_{\boldsymbol{S'}}(\boldsymbol{S'})}\right\} = E\left\{\log_2\left(\frac{f_{SS'}(S,S')}{f_S(S)\,f_{S'}(S')}\right)^N\right\}$$
$$= N\cdot E\left\{\log_2 \frac{f_{SS'}(S,S')}{f_S(S)\,f_{S'}(S')}\right\}$$
$$= N\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_S(s)\,g_1(s'|s)\,\log_2\frac{g_1(s'|s)}{f_{S'}(s')}\,\mathrm{d}s\,\mathrm{d}s' = N\cdot I_1(g) \qquad (238)$$

# Rate-Distortion Function for IID Sources

- For iid sources and additive distortion measures, we have

$$\delta_N(g_N^Q) = \delta_1(g^Q) \qquad \text{and} \qquad I_N(g_N^Q) = N \cdot I_1(g^Q) \tag{239}$$

- Rate-distortion function for iid sources and additive distortion measures

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\le D} \frac{I_N(g_N)}{N} = \inf_{g_1:\, \delta_1(g_1)\le D} I_1(g_1) \tag{240}$$

$\Longrightarrow$ Also called first-order rate-distortion function $R_1(D)$

- Distortion-rate function for iid sources and additive distortion measures

$$D(R) = \lim_{N\to\infty} \inf_{g_N:\, I_N(g_N)/N\le R} \delta_N(g_N) = \inf_{g_1:\, I_1(g_1)\le R} \delta_1(g_1) \tag{241}$$

$\Longrightarrow$ Also called first-order distortion-rate function $D_1(R)$

# $N$-th Order Rate-Distortion Functions

- Can define $N$-th order rate-distortion and distortion-rate functions

$$R_N(D) = \inf_{g_N:\, \delta_N(g_N) \leq D} \frac{I_N(g_N)}{N} \tag{242}$$

$$D_N(R) = \inf_{g_N:\, I_N(g_N)/N \leq R} \delta_N(g_N) \tag{243}$$

- In general, the rate-distortion and distortion-rate functions can be written as

$$\boxed{R(D) = \lim_{N\to\infty} R_N(D)} \quad \text{and} \quad \boxed{D(R) = \lim_{N\to\infty} D_N(R)} \tag{244}$$

- For **iid sources and additive distortion measures**, we have

$$\boxed{R(D) = R_1(D)} \quad \text{and} \quad \boxed{D(R) = D_1(R)} \tag{245}$$

# Discussion of Rate-Distortion Functions

## Operational rate-distortion function

$$R(D) = \inf_{Q:\,\delta(Q)\leq D} r(Q) \tag{246}$$

- Minimization over all possible source codes
- Easy to define, but impossible to evaluate

## Information rate-distortion function

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\delta_N(g_N)\leq D} \frac{I_N(g_N)}{N} \tag{247}$$

- Property of source: Don't need to consider all possible codes
- Still impossible to evaluate directly (minimization over all conditional pdfs)
- Numerical minimization for discrete sources: Blahut-Arimoto algorithm

How can we proceed?

- Can derive lower bound for (information) rate-distortion function
- For some sources and distortion measures (e.g., Gaussian and MSE):
  $\Longrightarrow$ Can show that lower bound is achievable

# Differential Entropy

- Mutual information between a continuous random vector $\boldsymbol{X}$ and a continuous or discrete random vector $\boldsymbol{Y}$

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= E\left\{\log_2 \frac{f_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{X},\boldsymbol{Y})}{f_{\boldsymbol{X}}(\boldsymbol{X})\, f_{\boldsymbol{Y}}(\boldsymbol{Y})}\right\} = E\left\{\log_2 \frac{f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}|\boldsymbol{Y})}{f_{\boldsymbol{X}}(\boldsymbol{X})}\right\} \\ &= E\{-\log_2 f_{\boldsymbol{X}}(\boldsymbol{X})\} - E\left\{-\log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}|\boldsymbol{Y})\right\} \\ &= h(\boldsymbol{X}) - h(\boldsymbol{X}|\boldsymbol{Y}) \end{aligned} \tag{248}$$

- Define: **Differential entropy** of a continuous random vector $\boldsymbol{X}$

$$h(\boldsymbol{X}) = E\{-\log_2 f_{\boldsymbol{X}}(\boldsymbol{X})\} = -\int_{\mathcal{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x})\, \log_2 f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{249}$$

- Define: **Conditional differential entropy** of $\boldsymbol{X}$ given $\boldsymbol{Y}$

$$\begin{aligned} h(\boldsymbol{X}|\boldsymbol{Y}) &= E\left\{-\log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}|\boldsymbol{Y})\right\} \\ &= -\int_{\mathcal{R}^N}\int_{\mathcal{R}^N} f_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})\, \log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y} \end{aligned} \tag{250}$$

# Example: Differential Entropy for an Uniform IID Source

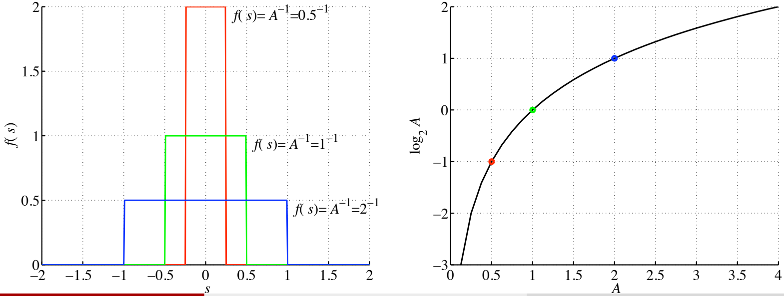
- For an continuous iid source $S$, differential entropy is defined as

$$h(S) = E\{-\log_2 f(S)\} = -\int_{-\infty}^{\infty} f(s)\log_2 f(s)\,\mathrm{d}s \qquad (251)$$

- $h(S)$ for uniform distribution $f(s) = 1/A$ for $-A/2 \le s \le A/2$

$$h(S) = -\int_{-A/2}^{A/2} \frac{1}{A}\log_2 \frac{1}{A}\mathrm{d}s = \frac{1}{A}\log_2 A \int_{-A/2}^{A/2}\mathrm{d}s = \log_2 A \qquad (252)$$

