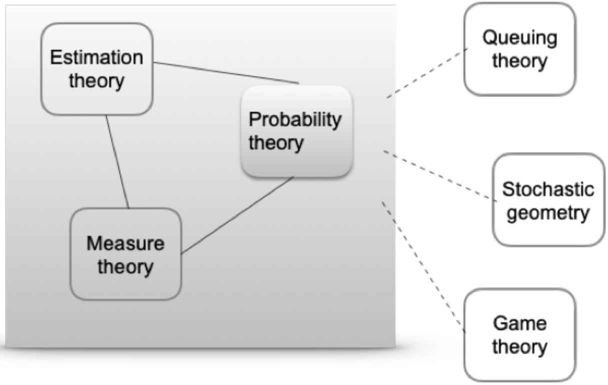
Selected topics in IT for communications

Jean-Marie Gorce
CITI, INSA Lyon, INRIA, Univ. Lyon
jean-marie.gorce@insa-lyon.fr

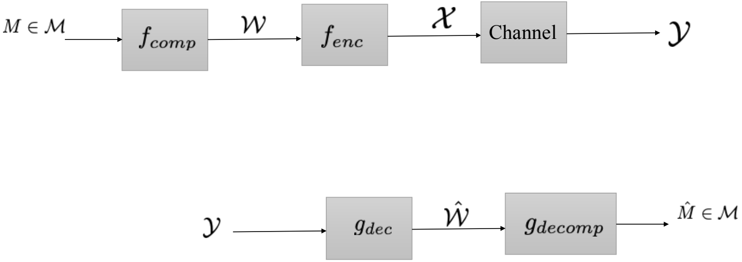
13, October, 2020

Table of contents

Information theory or communication theory ?

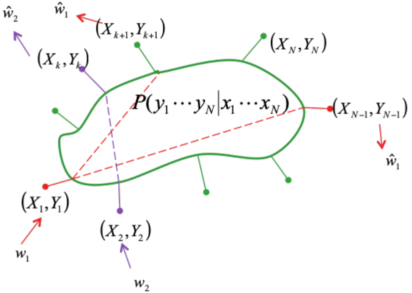


The standard P2P model



- Compression (source coding) aims at reducing the number of bits : $H(M)$.
- Coding (channel cod) adds redondancy to protect from errors $H(W) \leq C_{channel}$.
- Source and channel coding problems are separable (in the asymptotic regime).

Multi-user scenarios



- Multi-user compression/communication take advantage of spatial correlation.
- This is a hard problem in general.
- Fundamental limits are known only for some specific simplified scenarios.
- Challenging to comply with the explosion of decentralized networks : URLLC, caching, privacy,...

Some references

- The standard knowledge on IT for communications : *Elements of Information Theory* by Thomas Cover and Joy Thomas.

- An overview of problems related to Network Inf Theory : *Network Information Theory* by Abbas El Gamal et Young-han Kim.

- A modern vision on the IT method for communications : *Lecture notes on Informtion Theory* by Y. Polyanskiy and Y. Wu.
- A detailed analysis of the non asymptotic regime : *Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities* bu V. Tan.

Don't forget

- The path from fundamental limits to a practical code may be hard.

1948 : Claude Shannon founded ***Information Theory*** and provided, among many things his **second theorem on system capacity**:

claude shannon
1916-2016

$$C = W \cdot \log(1 + SNR)$$

But this limit was **achieved within 1bit** in 1993 only!!!

[turbo-codes, Berrou and Glavieux]

Maracas research group

Our research group is associated to Inria, located at Insa Lyon.

- Objective : contribute to the design of future communication systems.
- Researh line : from theory to practice.
- Hot topics : low latency, reliability, privacy, massive access.

Some research topics you may be interested in

- Theory : How can we deal with a large quantity of small information in a distributed system.
- Theory : Associate IT with other theoretical tools (stoch geometry, graphs, ...).
- Algorithms : machine learning is on the place. Quantum information is not far.
- Experimentation : around CorteXlab.

Table of contents

ide//

Information

Given, a countable measurable space $(\mathcal{X}, \mathcal{F})$
One probability measure P on $(\mathcal{X}, \mathcal{F})$
$X \sim P$ (discrete).

- Information
  (for a discrete variable only)

$$\imath_X(X) = \log_2\left(\frac{1}{P_X(X)}\right)$$

- Entropy : $H(X) = \mathbb{E}_X\left[\imath_X(X)\right]$.
- Varentropy : $V(X) = \mathrm{Var}\left(\imath_X(X)\right)$.

Throughout this document, $V(X)$ will be assumed finite.

Information

Given, a measurable space $(\mathcal{X}, \mathcal{F})$
Two probability measures P, Q on $(\mathcal{X}, \mathcal{F})$ with $P \ll Q$
$X \sim P$, $Y \sim Q$.

- Relative information

$$\imath_{X||Y}(x) = \log_2\left(\frac{dP}{dQ}(x)\right)$$

- Relative entropy :
$D(X||Y) = \mathbb{E}_X\left[\imath_{X||Y}(X)\right]$.

- Information density

$$\imath_{X;Y}(x;y) = \imath_{P_{XY}||P_X P_Y}(x;y)$$

- Mutual information :
$I(X;Y) = \mathbb{E}_{XY}[\imath_{X;Y}(X;Y)]$.

i.i.d sequence

Given a sequence of i.i.d. discrete random variables $\boldsymbol{S} = (X_1, X_2, \ldots, X_n)$.
By independency :

$$\imath_{\boldsymbol{S}}(\boldsymbol{S}) = \sum_{k=}^{n} \imath_X(X_k). \quad (1)$$

i.i.d sequence

Given a sequence of i.i.d. discrete random variables $\boldsymbol{S} = (X_1, X_2, \ldots, X_n)$.

By independency :

$$\imath_{\boldsymbol{S}}(\boldsymbol{S}) = \sum_{k=}^{n} \imath_X(X_k). \quad (1)$$

WLLN (foll. Chebyshev inequality) :

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{\boldsymbol{S}}{n} - H(X)\right| > \varepsilon\right] = 0, \quad (2)$$

for any $\varepsilon > 0$. When $n \to \infty$ the random vectors *concentrate* in a typical set.

i.i.d sequence (cont')

Second order properties (using CLN) :

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\boldsymbol{S} - nH(X)}{\sqrt{nV(X)}} < a\right] = \Phi(a), \tag{3}$$

where $\Phi(a)$ is the cdf of a r.v. $\sim \mathcal{N}(0,1)$.

i.i.d sequence (cont')

Second order properties (using CLN) :

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\boldsymbol{S} - nH(X)}{\sqrt{nV(X)}} < a\right] = \Phi(a), \tag{3}$$

where $\Phi(a)$ is the cdf of a r.v. $\sim \mathcal{N}(0,1)$. It means that :

$$\frac{\boldsymbol{S}}{n} \xrightarrow{d} \mathcal{N}\left(H(X), V(X)/n\right). \tag{4}$$

The information theory method

- A fundamental tradeoff is explored.
- Find a converse (the space of non achievable solutions).
- Find an achievable solution that approach the converse.
- If both matches, the fundamental limit is known.
- Explore constructive solution that approach the fundamental limit.

