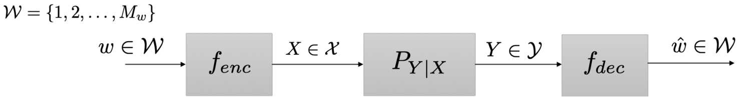
Selected topics in IT for communications - part II

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

20, October, 2020

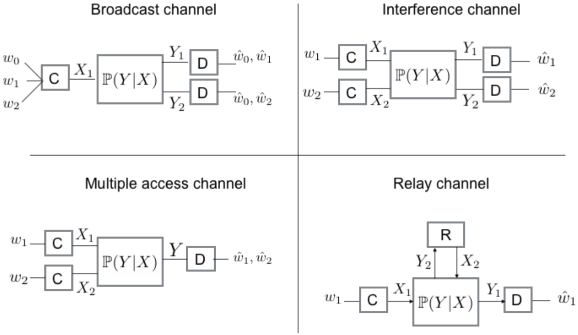
Table of contents

The standard P2P channel model



- Assumptions :
  1. The source is compressed, $W \sim \mathcal{U}$ on $\boldsymbol{\mathcal{W}}$.
  2. All messages are equiprobable, the source is discrete.
- Fundamental questions :
  1. What is the maximal number of channel uses we may need to transmit a message reliably ?
  2. What is tha maximal data rate in a given channel ?
  3. What is the fundamental latency-error tradeoff ?

Multi-user scenarios

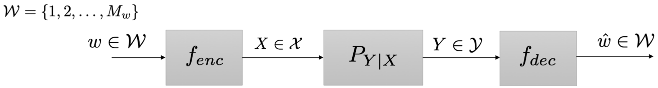


- Multi-user transmission explores the optimla usage of ressources.
- This is a hard problem in general.
- Fundamental limits are known only for some specific simplified scenarios.
- Challenging to comply with the explosion of decentralized networks : URLLC, caching, privacy

Table of contents

ide//

Channels from IT perspective



Definition 1 (P2P channel).

A point to point (P2P) single-shot communication channel is defined by the tuple $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, where :

- The input domain can be discrete or absolutely continuous : measurable space $(\boldsymbol{\mathcal{X}}, \mathcal{F})$.
- The output domain can be discrete or absolutely continuous : measurable space $(\boldsymbol{\mathcal{Y}}, \mathcal{G})$.
- The conditional probability measure, $P_{Y|X}$ relies the variables $X \in \boldsymbol{\mathcal{X}}$ and $Y \in \boldsymbol{\mathcal{Y}}$.

Channel properties

The following properties are usually defined :

- When $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are countable sets, $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, p_{Y|X})$ is a discrete channel.
- When $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are continuous sets, $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, p_{Y|X})$ is a continuous channel.
- A multiple use channel is defined by n channel uses, with $\boldsymbol{\mathcal{X}} = \mathcal{X}^n, \boldsymbol{\mathcal{Y}} = \mathcal{Y}^n$ and $P_{\mathbf{Y}|\mathbf{X}} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{Y}}$.

Channel properties

The following properties are usually defined :

- When $\mathcal{X}$ and $\mathcal{Y}$ are countable sets, $(\mathcal{X}, \mathcal{Y}, p_{Y|X})$ is a discrete channel.
- When $\mathcal{X}$ and $\mathcal{Y}$ are continuous sets, $(\mathcal{X}, \mathcal{Y}, p_{Y|X})$ is a continuous channel.
- A multiple use channel is defined by n channel uses, with $\boldsymbol{\mathcal{X}} = \mathcal{X}^n, \boldsymbol{\mathcal{Y}} = \mathcal{Y}^n$ and $P_{\mathbf{Y}|\mathbf{X}} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{Y}}$.

And for a multiple use channel :

- A channel is memoryless if $P_{\mathbf{Y}|\mathbf{X}} = \prod_{k=1}^{n} P_{Y_k|X_k}$.
- A channel is memoryless and stationary if $P_{\mathbf{Y}|\mathbf{X}} = P_{Y|X}^n$.
- A discrete memoryless channel (DMC) is defined by a matrix $P_{Y|X}$ on $|\mathcal{X}| \times |\mathcal{Y}|$ (bipartite graph).

Asymptotic regime

Study of the asymptotic regime :

- What are the properties of the channel when $n \to \infty$.
- Define the rate as $R = \frac{\log_2(M_W)}{n}$.

Examples

The following channels are widely used :

- Binary symmetric channel : $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} = \{0,1\}$.
- Additive White Gaussian noise channel : $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} = \mathbb{R}$, $Y = X + Z$.

Information capacity

Definition 2.

For a random channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, the information capacity is defined by :

$$C = \max_{P_X} I(X, Y)$$

Remarks

- This is an information theory property. No operational meaning for now.
- this is a *single letter* expression. i.e. apply to a single-shot channel properties. We will see how it is also relevant for stationary memoryless channels.

