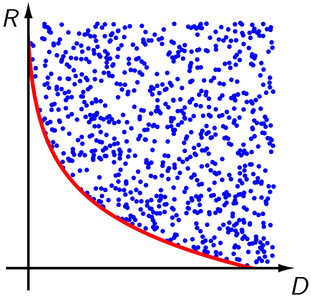
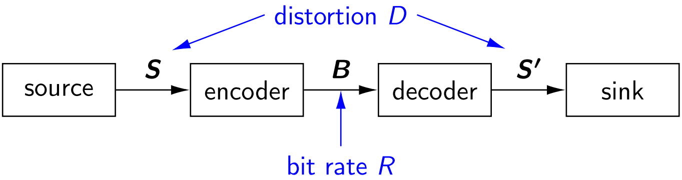
Rate-Distortion Theory II



Source Coding System

- Source: Statistical properties characterized by random process $\boldsymbol{S}$
- Encoder: Deterministic process for generating bitstream $\boldsymbol{B} = f_{\text{enc}}(\boldsymbol{S})$
- Decoder: Deterministic process for generating reconstruction $\boldsymbol{S}' = f_{\text{dec}}(\boldsymbol{B})$
- Sink: Receiver of reconstruction $\boldsymbol{S}'$



Characterization of Coding Efficiency

- Bit rate R: Average number of bits per sample (or per time unit)
- Distortion D: Measure characterizing difference between $\boldsymbol{S}$ and $\boldsymbol{S}'$

Distortion Measures

Definition: Distortion

Measure of deviation between a block of N input samples $\boldsymbol{s} = \{s_0, \ldots, s_{N-1}\}$ and the corresponding block of N reconstructed samples $\boldsymbol{s'} = \{s'_0, \ldots, s'_{N-1}\}$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) \geq 0 \qquad \text{(equality iff } \boldsymbol{s} = \boldsymbol{s'}\text{)}$$

- In media coding: Quality is eventually judged by human beings
- Ideally: Should use distortion measures that reflect human perception
- Very difficult to specify such distortion measures

➡ Will use simple (mathematically tractable) distortion measures

Additive Distortion Measures

- Average of single symbol distortions $d_1(s, s')$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \frac{1}{N} \sum_{k=0}^{N-1} d_1(s_k, s'_k) \tag{1}$$

Common Additive Distortion Measures

Difference Distortion Measures

- Single symbol distortion $d_1(s, s')$

$$d_1(s_k, s'_k) = \left| s_k - s'_k \right|^p \qquad \text{with} \qquad p > 0 \tag{2}$$

→ Average distortion for vectors of N samples

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \left\| \boldsymbol{s} - \boldsymbol{s'} \right\|_p^p = \frac{1}{N} \sum_{k=0}^{N-1} \left| s_k - s'_k \right|^p \tag{3}$$

Mean Squared Error

- In this course: Mainly **mean squared error** (MSE)

$$d_1(s_k, s'_k) = (s_k - s'_k)^2 \tag{4}$$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \left\| \boldsymbol{s} - \boldsymbol{s'} \right\|_2^2 = \frac{1}{N} \sum_{k=0}^{N-1} (s_k - s'_k)^2 \tag{5}$$

Operational Rate-Distortion Function

Consider given source / random process $\boldsymbol{S}$

- Each code Q is associated with a rate-distortion point

$$(R, D) = \big(r(Q), \delta(Q)\big)$$

Note: For a given source $\boldsymbol{S}$ and known encoder and decoder, $r(Q)$ and $\delta(Q)$ represent probabilistic averages for the rate and distortion

- **Achievable rate distortion point** (R, D)

$$\exists Q : \quad r(Q) \le R \quad \wedge \quad \delta(Q) \le D \tag{6}$$

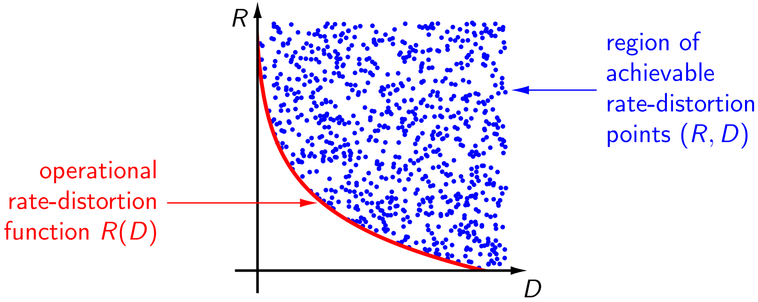
- **Operational rate-distortion function** $R(D)$

$$\boxed{R(D) = \inf_{Q:\,\delta(Q) \le D} r(Q)} \tag{7}$$

- Inverse: **Operational distortion-rate function** $D(R)$

$$\boxed{D(R) = \inf_{Q:\,r(Q) \le R} \delta(Q)} \tag{8}$$

Operational Rate-Distortion Function



Operational rate-distortion function $R(D)$ / **distortion-rate function** $D(R)$

- Divides R-D space into region of achievable and region of non-achievable rate-distortion points
- Specifies the greatest lower bound for lossy coding

➡ **Impossible to evaluate** (minimization over all possible codes)

➡ **Information rate-distortion function** (Shannon, 1948)

Mutual Information for Discrete Random Variables

Definition

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B) \tag{9}$$

Interpretation

- $H(A)$: Uncertainty about random variable A
- $H(A\,|\,B)$: Uncertainty about random variable A after observing B

➜ $I(A;B)$: Reduction of uncertainty about A due to the observation of B

Mutual information $I(A;B)$:

Average amount of information that a random variable B carries about another random variable A

Mutual Information for Discrete Random Variables

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\{-\log_2 p_A(A)\} - \mathrm{E}\{-\log_2 p_{A|B}(A\,|\,B)\} \\
&= \mathrm{E}\left\{\log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)}\right\} = \mathrm{E}\left\{\log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)}\right\} && (10) \\
&= \sum_a \sum_b p_{AB}(a,b) \log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} && (11)
\end{aligned}
$$