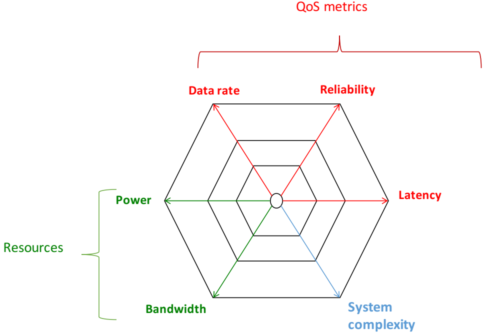


Table of contents

ide//

3 - Lossless compression

Single-shot compression

Dataflow compression (lossless)

Uniquely decodable codes

Lossless compressor

Consider :

- A single block source M defined on $\mathcal{M}$, a countable set.
- A coding space :
  $\mathcal{W} \subset \{0,1\}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \dots\}$.

Lossless compressor

Consider :

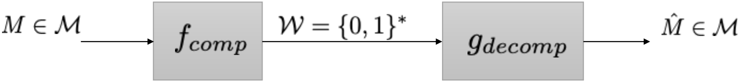
- A single block source M defined on $\boldsymbol{\mathcal{M}}$, a countable set.
- A coding space :
  $\boldsymbol{\mathcal{W}} \subset \{0,1\}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \dots\}$.

Definition 1 (Lossless compressor).

A lossless data compression code is a pair of mappings defined as :

- Compressor (encoder) : $f : \boldsymbol{\mathcal{M}} \to \boldsymbol{\mathcal{W}}$
- Decompressor (decoder) : $g : \boldsymbol{\mathcal{W}} \to \boldsymbol{\mathcal{M}}$.

such that $g(f(m)) = m$ (lossless constraint).



Remarks

- Lossless compression imposes that $\hat{m} = g(w)$ exists s.t. $\hat{m} = m$.
- Hence, $f(M)$ is injective : $|\mathcal{W}| \geq |\mathcal{M}|$.
- Single shot compressor : to compress M from a single observation.

Remarks

- Lossless compression imposes that $\hat{m} = g(w)$ exists s.t. $\hat{m} = m$.
- Hence, $f(M)$ is injective : $|\mathcal{W}| \geq |\mathcal{M}|$.
- Single shot compressor : to compress M from a single observation.

Definition 2 (Length function).

The length function $l : \mathcal{W} \to \mathbb{N}_1$ counts the number of bits for any element of $\{0, 1\}^*$.
We note the r.v. $L = l(f(M))$.

Remarks

- Lossless compression imposes that $\hat{m} = g(w)$ exists s.t. $\hat{m} = m$.
- Hence, $f(M)$ is injective : $|\mathcal{W}| \geq |\mathcal{M}|$.
- Single shot compressor : to compress M from a single observation.

Definition 2 (Length function).

The length function $l : \mathcal{W} \to \mathbb{N}_1$ counts the number of bits for any element of $\{0,1\}^*$.
We note the r.v. $L = l(f(M))$.

Optimal compressor : minimize some metric $Q = \mathbb{E}_M\left[q(l(f(M)))\right]$ ($q()$ is monotonically increasing).

Optimum compressor

Theorem 3 (Optimum lossless compressor).

WLOG, let assume that $P_M(1) \geq P_M(2) \geq \ldots P_M(|\mathcal{M}|)$.
The optimal compressor f^* *allocates the codes iteratively from* $m = 1$ *by assigning one of the smallest available codes to the message* m.
Then :

1. *The length of message* m *is given by* $l(f^*(m)) = \lfloor \log_2(m) \rfloor$.
2. *Any lossless compressor* $f(m)$ *verifies :*

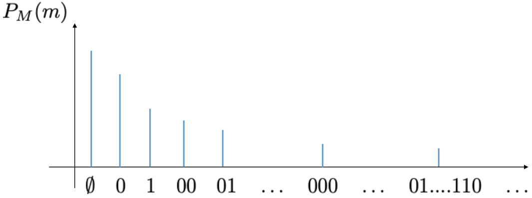
$$\mathbb{P}\left[l(f(M)) \leq a\right] \leq \mathbb{P}\left[l(f^*(M)) \leq a\right],$$

which can be written as $L \overset{st.}{\geq} L^*$.

L^* is the length random variable associated to the optimal compressor.

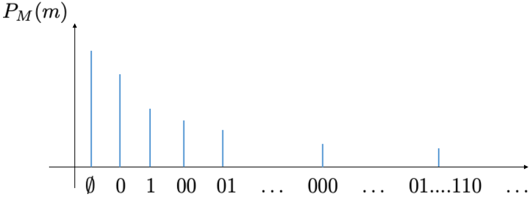
Illustration

The idea is to assign shorter codewords to most likely symbols :

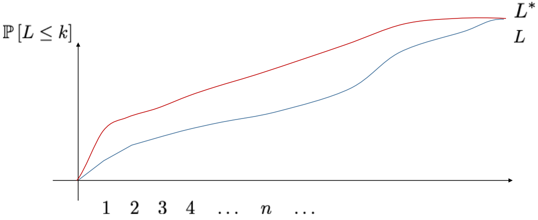


Illustration

The idea is to assign shorter codewords to most likely symbols :



Then, stochastic dominance $L \overset{st.}{\geq} L^*$ follows :



Proof of Th.3

- The first item comes directly from the compressor strategy. Nb of codes with $l(f(m)) < n : 2^n - 1$.
  Nb of codes with $l(f(m)) = n : 2^n$.

Proof of Th.3

- The first item comes directly from the compressor strategy. Nb of codes with $l(f(m)) < n : 2^n - 1$.
  Nb of codes with $l(f(m)) = n : 2^n$.
- The second item :
  1. For any compressor f : lossless $\Rightarrow$ Nb of msg s.t. $l(f(m)) < a$ to be $\leq 2^n - 1$.
  2. Note $\mathcal{A}^* = \{m; l(f^*(m)) \leq a\}$. It contains already the most prob. msg.
  3. Any permutation between $\mathcal{A}^*$ and $\overline{\mathcal{A}}^*$ $\Rightarrow$ reduction of $\mathbb{P}(\mathcal{A}^*)$.

Thus $Q = \mathbb{E}_M\left[q(l(f(M)))\right]$ is minimum for f^*.

Average length is (almost) Entropy

The av. nb of bits to store the message M is a useful operational metric. Its minimal value can be related to the entropy :

Average length is (almost) Entropy

The av. nb of bits to store the message M is a useful operational metric. Its minimal value can be related to the entropy :

Theorem 4 (Entropy vs minimum nb of bit).