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



# Differential Entropy for an Gaussian IID Source

- Gaussian iid process

$$f_S(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}} \tag{253}$$

- Differential entropy

$$\begin{aligned} h(S) &= -\int_{-\infty}^{\infty} f_S(s)\log_2 f_S(s)\mathrm{d}s \\ &= -\int_{-\infty}^{\infty} f_S(s)\left[-\frac{(s-\mu)^2}{2\sigma^2}\log_2 e - \log_2\sqrt{2\pi\sigma^2}\right]\mathrm{d}s \\ &= \frac{E\{(S-\mu)^2\}}{2\sigma^2}\cdot\log_2 e + \frac{1}{2}\log_2(2\pi\sigma^2) \\ &= \frac{1}{2}\log_2 e + \frac{1}{2}\log_2(2\pi\sigma^2) \\ &= \frac{1}{2}\log_2(2\pi e\sigma^2) \end{aligned} \tag{254}$$

# $N$-th Order Differential Entropy

- **$N$-th order differential entropy**

$$h_N(\boldsymbol{S}) = h(\boldsymbol{S}^{(N)}) = h(S_0, \cdots, S_{N-1}) = E\Big\{ -\log_2 f_{\boldsymbol{S}}(\boldsymbol{S}^{(N)}) \Big\} \qquad (255)$$

- **Differential entropy rate**

$$\bar{h}(\boldsymbol{S}) = \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} = \lim_{N\to\infty} \frac{h(S_0, \cdots, S_{N-1})}{N} \qquad (256)$$

- $N$-th order pdf of a stationary Gaussian process

$$f_G(\boldsymbol{s}) = \frac{1}{(2\pi)^{N/2}\,|\boldsymbol{C}_N|^{1/2}}\; e^{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)} \qquad (257)$$

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

$$\begin{aligned} h_N^{(G)}(\boldsymbol{S}) &= -\int_{\mathcal{R}^N} f_G(\boldsymbol{s})\, \log_2 f_G(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s} \\ &= \frac{1}{2}\log_2\big(\,(2\pi)^N |\boldsymbol{C}_N|\,\big) \\ &\quad + \frac{\log_2 e}{2} \int_{\mathcal{R}^N} f_G(\boldsymbol{s})\; (\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)\, \mathrm{d}\boldsymbol{s} \qquad (258) \end{aligned}$$

# $N$-th order Differential Entropy of Stationary Gaussian Process

- General stationary process with pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$, mean $\boldsymbol{\mu}_N$, covariance matrix $\boldsymbol{C}_N$

$$
\begin{aligned}
\int_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\,(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)\,\mathrm{d}\boldsymbol{s} & \\
&= E\big\{(\boldsymbol{S}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{S}-\boldsymbol{\mu}_N)\big\} \\
&= E\left\{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(S_i-\mu_i)(C^{-1})_{i,j}(S_j-\mu_j)\right\} \\
&= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} E\{(S_i-\mu_i)(S_j-\mu_j)\}\,(C^{-1})_{i,j} \\
&= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} C_{j,i}(C^{-1})_{i,j} \\
&= \sum_{i=0}^{N-1} (CC^{-1})_{j,j} \\
&= N \qquad (259)
\end{aligned}
$$

# $N$-th order Differential Entropy of Stationary Gaussian Process

- Showed for general pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$

$$\int_{\mathcal{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\,(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)\,\mathrm{d}\boldsymbol{s} = N \tag{260}$$

- Continue derivation for stationary Gaussian source

$$\begin{aligned} h_N^{(G)}(\boldsymbol{S}) &= \frac{1}{2}\log_2\big((2\pi)^N|\boldsymbol{C}_N|\big) \\ &\qquad + \frac{\log_2 e}{2}\int_{\mathcal{R}^N} f_G(\boldsymbol{s})\,(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)\,\mathrm{d}\boldsymbol{s} \\ &= \frac{1}{2}\log_2\big((2\pi)^N|\boldsymbol{C}_N|\big) + \frac{N}{2}\log_2 e \\ &= \frac{1}{2}\log_2\big((2\pi)^N|\boldsymbol{C}_N|\big) + \frac{1}{2}\log_2 e^N \\ &= \frac{1}{2}\log_2\big((2\pi e)^N|\boldsymbol{C}_N|\big) \end{aligned} \tag{261}$$

# $N$-th order Differential Entropy of Stat. Non-Gaussian Process

- Consider stationary non-Gaussian process with $N$-th order pdf $f(\boldsymbol{s})$
- Let $f_G(\boldsymbol{s})$ be the $N$-th order pdf of a Gaussian process with the same $N$-th order autocovariance matrix $\boldsymbol{C}_N$
- By applying the divergence inequality for pdfs, we obtain

$$
\begin{aligned}
h_N(\boldsymbol{S}) &= -\int_{\mathcal{R}^N} f(\boldsymbol{s}) \, \log_2 f(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} \\
&\leq -\int_{\mathcal{R}^N} f(\boldsymbol{s}) \, \log_2 f_G(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} \\
&= \frac{1}{2} \log_2\!\left( (2\pi)^N |\boldsymbol{C}_N| \right) + \frac{\log_2 e}{2} \int_{\mathcal{R}^N} f(\boldsymbol{s})(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1} (\boldsymbol{s}-\boldsymbol{\mu}_N) \, \mathrm{d}\boldsymbol{s} \\
&= \frac{1}{2} \log_2\!\left( (2\pi e)^N |\boldsymbol{C}_N| \right) = h_N^{(G)}(\boldsymbol{S}) \qquad (262)
\end{aligned}
$$

$\Longrightarrow$ **Gaussian process with a given $N$-th order autocovariance matrix $C_N$ has the largest $N$-th order differential entropy among all processes with the same autocovariance matrix $C_N$**

# Eigendecomposition of the Covariance Matrix

- Determinant $|\boldsymbol{C}_N|$: Product of the eigenvalues $\xi_i$ of the matrix $\boldsymbol{C}_N$,

$$\boldsymbol{C}_N = \boldsymbol{A}_N \boldsymbol{\Xi}_N \boldsymbol{A}_N^T \quad \rightarrow \quad |\boldsymbol{C}_N| = \underbrace{|\boldsymbol{A}_N|}_{=1} \cdot |\boldsymbol{\Xi}_N| \cdot \underbrace{|\boldsymbol{A}_N^T|}_{=1} = \prod_{i=0}^{N-1} \xi_i^{(N)} \qquad (263)$$

- $\boldsymbol{A}_N$: Orthogonal matrix with the $N$ unit-norm eigenvectors as columns

$$\boldsymbol{A}_N = \left( \boldsymbol{v}_0^{(N)}, \boldsymbol{v}_1^{(N)}, \cdots, \boldsymbol{v}_{N-1}^{(N)} \right) \qquad (264)$$