- Due to symmetry

$$I(A;B) = H(A) - H(A\,|\,B) = H(B) - H(B\,|\,A) \qquad (12)$$

➡ Mutual information $I(A;B)$ specifies

- Average amount of information that A carries about B
- Average amount of information that B carries about A

Mutual Information for Discrete Random Vectors

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B) \tag{13}$$
$$= H(B) - H(B\,|\,A) \tag{14}$$
$$= \sum_{a}\sum_{b} p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \tag{15}$$

➜ Straightforward extension to random vectors $\boldsymbol{A}$ and $\boldsymbol{B}$

$$I(\boldsymbol{A};\boldsymbol{B}) = H(\boldsymbol{A}) - H(\boldsymbol{A}\,|\,\boldsymbol{B}) \tag{16}$$
$$= H(\boldsymbol{B}) - H(\boldsymbol{B}\,|\,\boldsymbol{A}) \tag{17}$$
$$= \sum_{\boldsymbol{a}}\sum_{\boldsymbol{b}} p_{\boldsymbol{AB}}(\boldsymbol{a},\boldsymbol{b})\,\log_2 \frac{p_{\boldsymbol{AB}}(\boldsymbol{a},\boldsymbol{b})}{p_{\boldsymbol{A}}(\boldsymbol{a})\,p_{\boldsymbol{B}}(\boldsymbol{b})} \tag{18}$$

Note: $H(\cdot)$ and $H(\cdot\,|\,\cdot)$ specifies the corresponding block entropy and conditional block entropy

Properties of Mutual Information

- General formulation: Discrete random vectors $\boldsymbol{A}$ and $\boldsymbol{B}$

$$I(\boldsymbol{A};\boldsymbol{B}) = H(\boldsymbol{A}) - H(\boldsymbol{A}\,|\,\boldsymbol{B}) = H(\boldsymbol{B}) - H(\boldsymbol{B}\,|\,\boldsymbol{A}) \tag{19}$$

- Relationship to Kullback-Leibler Divergence

$$I(\boldsymbol{A};\boldsymbol{B}) = D\Big(p_{\boldsymbol{AB}}(\boldsymbol{a},\boldsymbol{b})\,\Big|\Big|\,p_{\boldsymbol{A}}(\boldsymbol{a})\,p_{\boldsymbol{B}}(\boldsymbol{b})\Big) \tag{20}$$

- Relationship to marginal (block) entropy

$$I(\boldsymbol{A};\boldsymbol{B}) \le H(\boldsymbol{A}) \qquad \text{and} \qquad I(\boldsymbol{A};\boldsymbol{B}) \le H(\boldsymbol{B}) \tag{21}$$

- Independent random variables / random vectors

$$I(\boldsymbol{A};\boldsymbol{B}) = 0 \tag{22}$$

- Deterministic functional relationship $\boldsymbol{B} = f(\boldsymbol{A})$

$$\boldsymbol{B} = f(\boldsymbol{A}) \quad \Longrightarrow \quad I(\boldsymbol{A};\boldsymbol{B}) = H(\boldsymbol{B}) \tag{23}$$

Mutual Information for Continuous Random Variables

Mutual information $I(A;B)$

- Discrete random variables A and B

$$\begin{aligned} I(A;B) &= H(A) - H(A\,|\,B) \\ &= H(B) - H(B\,|\,A) \end{aligned}$$

➡ Continuous random variables

- Discrete entropy and discrete conditional entropy are not defined
- $H(A)$ and $H(A\,|\,\cdot)$ both approach infinity if A is continuous

Define mutual information for continuous random variables

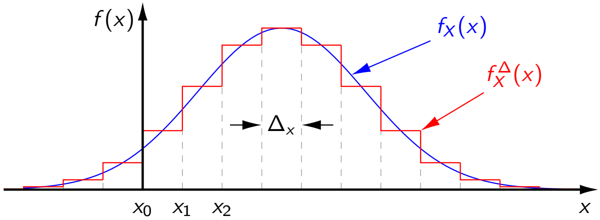
- Via approximation by discrete random variables

➡ Quantize pdfs with a quantizaton step size Δ

➡ Calculate mutual information for resulting discrete random variables

➡ Consider limit for $\Delta \rightarrow 0$

Discretization of Continuous Random Variables



- Discrete approximation $f_X^{\Delta}(x)$ of probability density function $f_X(x)$

$$\forall x : x_k \leq x < x_{k+1}, \qquad f_X^{\Delta}(x) = \frac{1}{\Delta_x} \int_{x_k}^{x_{k+1}} f_X(t)\,\mathrm{d}t \tag{24}$$

- Pmf $p_{X_\Delta}(x_k)$ for discrete random variable X_Δ with alphabet $\{x_k\}$

$$p_{X_\Delta}(x_k) = \int_{x_k}^{x_{k+1}} f_X(t)\,\mathrm{d}t = f_X^{\Delta}(x_k) \cdot \Delta_x \tag{25}$$

- Joint pmf $p_{X_\Delta Y_\Delta}(x_k, y_i)$ for two discrete approximations X_Δ and Y_Δ

$$p_{X_\Delta Y_\Delta}(x_k, y_i) = f_{XY}^{\Delta}(x_k, y_i) \cdot \Delta_x \cdot \Delta_y \tag{26}$$

Mutual Information for Continuous Random Variables

- Mutual information $I(X_\Delta; Y_\Delta)$ for discrete approximations X_Δ and Y_Δ

$$I(X_\Delta; Y_\Delta) = \sum_{\forall x_k} \sum_{\forall y_i} p_{X_\Delta Y_\Delta}(x_k, y_i) \cdot \log_2 \frac{p_{X_\Delta Y_\Delta}(x_k, y_i)}{p_{X_\Delta}(x_k)\, p_{Y_\Delta}(y_i)} \tag{27}$$

$$= \sum_{\forall x_k} \sum_{\forall y_i} f^{\Delta}_{XY}(x_k, y_i) \cdot \log_2 \frac{f^{\Delta}_{XY}(x_k, y_i)}{f^{\Delta}_{X}(x_k)\, f^{\Delta}_{Y}(y_i)} \cdot \Delta_x \cdot \Delta_y \tag{28}$$