The minimal average number of bits required to compress a single-block source M is given by

$$\overline{L}^* = \mathbb{E}_{L^*}[L^*] = \mathbb{E}_M[l(f^*(M))].$$

It is bounded by

$$H(M) - \log_2(e) - \log_2(H(M)+1) \leq \overline{L}^* \leq H(M).$$

First strong connection between a theoretical measure (entropy) and an operational one.

Proof of Theorem4

- RHS :
  1. from Th.3 : $l(f^*(m)) \leq \log_2(m)$.
  2. $P_M(m) \leq P_M(m-1) \Rightarrow P_m(M) \leq 1/m$ which means :

$$\log_2(m) \leq -\log_2(P_m(M)) = \imath_M(m).$$

  3. Then, for any m : $l(f^*(m)) \leq \imath_M(m)$, take expectation.

Proof of Theorem4 (cont)

- LHS : $H(M) = H(M, L^*) = H(M|L^*) + H(L^*)$
  1. Use $H(M|L^* = l) \leq l$. (2^l codes of length $l \Rightarrow$ entropy at most l).
  2. Then : $H(M|L^*) \leq \mathbb{E}_{L^*}[L]$
  3. For $H(L^*)$, use the following lemma (not proved here) with $Z = L^* + 1$:
     Given a r.v. Z on $\mathbb{N}_1$:
     $$H(Z) \leq h(1/\mathbb{E}[Z]) \cdot \mathbb{E}[Z]$$
     .
  4. Combining these equalities give the proof.

Bounds on the codelength distribution

Theorem 5 (Optimal length distribution).

Given the optimal lossless compressor f^*, the cdf of L^* is bounded by

$$\mathbb{P}\left[\imath_M(M) \leq k\right] \leq \mathbb{P}\left[L^* \leq k\right] \leq \mathbb{P}\left[\imath_M(M) \leq k + \tau\right] + 2^{-\tau+1}.$$

Proof : sketch-up

- LHS : same argument : $l(f^*(m)) \leq \imath_M(m)$.
- RHS : use conditionnng w.r.t. $\imath_M(M)$:
  1. : inf. spectrum splitting

$$\begin{aligned}\mathbb{P}\left[L^* \leq k\right] = &\ \mathbb{P}\left[L^* \leq k | \imath_M(M) \leq k + \tau\right] \cdot \mathbb{P}\left[\imath_M(M) \leq k + \tau\right] \\ &+ \mathbb{P}\left[L^* \leq k | \imath_M(M) > k + \tau\right] \cdot \mathbb{P}\left[\imath_M(M) > k + \tau\right] \\ &\leq \mathbb{P}\left[\imath_M(M) \leq k + \tau\right] + \mathbb{P}\left[L^* \leq k, \imath_M(M) > k + \tau\right].\end{aligned} \tag{5}$$

Bounds on the codelength distribution (cont')

2. The last term can be bounded as follows :

$$
\begin{aligned}
&\mathbb{P}\left[L^* \leq k, \imath_M(M) > k + \tau\right] \\
&= \sum_{m \in \mathcal{M}} P_M(m) \cdot \mathbb{1}_{\{l(f^*(m)) \leq k\}} \cdot \mathbb{1}_{\{\imath_M(m) > k+\tau\}} \\
&= \sum_{m \in \mathcal{M}} P_M(m) \cdot \mathbb{1}_{\{l(f^*(m)) \leq k\}} \cdot \mathbb{1}_{\left\{P_M(m) \leq 2^{-k-\tau}\right\}} \\
&\leq (2^{k+1} - 1) \cdot 2^{-k-\tau}.
\end{aligned}
\tag{6}
$$

3 - Lossless compression

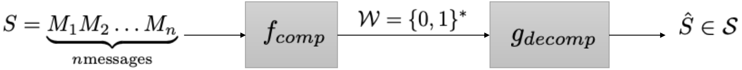
Problem formulation

$$S = \underbrace{M_1 M_2 \ldots M_n}_{n\text{messages}} \longrightarrow \boxed{f_{comp}} \xrightarrow{\mathcal{W}=\{0,1\}^*} \boxed{g_{decomp}} \longrightarrow \hat{S} \in \mathcal{S}$$

Dataflow : $\boldsymbol{S} = (M_1, \ldots, M_n) \in \boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{M}}^n$.

Without latency constraint, the compressor can apply the singleshot approach to the source $\boldsymbol{S}$ as a whole.

Problem formulation



Dataflow : $\boldsymbol{S} = (M_1, \ldots, M_n) \in \boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{M}}^n$.

Without latency constraint, the compressor can apply the singleshot approach to the source $\boldsymbol{S}$ as a whole.

Th.5 applies to $\boldsymbol{S}$. Taking $\tau = \sqrt{n}$ and $k = u \cdot n$:

$$\mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S})}{n} \leq u\right] \leq \mathbb{P}\left[\frac{L^*}{n} \leq u\right] \leq \mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S})}{n} \leq u + \frac{1}{\sqrt{n}}\right] + 2^{-\sqrt{n}+1}, \tag{7}$$

$\Rightarrow$ L^* and $\imath_{\boldsymbol{S}}(\boldsymbol{S})$ converge in distribution to a same random variable U_1.

Second order limit

With $\tau = n^{1/4}$ and $k = H(\boldsymbol{S}) + \sqrt{n} \cdot u$, Th.5 shows :

$$\mathbb{P}\left[\frac{L^* - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right] \geq \mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right]$$
$$\mathbb{P}\left[\frac{L^* - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right] \leq \mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}} \leq u + n^{-1/4}\right] + 2^{-n^{1/4}+1}. \tag{8}$$

$\Rightarrow \frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}}$ and $\frac{L* - H(\boldsymbol{S})}{\sqrt{n}}$ converge in distribution to a same random variable U_2.

Second order limit

With $\tau = n^{1/4}$ and $k = H(\boldsymbol{S}) + \sqrt{n} \cdot u$, Th.5 shows :

$$\mathbb{P}\left[\frac{L^* - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right] \geq \mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right]$$

$$\mathbb{P}\left[\frac{L^* - H(\boldsymbol{S})}{\sqrt{n}} \leq u\right] \leq \mathbb{P}\left[\frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}} \leq u + n^{-1/4}\right] + 2^{-n^{1/4}+1}. \tag{8}$$

$\Rightarrow \frac{\imath_{\boldsymbol{S}}(\boldsymbol{S}) - H(\boldsymbol{S})}{\sqrt{n}}$ and $\frac{L* - H(\boldsymbol{S})}{\sqrt{n}}$ converge in distribution to a same random variable U_2.

For an i.i.d source, the rate at the output of the decoder converges to :

$$R(\boldsymbol{S}) = \frac{L^*}{n} \sim \mathcal{N}\left(H(M), \frac{V(M)}{n}\right). \tag{9}$$

Homework

Consider a grey image of N_p pixels. Each pixel is coded with 8 bits. Compare the compression factor that can be achieved, if the pixels are coded independently or as a whole. The grey levels are assumed i.i.d. according to one of the following distributions :

1. uniform pmf.
2. some sampled exponential pmf.