- $\boldsymbol{\Xi}_N$: Diagonal matrix with the $N$ eigenvalues of $\boldsymbol{C}_N$ on its main diagonal

$$\boldsymbol{\Xi}_N = \begin{pmatrix} \xi_0^{(N)} & 0 & \dots & 0 \\ 0 & \xi_1^{(N)} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \xi_{N-1}^{(N)} \end{pmatrix} \qquad (265)$$

# Maximum Differential Entropy

- Determinant of a matrix is the product of its eigenvalues

$$|\boldsymbol{C}_N| = \prod_{i=0}^{N-1} \xi_i^{(N)} \tag{266}$$

- Trace of a matrix is the sum of its eigenvalues (trace is similarity invariant)

$$\mathrm{tr}\left(|\boldsymbol{C}_N|\right) = \sum_{i=0}^{N-1} \xi_i^{(N)} = N \cdot \sigma^2 \tag{267}$$

- Inequality of arithmetic and geometric means:

$$\left(\prod_{i=0}^{N-1} x_i\right)^{\frac{1}{N}} \leq \frac{1}{N} \sum_{i=0}^{N-1} x_i, \tag{268}$$

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

# Maximum Differential Entropy

- Apply inequality to determinant of autocovariance matrix

$$|\boldsymbol{C}_N| = \prod_{i=0}^{N-1} \xi_i \leq \left( \frac{1}{N} \sum_{i=0}^{N-1} \xi_i \right)^N = \sigma^{2N} \quad (269)$$

$\Longrightarrow$ Equality if and only if source is iid (all eigenvalues are the same)

- For $N$-th order differential entropy of any source $\boldsymbol{S}$, we get

$$\begin{aligned} h_N(\boldsymbol{S}) &\leq \frac{1}{2} \log_2 \left( (2\pi e)^N \, |\boldsymbol{C}_N| \right) && \text{(equality for Gaussian)} \\ &\leq \frac{N}{2} \log_2 \left( 2\pi e \sigma^2 \right) && \text{(equality for iid)} \end{aligned} \quad (270)$$

$\Longrightarrow$ Equality if and only if source is Gaussian iid

$\Longrightarrow$ **For a given variance $\sigma^2$, the $N$-th order differential entropy is maximized for Gaussian iid processes**

$$\boxed{h_N(\boldsymbol{S}) \leq \frac{N}{2} \log_2 \left( 2\pi e \sigma^2 \right)} \quad (271)$$

# Shannon Lower Bound

- Lower bound for rate-distortion function $R(D)$

$$
\begin{aligned}
R(D) &= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\boldsymbol{S};\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}) - h_N(\boldsymbol{S}|\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}|\boldsymbol{S}')}{N} \\
&= \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}'|\boldsymbol{S}')}{N} \qquad (272)
\end{aligned}
$$

- Define: **Shannon Lower Bound**

$$
\boxed{R_L(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}} \qquad (273)
$$

- Since conditioning does not increase differential entropy, we have

$$
\boxed{R(D) \ge R_L(D)} \qquad \text{(equality if } \boldsymbol{S}-\boldsymbol{S}' \text{ is independent of } \boldsymbol{S}'\text{)} \qquad (274)
$$

# Shannon Lower Bound for MSE Distortion

- For MSE distortion: Distortion is given by variance of differences

$$\delta_N(g_N) = \sigma_Z^2 \qquad \text{with} \qquad \boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S'} \qquad \text{and} \qquad \mu_Z = 0 \tag{275}$$

- Remember: Maximum differential entropy

$$h_N(\boldsymbol{S} - \boldsymbol{S'}) = h_N(\boldsymbol{Z}) \leq \frac{N}{2} \log_2\left(2\pi e \sigma_Z^2\right) = \frac{N}{2} \log_2\left(2\pi e D\right) \tag{276}$$

- **Shannon lower bound for MSE distortion**

$$\boxed{R_L(D) = \bar{h}(\boldsymbol{S}) - \frac{1}{2} \log_2\left(2\pi e D\right)} \tag{277}$$

$\Longrightarrow$ **For given $C_N$ or $\Phi_{SS}(\omega)$, maximized for Gaussian processes**
$\Longrightarrow$ **For given $\sigma^2$, maximized for Gaussian iid processes**

- When is the Shannon lower bound for MSE achievable?
  $\Longrightarrow$ Difference process $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S'}$ has to be zero-mean Gaussian iid
  $\Longrightarrow$ Difference process $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S'}$ has to be independent of $\boldsymbol{S'}$:

$$g_{\boldsymbol{Z}|\boldsymbol{S'}}(\boldsymbol{z}|\boldsymbol{s'}) = g_{\boldsymbol{Z}}(\boldsymbol{z})$$

# Shannon Lower Bound for IID Sources MSE Distortion

- **Shannon lower bound for MSE distortion**

$$\boxed{R_L(D) = \bar{h}(\boldsymbol{S}) - \frac{1}{2}\log_2\big(\,2\pi e D\,\big)} \qquad \boxed{D_L(R) = \frac{1}{2\pi e}\cdot 2^{\,2\,\bar{h}(\boldsymbol{S})}\cdot 2^{-2R}} \qquad (278)$$

- For iid sources $\boldsymbol{S}$, we have

$$\begin{aligned}
\bar{h}(\boldsymbol{S}) &= \lim_{N\to\infty}\frac{h_N(\boldsymbol{S})}{N} = \lim_{N\to\infty}\frac{1}{N}E\{-\log_2 f_{\boldsymbol{S}}(\boldsymbol{S})\} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}E\{-\log_2 f_S(S_i)\} = \lim_{N\to\infty}\frac{N}{N}E\{-\log_2 f_S(S)\} \\
&= E\{-\log_2 f_S(S)\} = h(S) \qquad (279)
\end{aligned}$$

- **Shannon lower bound for MSE distortion and iid sources**

$$\boxed{R_L(D) = h(S) - \frac{1}{2}\log_2\big(\,2\pi e D\,\big)} \qquad \boxed{D_L(R) = \frac{1}{2\pi e}\cdot 2^{\,2\,h(S)}\cdot 2^{-2R}} \qquad (280)$$