- Mutual information $I(X; Y)$ for continuous random variables X and Y

$$I(X; Y) = \lim_{\substack{\Delta_x \to 0 \\ \Delta_y \to 0}} I(X_\Delta; Y_\Delta) \tag{29}$$

➡ **Mutual information for continuous random variables**

$$I(X; Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x)\, f_Y(y)} \,\mathrm{d}x\, \mathrm{d}y \tag{30}$$

Mutual Information for Continuous Case

- Mutual information for continuous random variables X and Y

$$I(X;Y) = \mathrm{E}\left\{ \log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)} \right\} \tag{31}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) \log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)} \,\mathrm{d}x\,\mathrm{d}y \tag{32}$$

- Extension to random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ (with dimensions N and M)

$$I(\boldsymbol{X};\boldsymbol{Y}) = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{X}}(\boldsymbol{x})\, f_{\boldsymbol{Y}}(\boldsymbol{y})} \right\} \tag{33}$$

$$= \int_{\mathbb{R}^N} \int_{\mathbb{R}^M} f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) \log_2 \frac{f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{X}}(\boldsymbol{x})\, f_{\boldsymbol{Y}}(\boldsymbol{y})} \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \tag{34}$$

Mutual Information for Source Coding

- N original samples: Continuous or discrete random vector $\boldsymbol{S}$
- N reconstructed samples: Discrete random vector $\boldsymbol{S'}$

➞ Mutual information

$$I_N(\boldsymbol{S};\boldsymbol{S'}) = H_N(\boldsymbol{S'}) - H_N(\boldsymbol{S'}\,|\,\boldsymbol{S}) \le H_N(\boldsymbol{S'})$$

Note: Equality is achieved iff $\boldsymbol{S'}$ is a deterministic function of $\boldsymbol{S}$

- Recall: Fundamental lossless coding theorem

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

- **Lower bound for lossless coding using mutual information**

$$\boxed{r(Q) \ge \lim_{N\to\infty} \frac{H_N(\boldsymbol{S'})}{N} \ge \lim_{N\to\infty} \frac{I_N(\boldsymbol{S};\boldsymbol{S'})}{N}} \qquad (35)$$

Distortion and Mutual Information as Expected Values

- Consider given source $\boldsymbol{S}$ and source code Q

➡ N-th order distortion and N-th order mutual information

$$\delta_N(Q) = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}' \tag{36}$$

$$I_N(Q) = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}' \tag{37}$$

$$\text{with} \quad f_{\boldsymbol{S}'}(\boldsymbol{s}') = \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}$$

- Joint pdf $f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')$ can be written as

$$f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N(\boldsymbol{s}' \,|\, \boldsymbol{s}) \tag{38}$$

where

➡ N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$ specifies the statistical properties of the source $\boldsymbol{S}$

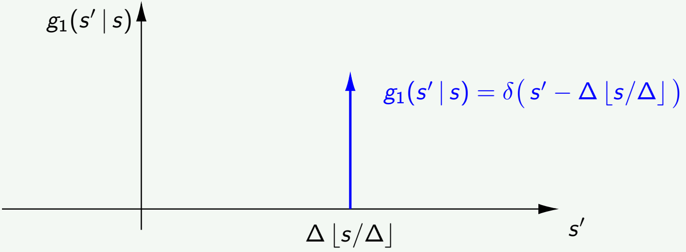
➡ N-th order **conditional pdf $g_N(\boldsymbol{s}' \,|\, \boldsymbol{s})$ specifies the statistical properties of the source code Q for given source $\boldsymbol{S}$**

Source Code as Conditional Pdf: Example

Example 1: Simple Scalar Quantization

- Mapping $s \rightarrow s'$: $\quad s' = \Delta \lfloor s/\Delta \rfloor \quad$ (quantization step size Δ)

➡ Conditional pdf $g_1(s' \,|\, s)$



- $N > 1$: Multivariate conditional pdf $g_N(\boldsymbol{s'} \,|\, \boldsymbol{s})$

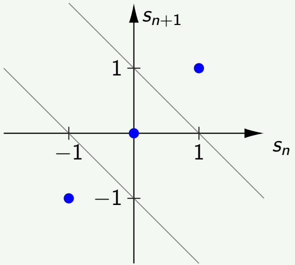
$$g_N(\boldsymbol{s'} \,|\, \boldsymbol{s}) = \delta\big(s'_0 - \Delta \lfloor s_0/\Delta \rfloor \big) \cdot \delta\big(s'_1 - \Delta \lfloor s_1/\Delta \rfloor \big) \cdot \ldots$$

Source Code as Conditional Pdf: Example

Example 2: Mapping of two Consecutive Samples

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

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



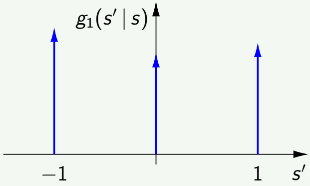
➡ Conditional pdf $g_2(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$g_2(\boldsymbol{s}' \,|\, \boldsymbol{s}) = \begin{cases} \delta(\boldsymbol{s}' + (1,1)) & : \quad s_0 + s_1 < -1 \\ \delta(\boldsymbol{s}' - (1,1)) & : \quad s_0 + s_1 > 1 \\ \delta(\boldsymbol{s}') & : \quad |s_0 + s_1| \leq 1 \end{cases}$$

➡ Conditional pdf $g_1(s' \,|\, s)$

$$g_1(s' \,|\, s) = a \cdot \delta(s' + 1) + b \cdot \delta(s') + c \cdot \delta(s' - 1)$$

with $a + b + c = 1$



Distortion and Mutual Information using Conditional Pdf

- Statistical properties of a source code Q for a given source $\boldsymbol{S}$ can be described using an N-th order conditional pdf $g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s})$

$$\text{source code } Q: \qquad f_{\boldsymbol{SS'}}(\boldsymbol{s}, \boldsymbol{s'}) = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s}) \tag{39}$$