Compute (numerically) the average length and plot the length distribution in each situation, as a function of n.

3 - Lossless compression

Single-shot compression

Dataflow compression (lossless)

Uniquely decodable codes

Code extension

Still a dataflow, but the messages M_n are compressed separatly, transmitted one after each other.

Definition 6 (Extension of a code).

Given a compressor $f : \mathcal{M} \to \mathcal{W}$ (see def.1), its extension is

$$f^+ : \mathcal{M}^+ \to \{0,1\}^*,$$

with $\mathcal{M}^+ = \bigcup_{n \geq 1} \mathcal{M}^n$ and
$f^+(m_1, m_2, \ldots, m_n) = (f(m_1), f(m_2), \ldots, f(m_n))$ (concatenation).
Note that $\mathcal{M}^+$ contains all non-empty finite-length strings of symbols from $\mathcal{M}$.

Uniquely decodable codes

- Uniquely decodable : guarantee the decodability of a sequence of codes.
- The compressors that fulfills this rule are called *uniquely decodable codes*.
- Prefix codes belong to this set codes :

Definition 7 (Prefix codes).

A compressor $f : \mathcal{M} \rightarrow \mathcal{W}$ (def.1) is a prefix code if no codeword is a prefix of another one.

Example :...

Kraft-McMillan inequality

Theorem 8 (Kraft-McMillan).

1. *Given $f : \boldsymbol{\mathcal{M}} \rightarrow \boldsymbol{\mathcal{W}}$ (def.1), uniquely decodable. Then f satisfies the Kraft inequality :*

$$\sum_{m \in \boldsymbol{\mathcal{M}}} 2^{-l_m} \leq 1,$$

where $l_m = l(f(m))$ is the code length.

2. *Conversly, for any set $\{\lambda(i); i \in \{1, 2, \ldots, |\boldsymbol{\mathcal{M}}|\}\}$, s.t. the Kraft inequality is verified, there exists a prefix code f, such that $l(f(m)) = \lambda(m)$.*

Proof of Th.8

Optimal uniquely decodable code

Definition 9 (Optimal average codelength for prefix codes).

The minimal average codelength among all prefix codes is the solution of the following problem

$$\overline{L}_{pc} \triangleq \min \sum_{m \in \mathcal{M}} P_m(M)\, l_m$$
$$\text{s.t.} \quad \sum_{m \in \mathcal{M}} 2^{-l_m} \leq 1$$
$$l_m \in \mathbb{N}_1.$$

This is a linear programming (IP) problem which can be solved with the Huffman algorithm.

Huffman algorithm

Bounds on the minimal average length

This optimal length can be bounded as follows :

Theorem 10 (fundamental bounds of the optimal codelength).

$$H(M) \leq \overline{L}_{pc} \leq H(M) + 1.$$

We can compare with the bounds found for L^* :

$$H(M) - \log_2(e) - \log_2 (H(M) + 1) \leq \overline{L}^* \leq H(M) \leq \overline{L}_{pc} \leq H(M) + 1.$$

Proof of Th10

1. Achievabilty ($\overline{L}_{pc} \leq H(M) + 1$) :
   Use the Shannon code :
   let define code lengths s.t. $l_m = \lceil \imath_M(m) \rceil$.
   It fulfills the Kraft inequality (prove it). Then, from the Kraft-McMillan theorem, such a code exists.

Proof of Th10 (cont')

2. Converse ($H(M) \leq \overline{L}_{pc}$) :
   Use the KL divergence between P_M and another prob. meas. on $\boldsymbol{\mathcal{M}}$: $Q_M(m) = \frac{2^{-l_m}}{\sum_{m \in \boldsymbol{\mathcal{M}}} 2^{-l_m}}$.
   Computes the KL divergence :

$$
\begin{aligned}
D(P_M || Q_M) &= \sum_{m \in \boldsymbol{\mathcal{M}}} P_M \log_2 \left(\frac{P_M}{Q_M} \right) \\
&= \sum_{m \in \boldsymbol{\mathcal{M}}} P_M \log_2 (P_M) - \sum_{m \in \boldsymbol{\mathcal{M}}} P_M \log_2 (Q_M) \\
&= -H(M) + \mathbb{E}_M [l(f(M))] + \log_2 \left(\sum_{m \in \boldsymbol{\mathcal{M}}} 2^{-l_m} \right).
\end{aligned}
\tag{10}
$$

The sum in the log is lower than 1, which concludes the proof.

Table of contents

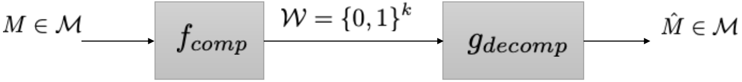
ide//

4 - Almost lossless compression

Fixed length codes

Multi-terminal compressors

Almost lossless codes



The compression/decompression constraint is now

$$g(f(m)) = \begin{cases} m & \text{if} \quad m \in \mathcal{D} \\ e & \text{ot.} \end{cases}$$

where $\mathcal{D}$ is the subset of decodable messages.

(k, ϵ)-code

Definition 11 ((k, ϵ)-code).

A compressor-decompressor code (f, g) is called a (k, ϵ)-code if

$$\begin{aligned} f: &\quad M \to \mathcal{W} \\ g: &\quad \mathcal{W} \to M \cup \{e\} \end{aligned}$$

with $\mathcal{W} = \{0, 1\}^k$, and such that $\mathbb{P}\left[g(f(M)) = e\right] \leq \epsilon$.

The fundamental limit is the minimal achievable error :

$$\epsilon^*(M, k) \triangleq \inf \left\{\epsilon; \exists (k, \epsilon)\text{-code for } M\right\}.$$

Minimal error

Given a source M, where the messages are ordered from the higest to the lowest probability.

Theorem 12 (Minimal error probability).

The min. error prob. for $k < \log_2(|\mathcal{M}|)$ *relies onto the optimal variable lenght compressor* $f^*(m)$ *(see def 3) :*

$$\begin{aligned}\epsilon^*(M,k) &= 1 - \sum_{m=1}^{2^k-1} P_M(m) \\ &= \mathbb{P}\left[l(f^*(M)) \geq k\right].\end{aligned}$$

Remarks

The proof is immediate. One code is reserved for the error message (all non detectable messages are compressed with this code), and the $2^k - 1$ messages with the highest probabilities are coded without errors.

- f maps the elements in $\mathcal{D}$ into the codes from $(00\ldots00)$ to $(11\ldots10)$.

Remarks

The proof is immediate. One code is reserved for the error message (all non detectable messages are compressed with this code), and the $2^k - 1$ messages with the highest probabilities are coded without errors.

- f maps the elements in $\mathcal{D}$ into the codes from $(00\ldots00)$ to $(11\ldots10)$.
- This theorem shows the link with the codelength distribution of the variable length coding.

Remarks

The proof is immediate. One code is reserved for the error message (all non detectable messages are compressed with this code), and the $2^k - 1$ messages with the highest probabilities are coded without errors.