# Shannon Lower Bound Selected IID Sources

- **Uniform pdf**:

$$h(S) = \frac{1}{2}\log_2(12\sigma^2) \quad \Longrightarrow \quad D_L(R) = \underbrace{\frac{6}{\pi e}}_{\approx 0.7} \sigma^2 \cdot 2^{-2R} \tag{281}$$

- **Laplacian pdf**:

$$h(S) = \frac{1}{2}\log_2(2e^2\sigma^2) \quad \Longrightarrow \quad D_L(R) = \underbrace{\frac{e}{\pi}}_{\approx 0.865} \sigma^2 \cdot 2^{-2R} \tag{282}$$

- **Gaussian pdf**:

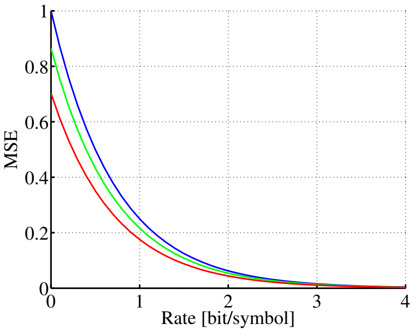
$$h(S) = \frac{1}{2}\log_2(2\pi e\sigma^2) \quad \Longrightarrow \quad D_L(R) = \sigma^2 \cdot 2^{-2R} \tag{283}$$

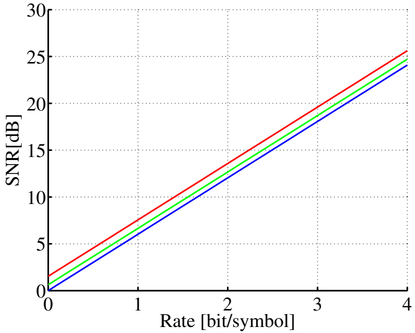
# Shannon Lower Bound Selected IID Sources

Shannon lower bound using MSE and SNR

$$\mathrm{SNR} = 10 \log_{10} \frac{\sigma^2}{\mathrm{MSE}} \tag{284}$$

- Uniform iid process: red
- Laplace iid process: green
- Gauss iid process: blue



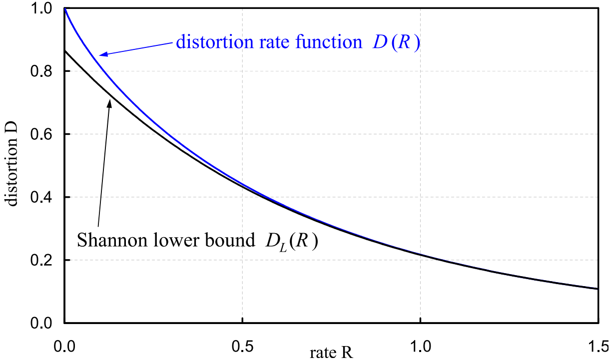



# Asymptotic Tightness of the Shannon Lower Bound

- Shannon lower bound approaches distortion rate function for small distortions or high rates

$$\lim_{D \to 0} R(D) - R_L(D) = 0. \qquad (285)$$

- Comparison of $D(R)$ with $D_L(R)$ for the Laplacian iid source



# Shannon Lower Bound for Gaussian Sources with Memory

- Differential entropy for Gaussian sources

$$h_N^{(G)}(\boldsymbol{S}) = \frac{1}{2}\log_2\big(\,(2\pi e)^N |\boldsymbol{C}_N|\,\big) \tag{286}$$

- Shannon lower bound for MSE distortion

$$\begin{aligned} R_L(D) &= \lim_{N\to\infty} \frac{h_N^{(G)}(\boldsymbol{S})}{N} - \frac{1}{2}\log_2\big(\,2\pi e D\,\big) \\ &= \lim_{N\to\infty} \frac{\log_2((2\pi e)^N |\boldsymbol{C}_N|)}{2N} - \frac{1}{2}\log_2\big(\,2\pi e D\,\big) \\ &= \frac{1}{2}\log_2(2\pi e) + \lim_{N\to\infty} \frac{\log_2(|\boldsymbol{C}_N|)}{2N} - \frac{1}{2}\log_2\big(\,2\pi e D\,\big) \\ &= \lim_{N\to\infty} \frac{\log_2 |\boldsymbol{C}_N|}{2N} - \frac{1}{2}\log_2 D \\ &= \lim_{N\to\infty} \frac{1}{2N}\sum_{i=0}^{N-1} \log_2 \xi_i^{(N)} - \frac{1}{2}\log_2 D \end{aligned} \tag{287}$$

# GRENANDER and SZEGÖ's theorem

- Assume zero-mean process: $\boldsymbol{C}_N = \boldsymbol{R}_N$
- Given the conditions
  - $\boldsymbol{R}_N$ is a sequence of Hermitian Toeplitz matrices with elements $\phi_k$ on the $k$-th diagonal
  - The infimum $\Phi_{\inf} = \inf_\omega \Phi(\omega)$ and supremum $\Phi_{\sup} = \sup_\omega \Phi(\omega)$ of the Fourier series are finite

$$\Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \, e^{-j\omega k} \tag{288}$$

  - The function $G$ is continuous in the interval $[\Phi_{\inf}, \Phi_{\sup}]$
- The following expression holds

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} G\left(\xi_i^{(N)}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G\left(\Phi(\omega)\right) \mathrm{d}\omega \tag{289}$$

where $\xi_i^{(N)}$, for $i = 0, 1, \ldots, N-1$, denote the eigenvalues of the $N$-th matrix $\boldsymbol{R}_N$

# Shannon Lower Bound for Gaussian Sources with Memory

- We have already derived

$$R_L(D) = \lim_{N\to\infty} \frac{1}{2N} \sum_{i=0}^{N-1} \log_2 \xi_i^{(N)} - \frac{1}{2} \log_2 D \tag{290}$$

- Applying GRENANDER and SZEGÖ's theorem

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} G\left(\xi_i^{(N)}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G\left(\Phi(\omega)\right) \, \mathrm{d}\omega \tag{291}$$

yields

$$\begin{aligned} R_L(D) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \Phi_{SS}(\omega) \, \mathrm{d}\omega - \frac{1}{2} \log_2 D \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \Phi_{SS}(\omega) \, \mathrm{d}\omega - \frac{1}{4\pi} \log_2 D \int_{-\pi}^{\pi} \mathrm{d}\omega \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \frac{\Phi_{SS}(\omega)}{D} \, \mathrm{d}\omega \end{aligned} \tag{292}$$