➡ N-th order distortion for given source $\boldsymbol{S}$

$$\delta_N(Q) = \delta_N(g_N^Q) = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s})\; d_N(\boldsymbol{s}, \boldsymbol{s'})\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s'} \tag{40}$$

➡ N-th order mutual information for given source $\boldsymbol{S}$

$$I_N(Q) = I_N(g_N^Q) = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s})\; \log_2 \frac{g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s})}{f_{\boldsymbol{S'}}(\boldsymbol{s'})}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s'} \tag{41}$$

with

$$f_{\boldsymbol{S'}}(\boldsymbol{s'}) = \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s'} \mid \boldsymbol{s})\; \mathrm{d}\boldsymbol{s} \tag{42}$$

Information Rate-Distortion Function

- Recall: Operational rate-distortion function

$$R(D) = \inf_{Q:\,\delta(Q)\leq D} r(Q) \tag{43}$$

- Lower bound for lossless coding using mutual information

$$r(Q) \geq \bar{H}(\boldsymbol{S}') = \lim_{N\to\infty} \frac{H_N(\boldsymbol{S}')}{N} \geq \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}')}{N} = \lim_{N\to\infty} \frac{I_N(g_N^Q)}{N} \tag{44}$$

➡ Lower bound: Infimum over all sources codes Q with $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$R(D) \geq \inf_{Q:\,\delta(Q)\leq D} \lim_{N\to\infty} \frac{I_N(g_N^Q)}{N} = \lim_{N\to\infty} \inf_{g_N^Q:\,\delta(g_N^Q)\leq D} \frac{I_N(g_N^Q)}{N} \tag{45}$$

- Conditional pdfs g_N^Q for source codes Q: Subset of all conditional pdfs g_N

➡ Lower bound of $R(D)$: **Information rate-distortion function** $R^I(D)$

$$\boxed{R(D) \geq R^I(D)} \quad \text{with} \quad \boxed{R^I(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta(g_N)\leq D} \frac{I_N(g_N)}{N}} \tag{46}$$

Rate-Distortion Function

Information vs Operational Rate-Distortion Function

1. We showed: $R^{\mathrm{I}}(D)$ is a lower bound for $R(D)$

$$R(D) \geq R^{\mathrm{I}}(D) \qquad \text{with} \qquad R^{I}(D) = \lim_{N \to \infty} \inf_{g_N \colon \delta(g_N) \leq D} \frac{I_N(g_N)}{N} \tag{47}$$

2. Can be shown: $R^{\mathrm{I}}(D)$ is asymptotically achievable
   - For any $D > 0$ and $\varepsilon > 0$, there exists a source code Q with

$$\delta(Q) \leq D \qquad \text{and} \qquad r(Q) \leq R^{\mathrm{I}}(D) + \varepsilon \tag{48}$$

   - Proof: Cover, Thomas, "Elements of Information Theory"

➡ **Information rate-distortion function $R^{\mathrm{I}}(D)$ coincides with operational rate-distortion function $R(D)$**

➡ Use term **rate-distortion function $R(D)$** for both

Fundamental Source Coding Theorem

Rate-Distortion Function

- **Greatest lower bound for lossy source coding** (source codes Q)

$$\forall Q : \; \delta(Q) \leq D, \quad r(Q) \geq R(D) \tag{49}$$

with (information) **rate-distortion function** $R(D)$ given by

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N) \leq D} \frac{I_N(\, g_N \,)}{N} \tag{50}$$

Distortion-Rate Function

- Distortion-rate function $D(R)$: Inverse of rate-distortion function $R(D)$

$$\forall Q : \; r(Q) \leq R, \quad \delta(Q) \geq D(R) \tag{51}$$

with (information) **distortion-rate function** $D(R)$ given by

$$D(R) = \lim_{N\to\infty} \inf_{g_N:\, I_N(g_N)/N \leq R} \delta_N(\, g_N \,) \tag{52}$$

Special Case: Lossless Source Coding Theorem

Lossless Coding of Discrete Source

- Rate-distortion function $R(D)$

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\,g_N\,)}{N} \tag{53}$$

- Lossless coding: $\boldsymbol{S}' = \boldsymbol{S}$

$$I_N(\,g_N\,) = I_N(\boldsymbol{S};\boldsymbol{S}) = H_N(\boldsymbol{S}) - H_N(\boldsymbol{S}\,|\,\boldsymbol{S}) = H_N(\boldsymbol{S}) \tag{54}$$

➡ Rate-distortion function for lossless coding ($D = 0$)

$$R(0) = \lim_{N\to\infty} \frac{H_N(\boldsymbol{S})}{N} = \bar{H}(\boldsymbol{S}) \tag{55}$$

➡ Lossless coding theorem: Special case of fundamental source coding theorem

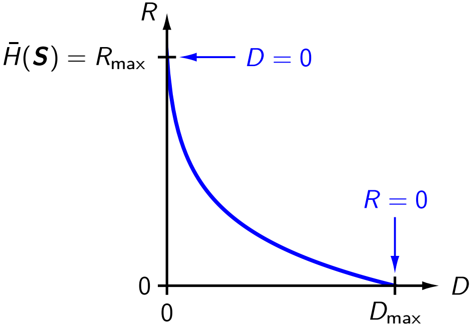
$$\forall Q:\ \delta(Q) = 0, \quad r(Q) \ge R(0) = \bar{H}(\boldsymbol{S}) \tag{56}$$