- f maps the elements in $\mathcal{D}$ into the codes from $(00\ldots00)$ to $(11\ldots10)$.
- This theorem shows the link with the codelength distribution of the variable length coding.
- Alternatively, if a detectable error is not imposed, 2^k codes can be used for detectable messages and the non detectable messages can be coded by any of the other codes. Gain is marginal.

Shannon theorem I

Theorem 13 (Shannon source coding).

Consider a sequence $\boldsymbol{S} = (M_1, \ldots, M_n)$ *of i.i.d. messages. Then, the following holds :*

$$\lim_{n\to\infty} \epsilon^*(\boldsymbol{S}, nR) = \begin{cases} 0 & R > H(M) \\ 1 & R < H(M) \end{cases}$$

Shannon theorem I

Theorem 13 (Shannon source coding).
Consider a sequence $\boldsymbol{S} = (M_1, \dots, M_n)$ of i.i.d. messages. Then, the following holds :

$$\lim_{n \to \infty} \epsilon^*(\boldsymbol{S}, nR) = \begin{cases} 0 & R > H(M) \\ 1 & R < H(M) \end{cases}$$

Note $R \triangleq k$represents the source rate (in bits per channel use) at the output of the compressor.
The proof follows directly form Th.12 and from the bounds of L^* in (9).

Second order

A second order approximation can be also obtained from the CLT on L^*, leading to :

$$\lim_{n\to\infty} \epsilon^*(\boldsymbol{S}, nR) = 1 - \Phi\left(\frac{\sqrt{n}\cdot(R - H(M))}{\sqrt{V(M)}}\right). \tag{11}$$

Second order

A second order approximation can be also obtained from the CLT on L^*, leading to :

$$\lim_{n\to\infty} \epsilon^*(\boldsymbol{S}, nR) = 1 - \Phi\left(\frac{\sqrt{n}\cdot(R - H(M))}{\sqrt{V(M)}}\right). \qquad (11)$$

Coming back to the single shot compressor, the relations obtained in Th.5, allow to bound the error probability as follows :

$$\mathbb{P}\left[\imath_M(M) > k + \tau\right] - 2^{-\tau} \leq \epsilon^*(M, k) \leq \mathbb{P}\left[\imath_M(M) > k - \tau\right] + 2^{-\tau}, \qquad (12)$$

$\forall\tau > 0.$

Proofs

- Converse : follows directly Th.5.
- Achievability : random coding argument (the method will be detailed later for another proof).

Second order : refine achievability

It is worh noting that a better achievability can be obtained :

$$\epsilon^*(M,k) \leq \mathbb{P}\left[\imath_M(M) \geq k\right]. \tag{13}$$

The proof follows directly from the achievability proof of Th.5.

Remarks

- The random coding argument based achievzbility wad initial Shannon's proof.
- This later is connected to the AEP for memoryless sources, concentration inequalities.
- Defining the typicality set :

$$T_n^\delta = \left\{ s = m^n : \left| \frac{1}{n} \imath_{\boldsymbol{S}}(s) - H(\boldsymbol{S}) \right| \leq \right\}$$

where $|T_n^\delta| \leq 2^{(H(\boldsymbol{S})+\delta)n} \ll |\boldsymbol{\mathcal{M}}|^n$.
Then $\mathbb{P}\left[\boldsymbol{S} \in T_n^\delta\right] \to 1$ as $n \to \infty$.

Notes on linear codes

- Among all possible codes, linear codes have interesting properties.
- Why : in order to find strutured codes (making decoding easier and faster).
- Key result : asysmptotically, linear codes can achieve the fundamental bounds.

Further work on this : coding theory in Galois fields.

4 - Almost lossless compression

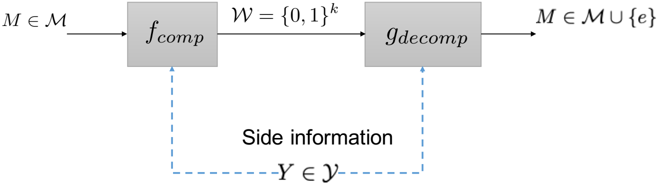
Fixed length codes

Multi-terminal compressors

Side information available at both sides

> Idea : what can we do if the two sides share a common information ?

Note : when a message is sent an explicite (correlated with M) information is know at both sides.



$$P_{MY} \neq P_M \cdot P_Y$$

Side information : definition

Definition 14 (Compression with side information).

Given P_{MY}, with M a disrete random variable.
A compressor with side information is given by :

- $f : \mathcal{M} \times \mathcal{Y} \to \mathcal{W}$
- $g : \mathcal{W} \times \mathcal{Y} \to \mathcal{M} \cup \{e\}$

such that $\mathbb{P}\left[g(f(M,Y),Y) \neq M\right] < \epsilon$
with $\mathcal{M} = \{0,1\}^K$.

Side information : definition

Definition 14 (Compression with side information).

Given P_{MY}, with M a disrete random variable.
A compressor with side information is given by :

- $f : \mathcal{M} \times \mathcal{Y} \to \mathcal{W}$
- $g : \mathcal{W} \times \mathcal{Y} \to \mathcal{M} \cup \{e\}$

such that $\mathbb{P}\left[g(f(M, Y), Y) \neq M\right] < \epsilon$
with $\mathcal{M} = \{0, 1\}^K$.

The fundamental limit is

$$\epsilon^*(M|Y, k) = \inf\left\{\epsilon; \exists(k, \epsilon)\text{-code}\right\} \tag{14}$$

Fundamental limits

Theorem 15 (Shannon source coding w side inf.).
Consider a sequence $(\boldsymbol{S}, \boldsymbol{T}) = (M_1, Y_1), \ldots, (M_n, Y_n)$ of i.i.d. messages. Then, the following holds :

$$\lim_{n\to\infty} \epsilon^*(\boldsymbol{S}|\boldsymbol{T}, nR) = \begin{cases} 0 & R > H(M|Y) \\ 1 & R < H(M|Y) \end{cases}$$

and for the error bounds, from (12), one have :

$$\mathbb{P}\left[\imath_{M|Y}(M|Y) > k + \tau\right] - 2^{-\tau} \leq \epsilon^*(M|Y, k) \leq \mathbb{P}\left[\imath_{M|Y}(M|Y) > k - \tau\right] + 2^{-\tau}, \tag{15}$$

$\forall \tau > 0$.

Proof

Apply Theorem 13 with the channel $P_{M|Y}$ instead of P_M :

- Start from the bounds obtained from P2P channel.
- Consider the optimal compressor under $P_{M|Y=y}$. That means that the compressor changes with y.
- Average over P_Y.

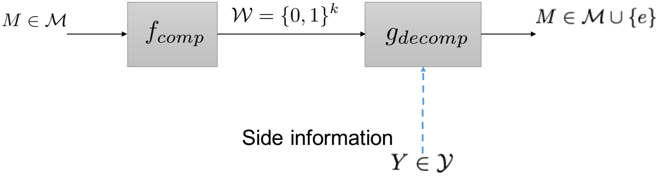
Application

If both Tx and Rx know a common information related to the source, they can take benefit from it.
Example : Consider images, wher both sides share some side information about the input distribution. This side information can be the realisation of the previous image.
The codebook can be modified according to y.