# Power Spectral Density of a Gauss-Markov Process

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

$$S_n = Z_n + \rho \cdot S_{n-1} \tag{293}$$

- Auto-correlation function

$$\phi[k] = \sigma^2 \cdot \rho^{|k|} \tag{294}$$

- Using the relationship

$$\sum_{k=1}^{\infty} a^k \, e^{-jkx} = \frac{a}{e^{-jx} - a} \tag{295}$$

we obtain

$$\begin{aligned} \Phi_{SS}(\omega) &= \sum_{k=-\infty}^{\infty} \phi[k] \cdot e^{-j\omega k} \\ &= \sum_{k=-\infty}^{\infty} \sigma^2 \cdot \rho^{|k|} \cdot e^{-j\omega k} \\ &= \sigma^2 \cdot \left( 1 + \frac{\rho}{e^{-j\omega} - \rho} + \frac{\rho}{e^{j\omega} - \rho} \right) \\ &= \sigma^2 \cdot \frac{1 - \rho^2}{1 - 2\rho\cos\omega + \rho^2} \end{aligned} \tag{296}$$

# Shannon Lower Bound for Gaussian-Markov Processes

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

$$\begin{aligned} R_L(D) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_2 \frac{\Phi_{SS}(\omega)}{D}\,\mathrm{d}\omega \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \log_2 \frac{\sigma^2(1-\rho^2)}{D}\,\mathrm{d}\omega - \\ &\qquad \underbrace{\frac{1}{4\pi}\int_{-\pi}^{\pi} \log_2(1 - 2\rho\cos\omega + \rho^2)\,\mathrm{d}\omega}_{=0} \\ R_L(D) &= \frac{1}{2}\log_2 \frac{\sigma^2\,(1-\rho^2)}{D} \end{aligned} \tag{297}$$

where we used

$$\int_0^{\pi} \ln(a^2 - 2ab\cos x + b^2)\,\mathrm{d}x = 2\pi \ln a \tag{298}$$

- Shannon lower bound as distortion-rate function

$$\boxed{D_L(R) = (1-\rho^2)\,\sigma^2\,2^{-2R}} \tag{299}$$

# Rate-Distortion Function for Gaussian IID Sources

- Consider Gaussian iid source

$$f_S(s) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}} \qquad (300)$$

- Shannon lower bound for Gaussian iid sources

$$\boxed{D_L(R) = \sigma^2 \cdot 2^{-2R}} \quad \Longleftrightarrow \quad \boxed{R_L(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \; D \leq \sigma^2 \\ 0 & : \; D > \sigma^2 \end{cases}} \qquad (301)$$

- For Gaussian iid sources: Rate-distortion function = Shannon lower bound

- How can we proof it?
  - Could show that Shannon lower bound is achievable

$\Longrightarrow$ Need to find $g_{S'|S}(s'|s)$ for which the Shannon lower bound is achieved

- Remember: Discussed that Shannon lower bound is achievable if
  - Difference signal $Z = S - S'$ is independent of $S'$
  - Difference signal $Z = S - S'$ has a zero-mean Gaussian distribution

# Rate-Distortion Function for Gaussian IID Sources

Consider conditional pdf $g_{Z|S'}(z|s') = g_{S-S'|S'}(s-s'|s')$ instead of $g_{S'|S}(s'|s)$

- Given $g_{Z|S'}(z|s')$, conditional pdf $g_{S'|S}(s'|s)$ can be derived by

$$g_{S'|S}(s'|s) = g_{S|S'}(s|s') \cdot \frac{f_{S'}(s')}{f_S(s)} \qquad \text{with} \qquad g_{S|S'}(s|s') = g_{Z|S'}(z+s'|s') \tag{302}$$

- Shannon lower bound coincides with rate-distortion function, only if the difference signal $Z = S - S'$ fulfills the conditions:
  - Difference signal $Z = S - S'$ is independent of $S'$
  - Difference signal $Z = S - S'$ has a zero-mean Gaussian distribution

- Hence, $g_{Z|S'}(z|s')$ has to have the form

$$g_{Z|S'}(z|s') = \frac{1}{\sqrt{2\pi\sigma_Z^2}}\, e^{-\frac{z^2}{2\sigma_Z^2}} = \frac{1}{\sqrt{2\pi D}}\, e^{-\frac{z^2}{2D}} = f_Z(z) \tag{303}$$

- Need to verify that this is a valid choice!

# Rate-Distortion Function for Gaussian IID Sources

- Question: Is the conditional pdf $g_{Z|S'}(z|s')$ a valid choice?

$$g_{Z|S'}(z|s') = f_Z(z) = \frac{1}{\sqrt{2\pi D}}\, e^{-\frac{z^2}{2D}} \tag{304}$$

- Source $S$ is the sum of two independent random variables $Z = S - S'$ and $S'$
- Hence, $f_S(s)$ is given by the convolution

$$f_S(s) = f_Z(z) * f_{S'}(s') \tag{305}$$

- Note: Convolution of two Gaussians $f(\mu_1, \sigma_1^2)$ and $f(\mu_2, \sigma_2^2)$ is a Gaussian with $\mu = \mu_1 + \mu_2$ and $\sigma = \sigma_1^2 + \sigma_2^2$
- Hence, the pdf of the reconstructed samples is

$$f_{S'}(s') = \frac{1}{\sqrt{2\pi\,(\sigma^2 - D)}}\, e^{-\frac{(s'-\mu)^2}{2(\sigma^2 - D)}} \tag{306}$$

- This is a valid pdf for $S'$ (no negative values)

$\Longrightarrow$ Our choice for $g_{Z|S'}(z|s')$ is valid

# Rate-Distortion Function for Gaussian IID Sources

Check distortion and rate (mutual information)

- Distortion given by variance of difference process $Z = S - S'$

$$\delta(g) = E\left\{(S - S')^2\right\} = E\left\{Z^2\right\} = D \qquad (307)$$

- Mutual information

$$\begin{aligned} I(g) &= h(S) - h(S|S') \\ &= h(S) - h(S - S'|S') \\ &= h(S) - h(Z|S') \\ &= h(S) - h(Z) \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) - \frac{1}{2}\log_2\left(2\pi e D\right) \\ &= R(D) = \frac{1}{2}\log_2\frac{\sigma^2}{D} \end{aligned} \qquad (308)$$

$\Longrightarrow$ **For Gaussian iid processes and MSE distortion, the rate-distortion function coincides with the Shannon lower bound**