Rate-Distortion Function for Discrete Sources

Properties of $R(D)$

- Domain: $[0, +\infty)$
- Non-increasing function
- Convex function
- Maximum rate (lossless coding)

$$R(0) = R_{\max} = \bar{H}(\boldsymbol{S})$$



- There exists is maximum value $D_{\max}$ for the distortion D

$$\exists\, D_{\max} : \qquad R(D) = \begin{cases} > 0 & : \quad D < D_{\max} \\ 0 & : \quad D \geq D_{\max} \end{cases} \tag{57}$$

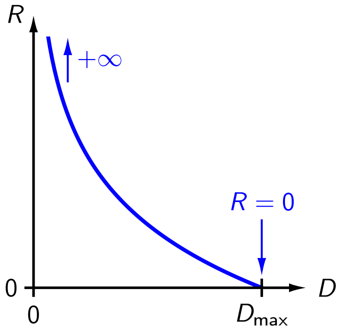
➡ MSE distortion measure: $D_{\max} = \sigma^2$

Rate-Distortion Function for Continuous Sources

Properties of $R(D)$

- Domain: $(\,0,\,+\infty\,)$
- Non-increasing function
- Convex function
- Unlimited rate

$$\lim_{D \to 0} R(D) = +\infty$$



- There exists is maximum value $D_{\max}$ for the distortion D

$$\exists\, D_{\max}: \qquad R(D) = \begin{cases} > 0 & : \quad D < D_{\max} \\ 0 & : \quad D \geq D_{\max} \end{cases} \tag{58}$$

➡ MSE distortion measure: $D_{\max} = \sigma^2$

Additive Distortion Measures & IID Sources

Additive Distortion Measures

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

$$\delta_N(\, g_N \,) \;=\; \mathrm{E}\{\, d_N(\boldsymbol{S}, \boldsymbol{S'})\,\} = \mathrm{E}\left\{ \frac{1}{N} \sum_{k=0}^{N-1} d_1(S_k, S'_k) \right\}$$
$$= \mathrm{E}\{\, d_1(S, S')\,\} = \delta_1(\, g_1 \,) \qquad (59)$$

IID Sources

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

$$I_N(\, g_N \,) \;=\; \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{SS'}}(\boldsymbol{S}, \boldsymbol{S'})}{f_{\boldsymbol{S}}(\boldsymbol{S})\, f_{\boldsymbol{S'}}(\boldsymbol{S'})} \right\} = \mathrm{E}\left\{ \log_2 \left(\frac{f_{SS'}(S, S')}{f_S(S)\, f_{S'}(S')} \right)^N \right\}$$
$$= N \cdot \mathrm{E}\left\{ \log_2 \frac{f_{SS'}(S, S')}{f_S(S)\, f_{S'}(S')} \right\} = N \cdot I_1(\, g_1 \,) \qquad (60)$$

N-th order Rate-Distortion Functions

IID sources S and additive distortion measure

- Rate-distortion function

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\le D} \frac{I_N(\, g_N\,)}{N} = \inf_{\mathbf{g_1}:\, \delta_{\mathbf{1}}(\mathbf{g_1})\le D} I_1(\, g_1\,) \tag{61}$$

N-th order rate-distortion / distortion-rate function

$$R_N(D) = \inf_{g_N:\, \delta_N(g_N)\le D} \frac{I_N(\, g_N\,)}{N} \quad \text{and} \quad D_N(R) = \inf_{g_N:\, I_N(g_N)/N\le R} \delta_N(\, g_N\,) \tag{62}$$

➡ Rate-distortion / distortion-rate function

$$R(D) = \lim_{N\to\infty} R_N(D) \quad \text{and} \quad D(R) = \lim_{N\to\infty} D_N(R) \tag{63}$$

➡ IID sources and additive distortion measure

$$R(D) = R_1(D) \quad \text{and} \quad D(R) = D_1(R) \tag{64}$$

Discussion of Rate-Distortion Functions

Greatest Lower Bound for Lossless Coding

- **Operational / information rate-distortion function**

$$R(D) = \inf_{Q:\,\delta(Q)\le D} r(Q) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\,g_N\,)}{N}$$

- Operational RDF: ➡ Obvious definition, but impossible to evaluate
- Information RDF: ➡ Property of source (no need to consider codes)
  - ➡ Still impossible to evaluate directly
  - ➡ Numerical minimization: Blahut-Arimoto algorithm

How can we proceed?

- Derive lower bound for rate-distortion function
- For some sources and distortion measures:
  - ➡ Show that lower bound is achievable

Differential Entropy

Mutual Information

- Mutual information: Continuous random vector $\boldsymbol{X}$ and random vector $\boldsymbol{Y}$

$$I(\boldsymbol{X};\boldsymbol{Y}) = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{XY}}(\boldsymbol{X},\boldsymbol{Y})}{f_{\boldsymbol{X}}(\boldsymbol{X})\, f_{\boldsymbol{Y}}(\boldsymbol{Y})} \right\} = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}\,|\,\boldsymbol{Y})}{f_{\boldsymbol{X}}(\boldsymbol{X})} \right\}$$

$$= \mathrm{E}\{ -\log_2 f_{\boldsymbol{X}}(\boldsymbol{X}) \} - \mathrm{E}\{ -\log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}\,|\,\boldsymbol{Y}) \}$$

$$= h(\boldsymbol{X}) - h(\boldsymbol{X}\,|\,\boldsymbol{Y}) \quad (65)$$

Definition of Differential Entropy

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

$$h(\boldsymbol{X}) = \mathrm{E}\{ -\log_2 f_{\boldsymbol{X}}(\boldsymbol{X}) \} \quad (66)$$

- **Conditional differential entropy** of continuous random vector $\boldsymbol{X}$ given $\boldsymbol{Y}$

$$h(\boldsymbol{X}\,|\,\boldsymbol{Y}) = \mathrm{E}\{ -\log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X}\,|\,\boldsymbol{Y}) \} \quad (67)$$

Differential Entropy

- Differential entropy of a continuous random vector $\boldsymbol{X}$ (dimension N)

$$h(\boldsymbol{X}) = \mathrm{E}\{-\log_2 f_{\boldsymbol{X}}(\boldsymbol{X})\} = -\int_{\mathbb{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x}) \log_2 f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \tag{68}$$