Slepian-Wolf : Side information available at the receiver only

Idea : How much is lost when only the receiver knows y ?

Note : the receiver knows some observation y that the transmitter does not know.



$$P_{MY} \neq P_M \cdot P_Y$$

Slepian-Wolf : definition

Definition 16 (Slepian-Wolf compression).

Given P_{MY}, with M a discrete random variable.
A Slepian-Wolf compressoris given by :

- $f : \mathcal{M} \rightarrow \mathcal{W}$
- $g : \mathcal{W} \times \mathcal{Y} \rightarrow \mathcal{M} \cup \{e\}$

such that $\mathbb{P}\left[g(f(M), Y) \neq M\right] < \epsilon$
with $\mathcal{W} = \{0, 1\}^k$.

Slepian-Wolf : definition

Definition 16 (Slepian-Wolf compression).

Given P_{MY}, with M a discrete random variable.
A Slepian-Wolf compressoris given by :

- $f : \mathcal{M} \to \mathcal{W}$
- $g : \mathcal{W} \times \mathcal{Y} \to \mathcal{M} \cup \{e\}$

such that $\mathbb{P}\left[g(f(M), Y) \neq M\right] < \epsilon$
with $\mathcal{W} = \{0, 1\}^k$.

The fundamental limit is

$$\epsilon^*_{SW}(M|Y, k) = \inf \{\epsilon; \exists (k, \epsilon)\text{-code}\} \tag{16}$$

Note that undetected errors are allowed.

Fundamental limits

Surprisingly : no loss ! !

Theorem 17 (Slepian-Wolf bounds).

Consider a sequence $(\boldsymbol{S}, \boldsymbol{T}) = (M_1, Y_1), \ldots, (M_n, Y_n)$ *of i.i.d. messages. Then, the following holds :*

$$\lim_{n \to \infty} \epsilon^*_{SW}(\boldsymbol{S}|\boldsymbol{T}, nR) = \begin{cases} 0 & R > H(M|Y) \\ 1 & R < H(M|Y) \end{cases}$$

Fundamental limits

Surprisingly : no loss ! !

Theorem 17 (Slepian-Wolf bounds).

Consider a sequence $(\boldsymbol{S}, \boldsymbol{T}) = (M_1, Y_1), \ldots, (M_n, Y_n)$ *of i.i.d. messages. Then, the following holds :*

$$\lim_{n\to\infty} \epsilon^*_{SW}(\boldsymbol{S}|\boldsymbol{T}, nR) = \begin{cases} 0 & R > H(M|Y) \\ 1 & R < H(M|Y) \end{cases}$$

and for the error bounds, one have :

$$\epsilon^*(M|Y,k) \leq \epsilon^*_{SW}(M|Y,k) \leq \mathbb{P}\left[\imath_{M|Y}(M|Y) > k - \tau\right] + 2^{-\tau}, \tag{17}$$

$\forall \tau > 0$.

Proof

- Converse : obvious
- Achievability : random codes approach :
  1. Generate a random codebook $\mathcal{C} = \{c_m \in \mathcal{W}; m \in \mathcal{M}\}$, independently of M, Y : $f(m) = c_m$.
  2. Let defined :
     $J(m, \mathcal{C}|y) \triangleq \{m' \neq m; c_{m'} = c_m, \imath_{M|Y}(m'|y) < k - \tau\}$
  3. And : $J(M, \mathcal{C}|Y) = \sum_{m,y} P_{MY}(m, y) \mathbb{1}_{\{J(m,\mathcal{C}|y) \neq 0\}}$.
  4. Decoding rule :

$$g(w, y) = \begin{cases} m & \exists! m \quad s.t. c_m = w, \imath_{M|Y}(m|y) < k - \tau \\ 0 & othw. \end{cases}$$

item The error associated to this codebook is :

$$\epsilon(\mathcal{C}) = \mathbb{P}\left[\imath_{M|Y}(M|Y) \geq k - \tau \mathrm{or} J(M, \mathcal{C}|Y) \neq 0\right] \tag{18}$$

Proof (cont')

- This error can be bounded (union bound) :

$$\epsilon(\mathcal{C}) \leq \underbrace{\mathbb{P}\left[\imath_{M|Y}(M|Y) \geq k - \tau\right]}_{\text{outage}} + \underbrace{\mathbb{P}\left[J(M, \mathcal{C}|Y) \neq 0\right]}_{\text{confusion}} \quad (19)$$

The confusion is upper bounded, considering
  - The number of codes belonging to the decodable set is at most $2^{k-\tau}$.
  - Averaging over all random codebooks : $\mathbb{P}\left[c_m = c_{m'}\right] = 2^{-k}$.

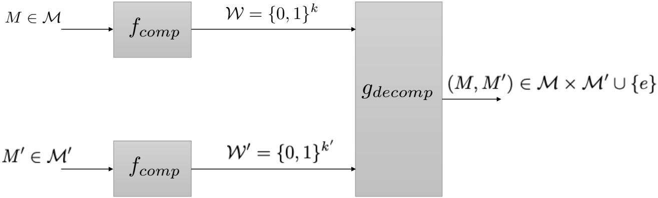
This concludes the proof.

Now, if a random code achieves this bound, there exists a deterministic code that achieves it as well.

The randomisation, and the duality confusion/outage is an usual approach in IT proofs.

Multi-terminals extension

Can we deal with correlated sources compressed independently ?



Multi-terminal Slepian-Wolf

Definition 18 (Multi-terminal compression).

Given $P_{MM'}$, with M, M' discrete random variables.
A biterminal compressor is :

- $f_1 : \mathcal{M} \to \mathcal{W}$
- $f_2 : \mathcal{M}' \to \mathcal{W}'$
- $g : \mathcal{W} \times \mathcal{W}' \to \mathcal{M} \times \mathcal{M}' \cup \{e\}$

such that $\mathbb{P}\left[(\hat{M}, \hat{M}') \neq (M, M')\right] \leq \epsilon$
with $\mathcal{W} = \{0,1\}^k$, $\mathcal{W}' = \{0,1\}^{k'}$.
The fundamental limit is

$$\epsilon^*(M, M', k, k') = \inf\left\{\epsilon; \exists(k, k', \epsilon)\text{-code}\right\}. \tag{20}$$

Fundamental limits

Theorem 19 (Multi-terminal bounds).