# Rate-Distortion Function for Gaussian IID Sources

- Considered Gaussian iid source with a variance $\sigma^2$ and MSE distortion
- Shannon lower bound coincides with the rate-distortion function
- The rate-distortion function $R(D)$ is given by

$$\boxed{R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases}} \tag{309}$$

- The distortion-rate function is given as

$$\boxed{D(R) = \sigma^2 \cdot 2^{-2R}} \tag{310}$$

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

$$\boxed{\mathrm{SNR}(R) = 10 \cdot \log_{10} \frac{\sigma^2}{D(R)} = 10 \cdot \log_{10} 2^{2R} \approx 6R \quad [\mathrm{dB}]} \tag{311}$$

- **For MSE distortion and a given variance $\sigma^2$, the rate-distortion function $R(D)$ is maximized for Gaussian iid processes**

$\Longrightarrow$ **Gaussian iid processes are the hardest to code**

# Rate-Distortion Function for Gaussian Sources with Memory

- $N$-th order pdf of stationary Gaussian random process

$$f_{\boldsymbol{S}}^{(G)}(\boldsymbol{s}) = \frac{1}{(2\pi)^{N/2}\,|\boldsymbol{C}_N|^{1/2}}\; e^{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu}_N)^T \boldsymbol{C}_N^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_N)} \tag{312}$$

- Eigendecomposition of covariance matrix $\boldsymbol{C}_N$,

$$\boldsymbol{C}_N = \boldsymbol{A}_N \cdot \boldsymbol{\Xi}_N \cdot \boldsymbol{A}_N^T \tag{313}$$

- $\boldsymbol{A}_N$: Matrix with columns are equal to the $N$ unit-norm eigenvectors

$$\boldsymbol{A}_N = \left( \boldsymbol{v}_0^{(N)}, \boldsymbol{v}_1^{(N)}, \cdots, \boldsymbol{v}_{N-1}^{(N)} \right) \tag{314}$$

- $\boldsymbol{\Xi}_N$: Diagonal matrix with eigenvalues of $\boldsymbol{C}_N$ on its main diagonal

$$\boldsymbol{\Xi}_N = \begin{pmatrix} \xi_0^{(N)} & 0 & \dots & 0 \\ 0 & \xi_1^{(N)} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \xi_{N-1}^{(N)} \end{pmatrix} \tag{315}$$

# Signal Space Rotation

- Given stationary Gaussian source $\{S_n\}$: Construct source $\{U_n\}$ by decomposing $\{S_n\}$ into vectors $\boldsymbol{S}$ of size $N$ and applying the transform

$$\boldsymbol{U} = \boldsymbol{A}_N^{-1} \left( \boldsymbol{S} - \boldsymbol{\mu}_N \right) = \boldsymbol{A}_N^T \left( \boldsymbol{S} - \boldsymbol{\mu}_N \right) \qquad (316)$$

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

$$f_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{(2\pi)^{N/2} \left| \boldsymbol{\Xi}_N \right|^{1/2}} \, e^{-\frac{1}{2} \boldsymbol{u}^T \boldsymbol{\Xi}_N^{-1} \boldsymbol{u}} = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi \xi_i^{(N)}}} \, e^{-\frac{u_i^2}{2\,\xi_i^{(N)}}} \qquad (317)$$

- Mean

$$E\{\boldsymbol{U}\} = \boldsymbol{A}_N^T \left( E\{\boldsymbol{S}\} - \boldsymbol{\mu}_N \right) = \boldsymbol{A}_N^T \left( \boldsymbol{\mu}_N - \boldsymbol{\mu}_N \right) = \boldsymbol{0} \qquad (318)$$

- Covariance

$$\begin{aligned} E\left\{ \boldsymbol{U}\,\boldsymbol{U}^T \right\} &= \boldsymbol{A}_N^T \, E\left\{ \left( \boldsymbol{S} - \boldsymbol{\mu}_N \right) \left( \boldsymbol{S} - \boldsymbol{\mu}_N \right)^T \right\} \boldsymbol{A}_N \\ &= \boldsymbol{A}_N^T \, \boldsymbol{C}_N \, \boldsymbol{A}_N = \boldsymbol{\Xi}_N \end{aligned} \qquad (319)$$

# Distortion and Mutual Information

- Inverse transform after compression identical to forward transform

$$\boldsymbol{S'} = \boldsymbol{A}_N \, \boldsymbol{U'} + \boldsymbol{\mu}_N, \tag{320}$$

- With

$$\big(\boldsymbol{U'} - \boldsymbol{U}\big) = \boldsymbol{A}_N^T\big(\boldsymbol{S'} - \boldsymbol{S}\big) \quad \Longleftrightarrow \quad \big(\boldsymbol{S'} - \boldsymbol{S}\big) = \boldsymbol{A}_N\big(\boldsymbol{U'} - \boldsymbol{U}\big) \tag{321}$$

- MSE distortion between any realization $\boldsymbol{s}$ of $\boldsymbol{S}$ and its reconstruction $\boldsymbol{s'}$

$$\begin{aligned} d_N(\boldsymbol{s};\boldsymbol{s'}) &= \frac{1}{N}\sum_{i=0}^{N-1}(s_i - s'_i)^2 = \frac{1}{N}(\boldsymbol{s}-\boldsymbol{s'})^T(\boldsymbol{s}-\boldsymbol{s'}) \\ &= \frac{1}{N}(\boldsymbol{u}-\boldsymbol{u'})^T\boldsymbol{A}_N^T\boldsymbol{A}_N(\boldsymbol{u}-\boldsymbol{u'}) = \frac{1}{N}(\boldsymbol{u}-\boldsymbol{u'})^T(\boldsymbol{u}-\boldsymbol{u'}) \\ &= \frac{1}{N}\sum_{i=0}^{N-1}(u_i - u'_i)^2 = d_N(\boldsymbol{u};\boldsymbol{u'}) \end{aligned} \tag{322}$$

- Since coordinate transform is invertible,

$$I_N(\boldsymbol{S};\boldsymbol{S'}) = I_N(\boldsymbol{U};\boldsymbol{U'}) \tag{323}$$

# Distortion-Rate Function

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

$$I_N(g_N^Q) = \sum_{i=0}^{N-1} I_1(g_i^Q) \qquad \text{and} \qquad \delta_N(g_N^Q) = \frac{1}{N}\sum_{i=0}^{N-1} \delta_1(g_i^Q) \tag{324}$$