- Differential entropy of a continuous random vector $\boldsymbol{X}$ (dimension N) given a random vector $\boldsymbol{Y}$ (dimension M)

$$h(\boldsymbol{X} \mid \boldsymbol{Y}) = \mathrm{E}\{-\log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{X} \mid \boldsymbol{Y})\} = -\int_{\mathbb{R}^N}\int_{\mathbb{R}^M} f_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y}) \log_2 f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \tag{69}$$

Example: Differential Entropy

Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases} \tag{70}$$

→ Differential entropy:

$$\begin{aligned} h(S) &= -\int_{-\infty}^{\infty} f(s) \log_2 f(s) \, \mathrm{d}s \\ &= -\int_{-A/2}^{A/2} \frac{1}{A} \log_2 \frac{1}{A} \, \mathrm{d}s \\ &= \frac{1}{A} \log_2 A \cdot \int_{-A/2}^{A/2} \mathrm{d}s \\ &= \log_2 A \end{aligned} \tag{71}$$

→ Note: Differential entropy $h(S)$ can become negative

Mutual Information and Entropy

Summary of Relationships

- Mutual information

$$\text{discrete } X: \quad I(X;Y) = H(X) - H(X\,|\,Y)$$

$$\text{continuous } X: \quad I(X;Y) = h(X) - h(X\,|\,Y)$$

- Discrete and differential entropies

$$H(X) = \mathrm{E}\{-\log_2 p_X(X)\} \qquad H(X\,|\,Y) = \mathrm{E}\{-\log_2 p_{X|Y}(X\,|\,Y)\}$$

$$h(X) = \mathrm{E}\{-\log_2 f_X(X)\} \qquad h(X\,|\,Y) = \mathrm{E}\{-\log_2 f_{X|Y}(X\,|\,Y)\}$$

- Relationship between discrete and differential entropy:
  Quantization X_Δ of continuous random variable X (with step size Δ)

$$\lim_{\Delta \to 0} \left(H(X_\Delta) + \log_2 \Delta \right) = h(X)$$

Differential Entropy Rate

N-th Order Differential Entropy

- Differential entropy for N consecutive random variables $S_k, \cdots, S_{k+N-1}$

$$\begin{aligned} h_N(\boldsymbol{S}) &= h(S_k, S_{k+1}, \cdots, S_{k+N-1}) \\ &= \mathrm{E}\{ -\log_2 f_{\boldsymbol{S}}(S_k, S_{k+1}, \cdots, S_{k+N-1}) \} \end{aligned} \quad (72)$$

Differential Entropy Rate

- Definition (analog to discrete entropy rate)

$$\bar{h}(\boldsymbol{S}) = \lim_{N \to \infty} \frac{h_N(\boldsymbol{S})}{N} \quad (73)$$

➜ Special sources (proof: same as for discrete entropy)

$$\text{IID:} \quad \bar{h}(\boldsymbol{S}) = h(S) \quad (74)$$

$$\text{Markov:} \quad \bar{h}(\boldsymbol{S}) = h(S_n \,|\, S_{n-1}) \quad (75)$$

Differential Entropy Rate for Gaussian IID

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(s-\mu)^2}{2\sigma^2}} \tag{76}$$

→ Differential entropy rate for Gaussian IID source

$$\begin{aligned} \bar{h}^{Giid}(\boldsymbol{S}) = h^G(S) &= \mathrm{E}\{ -\log_2 f_G(S) \} \\ &= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2} \log_2 e \right\} \\ &= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\mathrm{E}\{ (S-\mu)^2 \} \log_2 e \\ &= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\sigma^2 \log_2 e \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) \end{aligned} \tag{77}$$

N-th Order Differential Entropy for Stationary Gaussian

- N-th order pdf for stationary Gaussian with covariance matrix $\boldsymbol{C}_N$

$$f_G(\boldsymbol{s}) = \frac{1}{\sqrt{(2\pi)^N\,|\boldsymbol{C}_N|}}\, e^{-\frac{1}{2}\,(\boldsymbol{s}-\boldsymbol{\mu})^{\mathrm{T}}\,\boldsymbol{C}_N^{-1}\,(\boldsymbol{s}-\boldsymbol{\mu})} \tag{78}$$

➡ N-th order differential entropy for Gaussian with covariance matrix $\boldsymbol{C}_N$

$$\begin{aligned} h_N^G(\boldsymbol{S}) &= \mathrm{E}\{-\log_2 f_G(\boldsymbol{S})\} \\ &= \frac{1}{2}\log_2\Big((2\pi)^N\,|\boldsymbol{C}_N|\Big) + \\ &\qquad \frac{\log_2 e}{2}\cdot\mathrm{E}\{(\boldsymbol{S}-\boldsymbol{\mu})^{\mathrm{T}}\,\boldsymbol{C}_N^{-1}\,(\boldsymbol{S}-\boldsymbol{\mu})\} \end{aligned} \tag{79}$$

N-th Order Differential Entropy for Stationary Gaussian

- General stationary process with covariance matrix $\boldsymbol{C}_N$

$$
\begin{aligned}
&\mathrm{E}\{\,(\boldsymbol{S}-\boldsymbol{\mu})^{\mathrm{T}}\,\boldsymbol{C}_N^{-1}\,(\boldsymbol{S}-\boldsymbol{\mu})\,\} \\
&\qquad = \mathrm{E}\left\{\sum_{k=0}^{N-1}\sum_{i=0}^{N-1}(S_i-\mu)\,(\boldsymbol{C}_N^{-1})_{i,k}\,(S_k-\mu)\right\} \\
&\qquad = \sum_{k=0}^{N-1}\sum_{i=0}^{N-1}\mathrm{E}\{\,(S_i-\mu)(S_k-\mu)\,\}\;(\boldsymbol{C}_N^{-1})_{i,k} \\
&\qquad = \sum_{k=0}^{N-1}\left(\sum_{i=0}^{N-1}(\boldsymbol{C}_N)_{k,i}\,(\boldsymbol{C}_N^{-1})_{i,k}\right) \\
&\qquad = \sum_{k=0}^{N-1}(\boldsymbol{C}_N\,\boldsymbol{C}_N^{-1})_{k,k} = \sum_{k=0}^{N-1}1 \\
&\qquad = N
\end{aligned}
\tag{80}
$$