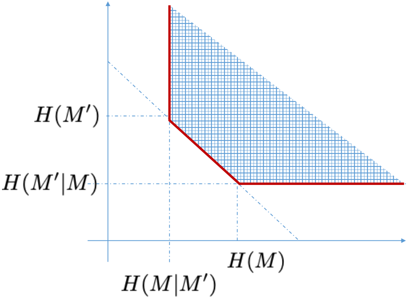
Consider a sequence $(\boldsymbol{S}, \boldsymbol{T}) = (M_1, M_1'), \ldots, (M_n, M_n')$ *of i.i.d. messages. Then, the following holds :*

$$\lim_{n\to\infty} \epsilon^*_{SW}(\boldsymbol{S}, \boldsymbol{T}, n, R, R') = \begin{cases} 0 & (R, R') \in int(\boldsymbol{\mathcal{R}}_{SW}) \\ 1 & (R, R') \notin \boldsymbol{\mathcal{R}}_{SW} \end{cases},$$

where the Slepian-Wolf region is :

$$\boldsymbol{\mathcal{R}}_{SW} = \{x, y\}; \begin{cases} x \geq H(M|M') \\ y \geq H(M'|M) \\ x + y \geq H(M, M') \end{cases}$$

Achievable region : converse



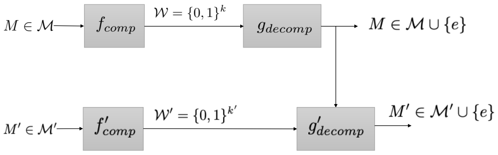
Converse :

- $R < H(M|M')$: this bound holds with side information at both sides.
- $R' < H(M'|M)$: idem.
- $R + R' < H(M, M')$: compression cannot do better than joint compression.

Achievability proof

Achievability :

- Achieve corner points by sort of successive decompressor.
- Combine these two compressors by time sharing : alternate codebooks with the two strategies determinstically or randomly (common randomness).



> The idea of time-sharing is usual. Convexity of achievable rate regions is valid in many scenarios.

Remarks

- Slepian-Wolf opened a lot of fundamental applications (security, water-marking, differential compression).
- The non asymptotic regime received a lot of recent attention, see *The dispersion of Slepian-Wolf coding* by V. Tan and O. Kosut (2012).
- Recent use for immersive video application, see *Rate-storage regions for Extractable Source Coding with side information* by E. Dupraz, A. Roumy, T. Maugey and M. Kieffer (2020).

Table of contents

ide//

- Lossy compression is necessary for continuous signals, and anyway useful for discrete signals (jpg, mpeg,...).
- A continous signal takes values on a continuous space and is defined on a continuous time $s(t)$.
- The discretiation involves two steps : sampling and quantization.
- Sampling theory is out of the scope of this course (rely on Fourier transform, filtering,...).
- Introduction to quantization in IT : Rate-distortion theory.
- Again a fundamental tradeoff : rate versus distortion.

General ideas on quantization

- Scalar quantizer.
- Vector quantizer (Lloyd's algorithm).

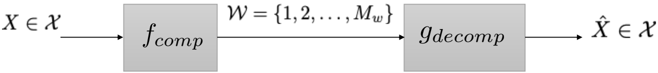
Hamming game

Given 100 unbiased bits, drawn indepndently.
What is the maximum information we can store on 50bits ?
From the 50 bits, you want to guess the 100 bit values.

- Scalar coding : store 50 values, and guess the other randomly : $P_{err} = 25\%$.
- Vectorial coding : thanks to concentration of measures, we can optimize the joint coding and obtain $P_{err} = 11\%$.

Illustration.

Lossy compressor setup



Definition 20 (Lossy compressor).

A lossy compressor/decompressor pair (f, g) is defined by :

- A compressor : $f : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{W}}$, with $w = f(x)$
- A decompressor : $g : \boldsymbol{\mathcal{W}} \to \hat{\boldsymbol{\mathcal{X}}}$, with $\hat{x} = g(w)$
- a distortion metric (loss function) : $d(X, \hat{X}) : \boldsymbol{\mathcal{X}} \times \hat{\boldsymbol{\mathcal{X}}} \to \mathbb{R} \cup \{+\infty\}$
- a cost function $Q(d) = \mathbb{E}_{X\hat{X}} \left[q(d(X, \hat{X})) \right]$ that measures the performance of the compressor.

where $\boldsymbol{\mathcal{W}}$ is a countable set, and $\boldsymbol{\mathcal{X}}$is a continuous set.

Rate-distortion formulations

Different possible settings :

- fixed length, average distortion : $W \in \{1, 2, \ldots, M_w\}$, $Q_{moy}(d) = \mathbb{E}_{X\hat{X}}\left[d(X, \hat{X})\right]$.
- fixed length, excess distortion : $W \in \{1, 2, \ldots, M_w\}$, $Q_{exc}(d) = \mathbb{E}_{X\hat{X}}\left[\mathbb{1}_{\{d(X,\hat{X})>D\}}\right]$.
- variable length, max distortion : $W \in \{0, 1\}^*$, and $\min(H(W))$, s.t. $Q_{exc}(d) = 0$.

Rate-distortion formulations

Different possible settings :

- fixed length, average distortion : $W \in \{1, 2, \ldots, M_w\}$, $Q_{moy}(d) = \mathbb{E}_{X\hat{X}}\left[d(X, \hat{X})\right]$.
- fixed length, excess distortion : $W \in \{1, 2, \ldots, M_w\}$, $Q_{exc}(d) = \mathbb{E}_{X\hat{X}}\left[\mathbb{1}_{\{d(X,\hat{X})>D\}}\right]$.
- variable length, max distortion : $W \in \{0, 1\}^*$, and $\min(H(W))$, s.t. $Q_{exc}(d) = 0$.

Definition 21 (Separable distortion metric).

For a sequence of $\boldsymbol{\mathcal{X}}$-valued random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, a separable distortion metric is defined as the average of single-letter distortions :

$$d(\boldsymbol{x}, \hat{\boldsymbol{x}}) \triangleq \frac{1}{n} \sum_{k=1}^{n} d(x_k, \hat{x}_k).$$

Lossy compressor to be studied

Definition 22 ((n, M, D)-code).

An (n, M, D)-codeis a lossy compressor with $\mathcal{W} = \{1, 2, \ldots, M\}$ and average distortion metric.

The fundamental limit (operational metric) is given by :

$$M^*(n, D) = min \{M : \exists (n, M, D)\text{-code}\}$$
$$R(D) := \lim_{n\to\infty} \sup \frac{1}{n} \log_2 M^*(n, D) \tag{21}$$

Minimal information

Definition 23 (Minimal information under distortion constraint).

Given a continuous source $X \in \boldsymbol{\mathcal{X}}$, the minimal mutual information under distortion constraint D is defined as :

$$\varphi_X(D) = \inf_{P_{Y|X};\mathbb{E}_{X,Y}[d(X,Y)]\leq D} I(X;Y), \tag{22}$$

where Y is defined on some domain $\boldsymbol{\mathcal{Y}}$, with the unique constraint that $d(X,Y)$ is defined.

Let be denoted the minimal distortion such that a *quantization* is feasible :

$$D_0 = \inf \{D; \varphi_X(D) < \infty\}$$

The solution Y for which $\varphi_X(D)$ is achieved, contains the minimal information on X, such that the distortion is bounded.