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

$$D_N(R) = \frac{1}{N}\sum_{i=0}^{N-1} D_i(R_i) \qquad \text{with} \qquad R = \frac{1}{N}\sum_{i=0}^{N-1} R_i \tag{325}$$

- $D_i(R_i)$: Distortion-rate function for Gaussian iid processes for component $U_i$

$$D_i(R_i) = \sigma_i^2\, 2^{-2R_i} = \xi_i^{(N)}\, 2^{-2R_i} \tag{326}$$

with $\xi_i^{(N)}$ being the eigenvalues of $\boldsymbol{C}_N$

# Optimal Bit Allocation

- Have to distribute the bit rate in an optimal way

$$\min_{R_0,R_1,\ldots,R_{N-1}} D_N(R) = \frac{1}{N}\sum_{i=0}^{N-1} \xi_i^{(N)}\, 2^{-2R_i} \qquad \text{such that} \qquad R \geq \frac{1}{N}\sum_{i=0}^{N-1} R_i$$

- Comparison on different types of mean computations

$$D_N(R) = \frac{1}{N}\sum_{i=0}^{N-1} \xi_i^{(N)}\, 2^{-2R_i} \geq \left(\prod_{i=0}^{N-1} \xi_i^{(N)}\, 2^{-2R_i}\right)^{\frac{1}{N}} = \underbrace{\left(\prod_{i=0}^{N-1} \xi_i^{(N)}\right)^{\frac{1}{N}}}_{=|\boldsymbol{C}_N|^{\frac{1}{N}}=\tilde{\xi}^{(N)}} \cdot 2^{-2R}$$

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

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

$$R_i = R + \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\tilde{\xi}^{(N)}} = \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\tilde{\xi}^{(N)} 2^{-2R}} \qquad \text{with} \qquad \tilde{\xi}^{(N)} = \left(\prod_{i=0}^{N-1} \xi_i^{(N)}\right)^{\frac{1}{N}}$$

# Condition for Partial Bit Rates

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

$$R_i = \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\tilde{\xi}^{(N)} 2^{-2R}} \geq 0 \quad \Longrightarrow \quad R_i = 0 \quad \text{if} \quad \xi_i^{(N)} \leq \tilde{\xi}^{(N)} 2^{-2R} \tag{327}$$

- Introducing the parameter $\theta$, with $0 \leq \theta \leq D$, yields

$$R_i = \begin{cases} \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\theta} & : \ \theta \leq \xi_i^{(N)} \\ 0 & : \ \theta > \xi_i^{(N)} \end{cases} \tag{328}$$

and

$$D_i = \begin{cases} \theta & : \ \theta \leq \xi_i^{(N)} \\ \xi_i^{(N)} & : \ \theta > \xi_i^{(N)} \end{cases} \tag{329}$$

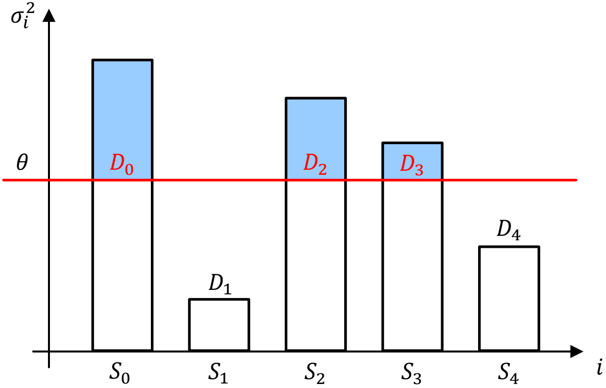
- Can also be written as

$$R_i(\theta) = \max\left(0, \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\theta}\right) \quad \text{and} \quad D_i(\theta) = \min\left(\xi_i^{(N)}, \theta\right) \tag{330}$$

- This rate allocation concept is also referred to as **reverse water filling**

# Reverse Water Filling for Independents Gaussian RV

$$D_i = \min\left(\sigma_i^2, \theta\right)$$




- Optimal rate allocation for independent Gaussian RV and MSE distortion
- Code random variable with $\sigma_i^2 > \theta$ so that the same distortion is obtained
- Do not assign any rate to random variables with $\sigma_i^2 \leq \theta$

# $N$-th Order Rate-Distortion Function

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

$$D_N(R) = \frac{1}{N} \sum_{i=0}^{N-1} D_i(R_i) \qquad \text{with} \qquad R = \frac{1}{N} \sum_{i=0}^{N-1} R_i \tag{331}$$

- Optimal rate allocation

$$R_i(\theta) = \max\left(0, \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\theta}\right) \quad \text{and} \quad D_i(\theta) = \min\left(\xi_i^{(N)}, \theta\right) \tag{332}$$

- Parametric expressions for $N$-th order rate-distortion function

$$D_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} D_i = \frac{1}{N} \sum_{i=0}^{N-1} \min\left(\xi_i^{(N)}, \theta\right) \tag{333}$$

$$R_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} R_i = \frac{1}{N} \sum_{i=0}^{N-1} \max\left(0, \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\theta}\right) \tag{334}$$

# Parametric Rate-Distortion Function

- Rate-distortion function is given by limit for $N \to \infty$

$$D(\theta) = \lim_{N\to\infty} D_N(\theta) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \min\left(\xi_i^{(N)}, \theta\right) \quad (335)$$

$$R(\theta) = \lim_{N\to\infty} R_N(\theta) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \max\left(0, \frac{1}{2}\log_2 \frac{\xi_i^{(N)}}{\theta}\right) \quad (336)$$

- Recall: Grenander and Szegös theorem for infinite Toeplitz matrices

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} G(\xi_i^{(N)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Phi(\omega)) d\omega \quad (337)$$

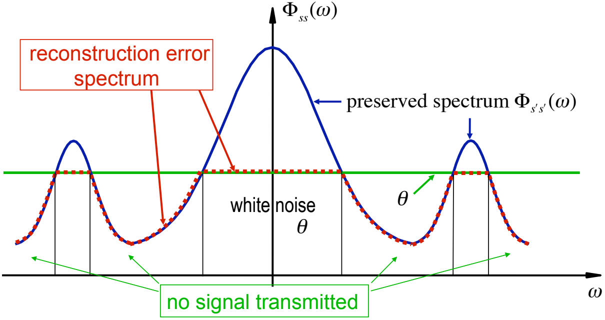
$\Longrightarrow$ **Rate-distortion function $R(D)$ for Gaussian sources with memory**

$$D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\Phi_{SS}(\omega), \theta\} d\omega$$