N-th Order Differential Entropy for Stationary Gaussian

- Showed: General stationary process with covariance matrix $\boldsymbol{C}_N$

$$\mathrm{E}\left\{ (\boldsymbol{S}-\boldsymbol{\mu})^{\mathrm{T}}\, \boldsymbol{C}_N^{-1}\, (\boldsymbol{S}-\boldsymbol{\mu}) \right\} = N \tag{81}$$

- Continue: N-th order differential entropy for stationary Gaussian

$$\begin{aligned} h_N^G(\boldsymbol{S}) &= \mathrm{E}\left\{ -\log_2 f_G(\boldsymbol{S}) \right\} \\ &= \frac{1}{2}\log_2\left((2\pi)^N\, |\boldsymbol{C}_N| \right) + \\ &\qquad\qquad \frac{\log_2 e}{2}\cdot \mathrm{E}\left\{ (\boldsymbol{S}-\boldsymbol{\mu})^{\mathrm{T}}\, \boldsymbol{C}_N^{-1}\, (\boldsymbol{S}-\boldsymbol{\mu}) \right\} \\ &= \frac{1}{2}\log_2\left((2\pi)^N\, |\boldsymbol{C}_N| \right) + \frac{N}{2}\log_2 e \\ &= \frac{1}{2}\log_2\left((2\pi e)^N\, |\boldsymbol{C}_N| \right) \end{aligned} \tag{82}$$

N-th Order Differential Entropy for Stationary Gaussian

- N-th order differential entropy for stationary Gaussian with $\boldsymbol{C}_N$

$$\begin{aligned} h_N^G(\boldsymbol{S}) &= \frac{1}{2}\log_2\left((2\pi e)^N\,|\boldsymbol{C}_N|\right) \\ &= \frac{N}{2}\log_2\left(2\pi e\,\sqrt[N]{|\boldsymbol{C}_N|}\right) \end{aligned} \tag{83}$$

- N-th order differential entropy for Gaussian iid with σ^2

$$\begin{aligned} h_N^{Giid}(\boldsymbol{S}) &= \mathrm{E}\{-\log_2 f_{Giid}(S_0, S_1, \cdots, S_{N-1})\} \\ &= \mathrm{E}\{-\log_2\left(f_G(S_0)\cdot f_G(S_1)\cdot\ldots\cdot f_G(S_{N-1})\right)\} \\ &= N\cdot\mathrm{E}\{-\log_2 f_G(S)\} \\ &= N\cdot h^G(S) \\ &= \frac{N}{2}\log_2\left(2\pi e\,\sigma^2\right) \end{aligned} \tag{84}$$

➡ **Question:** What is the relationship between h_N^G and h_N^{Giid} ?

Determinant and Trace of Autocovariance Matrix

- Determinant of covariance matrix $\boldsymbol{C}_N$: Product of eigenvalues

$$|\boldsymbol{C}_N| = \prod_{k=0}^{N-1} \xi_k \qquad \Longrightarrow \qquad \sqrt[N]{|\boldsymbol{C}_N|} = \sqrt[N]{\prod_{k=0}^{N-1} \xi_k} \tag{85}$$

- Trace of covariance matrix $\boldsymbol{C}_N$: Sum of eigenvalues

$$\mathrm{tr}\big(\boldsymbol{C}_N\big) = N \cdot \sigma^2 = \sum_{k=0}^{N-1} \xi_k \qquad \Longrightarrow \qquad \sigma^2 = \frac{1}{N} \sum_{k=0}^{N-1} \xi_k \tag{86}$$

- Inequality of arithmetic and geometric means

$$\sqrt[N]{\prod_{k=0}^{N-1} x_k} \le \frac{1}{N} \sum_{k=0}^{N-1} x_k \tag{87}$$

➡ Equality iff $x_0 = x_1 = x_2 = \cdots = x_{N-1}$

➡ Proofs can be found in literature (or Wikipedia)

N-th Order Differential Entropy for Stationary Gaussian

- Consider: Stationary Gaussian processes with given variance σ^2
- Inequality of arithmetic and geometric means for eigenvalues $\{\xi_k\}$ of $\boldsymbol{C}_N$

$$
\begin{aligned}
\sqrt[N]{\prod_{k=0}^{N-1} \xi_k} &\leq \frac{1}{N} \sum_{k=0}^{N-1} \xi_k \\
\sqrt[N]{|\boldsymbol{C}_N|} &\leq \sigma^2 \\
\frac{N}{2} \log_2 \left(2\pi e \sqrt[N]{|\boldsymbol{C}_N|} \right) &\leq \frac{N}{2} \log_2 \left(2\pi e \, \sigma^2 \right) \\
h_N^G(\boldsymbol{S}) &\leq h_N^{Giid}(\boldsymbol{S})
\end{aligned} \tag{88}
$$

For given variance σ^2

➡ Gaussian iid is the Gaussian process with max. N-th order differential entropy

➡ Gaussian iid is the Gaussian process with max. differential entropy rate

Kullback-Leibler Divergence

Recall: Discrete Random Variables

- Divergence inequality for pmfs p and q

$$D(\, p \,||\, q \,) = \sum_{\forall k} p_k \, \log_2 \left(\frac{p_k}{q_k} \right) \, \geq \, 0 \tag{89}$$

with equality if and only if $p = q$ (both pmfs are the same)

Continuous Random Variables

- Divergence inequality for pdfs f and g

$$D(\, f \,||\, g \,) = \int_{-\infty}^{\infty} f(s) \, \log_2 \frac{f(s)}{g(s)} \, \mathrm{d}s \, \geq \, 0 \tag{90}$$

with equality if and only if $f = g$ (both pdfs are the same)