Properties of $\varphi_X(D)$

1. φ_X is convex, non increasing.
   thanks to convexity of $P_{Y|X} \to I(P_X, P_{Y|X})$.

Properties of $\varphi_X(D)$

1. φ_X is convex, non increasing.
   thanks to convexity of $P_{Y|X} \to I(P_X, P_{Y|X})$.
2. φ_X is continuous on (D_0, ∞).
   follows form convexity.

Properties of $\varphi_X(D)$

1. φ_X is convex, non increasing.
   thanks to convexity of $P_{Y|X} \to I(P_X, P_{Y|X})$.
2. φ_X is continuous on (D_0, ∞).
   follows form convexity.
3. If $d(x,x) = D_0$ and $d(x, y \neq x) > D_0$, then $\varphi_X(D_0) = I(X;X)$.
   all information is needed.

Properties of $\varphi_X(D)$

1. φ_X is convex, non increasing.
   thanks to convexity of $P_{Y|X} \to I(P_X, P_{Y|X})$.
2. φ_X is continuous on (D_0, ∞).
   follows form convexity.
3. If $d(x,x) = D_0$ and $d(x, y \neq x) > D_0$, then
   $\varphi_X(D_0) = I(X;X)$.
   all information is needed.
4. Given $D_{max} \triangleq \inf_{\hat{x}} \mathbb{E}_X\left[d(X,\hat{x})\right]$ then $\varphi_X(D > D_{max}) = 0$.
   no information is needed.

Information Rate-Distortion

Theorem 24 (Single letterization).

Given an i.i.d. sequence $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, and a separable distortion

$$R_i(D) = \varphi_X(D)$$

where

Definition 25 ($R_i(D)$).

The information rate-distortion function is defined by

$$R_i(D) := \lim_{n\to\infty} \sup \frac{1}{n} \varphi_{\boldsymbol{X}}(D)$$

Proofs

- Achievability : choose $P_{\boldsymbol{Y}|\boldsymbol{X}} = P^n_{Y|X}$, then $\varphi_{\boldsymbol{X}}(D) \leq n\varphi_X(D)$ (M.I property)

Proofs

- Achievability : choose $P_{\boldsymbol{Y}|\boldsymbol{X}} = P^n_{Y|X}$, then $\varphi_{\boldsymbol{X}}(D) \leq n\varphi_X(D)$ (M.I property)
- Converse :

$$
\begin{aligned}
I(\boldsymbol{X}, \hat{\boldsymbol{X}}) &\geq \sum_{j=1}^{n} I(X_i, \hat{X}_i) \\
&\geq \sum_{j=1}^{n} \varphi_X(\mathbb{E}_{\cdot}\left[d(X_i, \hat{X}_i)\right]) \\
&\geq n\varphi_X\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}_{\cdot}\left[d(X_i, \hat{X}_i)\right])\right) \\
&\geq n\varphi_X(D)
\end{aligned}
$$

Shannon's rate distortion theorem

Theorem 26 (rate distortion limit).

Given a i.i.d. sequence $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ (stationary and memoryless), given a non negative and separable distortion metric $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$, such that $D_{max} < \infty$, and a target distortion $D > D_0$, Then

$$R(D) = R_i(D) = \inf_{P_{\hat{X}t|X};\mathbb{E}_{X\hat{X}}\left[d(X,\hat{X})\right]\leq D} I(X;\hat{X})$$

Shannon's rate distortion theorem

Theorem 26 (rate distortion limit).

Given a i.i.d. sequence $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ (stationary and memoryless), given a non negative and separable distortion metric $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$, such that $D_{max} < \infty$, and a target distortion $D > D_0$, Then

$$R(D) = R_i(D) = \inf_{P_{\hat{X}t|X}; \mathbb{E}_{X\hat{X}}[d(X,\hat{X})] \leq D} I(X; \hat{X})$$

This theorem links the operational and the information rate-distortion functions.

General converse proof

The converse is immediate from information theory properties :

$$
\begin{aligned}
\log_2(M) &\geq H(W) \quad \text{entropy max for equiprobable source} \\
&\geq I(X;W) \quad \text{M.I. always lower than entropy} \\
&\geq I(X;\hat{X}) \quad \text{processing reduces entropy} \\
&\geq \varphi_X(D). \quad \text{def of} \varphi.
\end{aligned}
\tag{23}
$$

Intuition on the achievability

- Random coding argument for a sequence : create a random codebook $\mathcal{C}\{c_1, \ldots, c_M\}$, drawn i.i.d. $\sim Q^n_{\hat{X}}$ on $\boldsymbol{\mathcal{X}}$.
- Define the following compressor-decompressor pair :
  - $f(x^n) = \arg\min_w d(x^n, c_w)$.
  - $g(w) = c_w$.
- Consider the excess ditortion metric : $P_{err} \triangleq \mathbb{P}\left[\forall c_w \in \mathcal{C}, d(\boldsymbol{X}, c_w p) > D\right]$.
- which can be bounded : $P_{err} \approx (1 - \mathbb{P}\left[d(\boldsymbol{X}, \hat{X}tb) \leq D\right]^M$.
- Then using large deviation, take enough codes to make this probability low (i.e ; to $I(S; \hat{S}) + \delta$).
- Then choose $Q_{\hat{X}}$ which minimizes this number and give the achievability.
- Compare excess and average distorsions.

Homework

- Don't forget to answer the question asked at last course.
- Write the complete proof of Shannon's rate-distortion theorem, in a very comprehensive way, using the definitions and theorems presented in the course and the first part.
- You can start from the lecture notes of Yuri Polianskiy, but write it from your own, with proper justification of all steps.
- Latex file is preferable.
- Illustrations are welcome.
- As well as pedagogic comments on the different calculation steps.

Application example : Gaussian

- A Gaussian source is defined on $\boldsymbol{\mathcal{X}} = \mathbb{R}$, and $X \sim \mathcal{N}\left((,0), \sigma^2\right)$.
- We want to minimize MSE distortion : $d(x, \hat{x}) = |x - \hat{x}|^2$.
- Result : $R(D) = \frac{1}{2} \log^+ \frac{\sigma^2}{D}$.

Tips : this gives the rate required to achieve some distortion target.

- The proof uses an achievability/converse approach.
- The achievability uses a code such that $S = \hat{S} + Z$.
- The converse uses a change of measure.

Recent references

- *Fixed-length lossy compression in the finite blocklength regime*, by Kostina, V., and Verdú, S. (2012).
- *Rate-distortion performance of lossy compressed sensing of sparse sources* by Leinonen, M., Codreanu, M., Juntti, M., and Kramer, G. (2018).

Table of contents

ide//

What has been done so far

Focus on source coding :

- Show the path from theoretical bounds to operational bounds.
- Use classical techniques in IT : achievabnility/converse, use of change of measure, thresholding the information measure, outage versus confusion.
- Introduction to multi-user settings.

Homework :

- Two problems have been introduced.
- Few references have been reported. Some fo you may choose a paper related to this topic.
- Next week : go to channel coding.