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left\{0, \frac{1}{2}\log_2 \frac{\Phi_{SS}(\omega)}{\theta}\right\} d\omega \quad (338)$$

$\Longrightarrow$ **Specifies upper bound for $R(D)$ of all processes with the same $\Phi_{SS}(\omega)$**

# Illustration of Minimization Approach




Similar to reverse water filling

- At each frequency, the variance of the frequency component as given by the spectral density $\Phi_{SS}(\omega)$ is compared to the parameter $\theta$, which represents the target mean squared error of that frequency component
- 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

# Rate-Distortion Function for Gauss-Markov Sources

- $R(D)$ for zero-mean Gauss-Markov process with $|\rho| < 1$ and variance $\sigma^2$

$$S_n = Z_n + \rho \cdot S_{n-1} \tag{339}$$

- Auto-correlation function and spectral density function are given as

$$\phi[k] = \sigma^2 |\rho|^k \qquad \Phi(\omega) = \sum_{k=-\infty}^{\infty} \phi[k] e^{-jk\omega} = \frac{\sigma^2(1-\rho^2)}{1 - 2\rho\cos\omega + \rho^2} \tag{340}$$

- If we choose

$$\theta \geq \min_{\forall\omega} \Phi_{SS}(\omega) = \sigma^2 \frac{1-\rho^2}{1-2\rho+\rho^2} = \sigma^2 \frac{1-\rho}{1+\rho} \tag{341}$$

we obtain

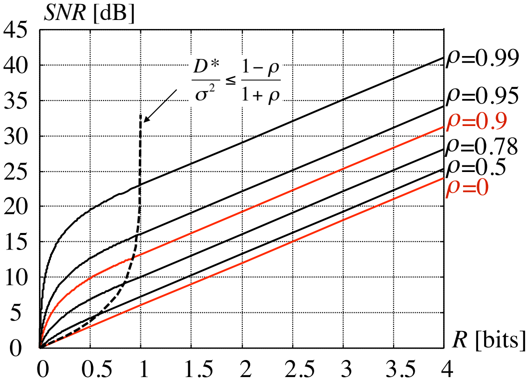
$$\boxed{R(D) = \frac{1}{2}\log_2 \frac{\sigma^2(1-\rho^2)}{D}} \tag{342}$$

# Rate-Distortion Function for Gauss-Markov Sources

- Corresponding distortion rate function for $R \geq \log_2(1+\rho)$ is given by

$$\boxed{D(R) = (1-\rho^2) \cdot \sigma^2 \cdot 2^{-2R}} \qquad (343)$$

- Includes result for Gaussian iid sources ($\rho = 0$)



# Chapter Summary

Rate-distortion theory

- Determine minimum rate $R$ for a given distortion $D$ and source
- Determine minimum distortion $D$ for a given distortion $R$ and source

Operational rate-distortion function

- Fundamental bound as minimum over all possible source codes

Information rate-distortion function

- Minimum over all conditional pdfs $g_{S'|S}(s'|s)$
- Coincides with operational rate-distortion function
- Use term rate-distortion function $R(D)$ for both
- Fundamental bound for lossless coding is given by $R(0)$
- Discrete sources: $R(D)$ is a convex function with $R(0) = \bar{H}(\boldsymbol{S})$
- Continuous sources: $R(D)$ is a convex function with $R(0) \rightarrow \infty$
- MSE distortion measure: $D(0) = \sigma^2$

# Chapter Summary

Shannon lower bound

- Lower bound of rate-distortion function
- Asymptotically tight for high rates
- Suitable reference for performance evaluation at high rates
- Shannon lower bound $R_L(D)$ can often be computed analytically
- Computed $R_L(D)$ for several iid sources and Gaussian source with memory

Rate-distortion function for Gaussian sources and MSE distortion

- $R(D)$ for Gaussian iid sources coincides with Shannon lower bound
- Any other source than the Gaussian iid source with the same variance requires less bits for same MSE distortion
- $R(D)$ for Gaussian source with memory can be specified as parametric expression using the power spectral density $\Phi_{SS}(\omega)$
- Derived analytic expression for Gauss-Markov source and $R \geq \log_2(1+\rho)$
- $R(D)$ for Gaussian source with memory and a spectral density $\Phi_{SS}(\omega)$ specifies an upper bound for all other sources with the same spectral density

$\Longrightarrow$ Gaussian sources are the most difficult to code

# Exercise 11

A fair die is rolled at the same time as a fair coin is tossed. Let $A$ be the number on the upper surface of the die and let $B$ describe the outcome of the coin toss, where $B$ is equal to 1 if the result is "head" and it is equal to 0 if the result if "tail". The random variables $X$ and $Y$ are given by $X = A + B$ and $Y = A - B$, respectively.

Calculate:

- the joint entropy $H(X, Y)$,
- the marginal entropies $H(X)$ and $H(Y)$,
- the conditional entropies $H(X|Y)$ and $H(Y|X)$,
- the mutual information $I(X; Y)$.

# Exercise 12

Consider a stationary Gauss-Markov process $\mathbf{X} = \{X_n\}$ with mean $\mu$, variance $\sigma^2$, and the correlation coefficient $\rho$ (correlation coefficient between two successive random variables).

Determine the mutual information $I(X_k; X_{k+N})$ between two random variables $X_k$ and $X_{k+N}$, where the distance between the random variables is $N$ times the sampling interval.

Interpret the results for the special cases $\rho = -1$, $\rho = 0$, and $\rho = 1$.

*Hint: In the lecture, we showed*

$$E\left\{(\mathbf{X} - \mu_N)^{\mathrm{T}} \cdot \mathbf{C}_N^{-1} \cdot (\mathbf{X} - \mu_N)\right\} = N,$$

*which can be useful for the problem.*

# Exercise 13

Show that for discrete random processes the fundamental bound for lossless coding is a special case of the fundamental bound for lossy coding.

# Exercise 14

Determine the Shannon lower bound with MSE distortion, as distortion-rate function, for iid processes with the following pdfs:

- The exponential pdf $f_E(x) = \lambda \cdot e^{-\lambda \cdot x}$, with $x \geq 0$
- The zero-mean Laplace pdf $f_L(x) = \frac{\lambda}{2} \cdot e^{-\lambda \cdot |x|}$

Express the distortion-rate function for the Shannon lower bound as a function of the variance $\sigma^2$.

Which of the given pdfs is easier to code (if the variance is the same)?