- Proof is analog to discrete case
- Straightforward extension to random vectors (multi-variate pdfs)

Inequality for Given Covariance Function

- Consider a stationary random processes $\boldsymbol{S}$ with N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$ and auto-covariance function

$$\phi(k) = \phi_k = \mathrm{E}\big\{ \big(S_n - \mu\big)\big(S_{n+k} - \mu\big) \big\} \tag{91}$$

➡ $\phi(k)$ is the value on the k-th diagonal of auto-covariance matrices $\boldsymbol{C}_N$

- Let $\boldsymbol{G}$ represent a stationary Gaussian process with the N-th order pdf $f_{\boldsymbol{G}}(\boldsymbol{s})$ and the same auto-covariance function $\phi(k)$ (i.e., same matrices $\boldsymbol{C}_N$)

- Consider expected value $\mathrm{E}\{ -\log_2 f_{\boldsymbol{G}}(\boldsymbol{S}) \}$ taken over $f_{\boldsymbol{S}}(\boldsymbol{s})$

$$\begin{aligned}
\mathrm{E}\{ -\log_2 f_{\boldsymbol{G}}(\boldsymbol{S}) \} &= -\int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s}) \log_2 f_{\boldsymbol{G}}(\boldsymbol{s}) \,\mathrm{d}\boldsymbol{s} \\
&= \frac{1}{2}\log_2\Big((2\pi)^N|\boldsymbol{C}_N|\Big) + \frac{\log_2 e}{2}\,\mathrm{E}\Big\{ (\boldsymbol{S}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}_N^{-1}(\boldsymbol{S}-\boldsymbol{\mu}) \Big\} \\
&= \frac{1}{2}\log_2\Big((2\pi e)^N |\boldsymbol{C}_N| \Big)
\end{aligned} \tag{92}$$

➡ Same expression as for $h_N^G(\boldsymbol{G}) = \mathrm{E}\{ -\log_2 f_{\boldsymbol{G}}(\boldsymbol{G}) \}$

Inequality for Given Covariance Function

- N-th order differential entropy for stationary random process $\boldsymbol{S}$

$$
\begin{aligned}
h_N(\boldsymbol{S}) &= \mathrm{E}\{-\log_2 f_{\boldsymbol{S}}(\boldsymbol{S})\} \\
&= \mathrm{E}\{-\log_2 f_{\boldsymbol{S}}(\boldsymbol{S}) + \log_2 f_{\boldsymbol{G}}(\boldsymbol{S}) - \log_2 f_{\boldsymbol{G}}(\boldsymbol{S})\} \\
&= \mathrm{E}\{-\log_2 f_{\boldsymbol{G}}(\boldsymbol{S})\} - \mathrm{E}\left\{\log_2 \frac{f_{\boldsymbol{S}}(\boldsymbol{S})}{f_{\boldsymbol{G}}(\boldsymbol{S})}\right\} \\
&= \frac{1}{2}\log_2\left((2\pi e)^N\,|\boldsymbol{C}_N|\right) - D\left(f_{\boldsymbol{S}}\,||\,f_{\boldsymbol{G}}\right) \qquad (93)
\end{aligned}
$$

- Applying the divergence inequality $D(f\,||\,g) \geq 0$ (equality iff $f = g$) yields

$$
h_N(\boldsymbol{S}) \leq \frac{1}{2}\log_2\left((2\pi e)^N\,|\boldsymbol{C}_N|\right) = h_N^G(\boldsymbol{S}) \qquad (94)
$$

➡ **Gaussian process has the largest N-th order differential entropy among all stationary random processes with the same covariance function**

Inequalities for Differential Entropy Rate

1 Differential entropy rate for given auto-covariance function $\phi(k)$

$$\bar{h}(\boldsymbol{S}) = \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} \leq \lim_{N\to\infty} \frac{h_N^G(\boldsymbol{S})}{N}$$

$$\bar{h}(\boldsymbol{S}) \leq \bar{h}^G(\boldsymbol{S}) = \frac{1}{2}\log_2(2\pi e) + \frac{1}{2}\lim_{N\to\infty} \log_2 \sqrt[N]{|\boldsymbol{C}_N|} \tag{95}$$

➡ **Maximized for Gaussian process**

2 Differential entropy rate for given variance $\sigma^2 = \phi(0)$

$$\bar{h}(\boldsymbol{S}) = \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} \leq \lim_{N\to\infty} \frac{h_N^G(\boldsymbol{S})}{N} \leq \lim_{N\to\infty} \frac{h_N^{Giid}(\boldsymbol{S})}{N}$$

$$\bar{h}(\boldsymbol{S}) \leq \bar{h}^{Giid}(\boldsymbol{S}) = \frac{1}{2}\log_2(2\pi e\sigma^2) \tag{96}$$

➡ **Maximized for Gaussian iid process**

Summary

Mutual Information

- Information that a random variable carries about another random variable
- Discrete random variables A: $I(A;B) = H(A) - H(A\,|\,B)$
- Continuous random variables A: $I(A;B) = h(A) - h(A\,|\,B)$

Fundamental Source Coding Theorem

- Greatest lower bound for lossy coding: Rate-Distortion Function $R(D)$

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\delta(g_N)\le D} \frac{I(g_N)}{N}$$

- $R(D)$ is convex function
- Special case (discrete sources with $D = 0$): Lossless coding theorem

Differential Entropy

- N-th order differential entropy $h_N(\boldsymbol{S})$ / differential entropy rate $\bar{h}(\boldsymbol{S})$
- Given $\phi(k)$: $h_N(\boldsymbol{S})$ and $\bar{h}(\boldsymbol{S})$ are maximized for Gaussian process
- Given σ^2: $h_N(\boldsymbol{S})$ and $\bar{h}(\boldsymbol{S})$ are maximized for Gaussian iid process