Channel codes

Definition 3 (M-code).

An M-code for $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, is an encoder-decoder pair (f, g) of (possibly randomized) functions :

- encoder : $f : \mathcal{W} = \{1, \ldots, M\} \to \mathcal{X}$,
- decoder : $g : \mathcal{Y} \to \mathcal{W}$.

The underlying model is

$$W \xrightarrow{f} X \xrightarrow{P_{Y|X}} Y \xrightarrow{g} \hat{W}.$$

Remarks :

- if the fonction f is determinist, $f(w) = c_w$ are codewords and $\mathcal{C} = \{c_1, c_2, \ldots, c_M\}$ is the codebook.
- the decision region associated to each codeword is $\mathcal{D}_w = g^{-1}(w)$.

Reliability metrics

The following reliability metrics can be used :

- Average error probability : $P_e \triangleq \mathbb{P}\left[\hat{W} \neq W\right]$.
- Max. error probability : $P_{e,max} \triangleq \max_{w \in \mathcal{W}} \mathbb{P}\left[\hat{W} \neq w | W = w\right]$.
- Bit error probability : $P_b \triangleq \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}\left[s_j \neq \hat{s}_j\right]$, with $W = S^k \in \mathbb{F}_2^k$.

Table of contents

ide//

3 - P2P discrete channels

General properties

Single shot achievability bounds

Stationary memoryless channels

Error constrained codes

Definition 4 ((M, ϵ)-code).

A cchannel code (f, g) is called an (M, ϵ)-code for a channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ if

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

such that $P_e \leq \epsilon$.

Similarly an $(M, \epsilon)_{max}$-code for the maximum error, can be defined. The fundamental limit is the maximal achievable alphabet size :

$$M_\epsilon^* \triangleq \max \{M; \exists (M, \varepsilon)\text{-code}\}.$$

The maximal entropy of the source that can be transmitted under message error probability, is $H(W) \leq Q^* = \log(M^*)$.

Relation with hypothesis testing

What is the connection with hypothesis testing ?

1. Fix P_X (i.e. choose the encoding function $f : \boldsymbol{\mathcal{W}} \to \boldsymbol{\mathcal{X}}$.
2. What is the optimal decoder w.r.t average error ?

Relation with hypothesis testing

What is the connection with hypothesis testing ?

1. Fix P_X (i.e. choose the encoding function $f : \boldsymbol{\mathcal{W}} \rightarrow \boldsymbol{\mathcal{X}}$.
2. What is the optimal decoder w.r.t average error ?
   MAP (maximum a posteriori) :
   $g^*(y) = \arg\max_{w \in \boldsymbol{\mathcal{W}}} \mathbb{P}[W = w | Y = y]$.
   Associated error : $P_e(y) = \mathbb{P}[W \neq w | Y = y]$.

Relation with hypothesis testing

What is the connection with hypothesis testing ?

1. Fix P_X (i.e. choose the encoding function $f : \mathcal{W} \to \mathcal{X}$.
2. What is the optimal decoder w.r.t average error ?
   MAP (maximum a posteriori) :
   $g^*(y) = \arg\max_{w \in \mathcal{W}} \mathbb{P}[W = w | Y = y]$.
   Associated error : $P_e(y) = \mathbb{P}[W \neq w | Y = y]$.
3. What does happen if all messages are equiprobable ?

Relation with hypothesis testing

What is the connection with hypothesis testing ?

1. Fix P_X (i.e. choose the encoding function $f : \mathcal{W} \to \mathcal{X}$.
2. What is the optimal decoder w.r.t average error ?
   MAP (maximum a posteriori) :
   $g^*(y) = \arg\max_{w \in \mathcal{W}} \mathbb{P}[W = w | Y = y]$.
   Associated error : $P_e(y) = \mathbb{P}[W \neq w | Y = y]$.
3. What does happen if all messages are equiprobable ?
   $\mathbb{P}[W = w | Y = y] \propto \mathbb{P}[Y = y | W = w] \cdot \mathbb{P}[W = w]$ (from Bayes).
   ML (maximum likelihood) :
   $g^*(y) = \arg\max_{w \in \mathcal{W}} \mathbb{P}[Y = y | W = w]$.

Relation with hypothesis testing

What is the connection with hypothesis testing ?

1. Fix P_X (i.e. choose the encoding function $f : \boldsymbol{\mathcal{W}} \to \boldsymbol{\mathcal{X}}$.
2. What is the optimal decoder w.r.t average error ?
   MAP (maximum a posteriori) :
   $g^*(y) = \arg\max_{w \in \boldsymbol{\mathcal{W}}} \mathbb{P}[W = w | Y = y]$.
   Associated error : $P_e(y) = \mathbb{P}[W \neq w | Y = y]$.
3. What does happen if all messages are equiprobable ?
   $\mathbb{P}[W = w | Y = y] \propto \mathbb{P}[Y = y | W = w] \cdot \mathbb{P}[W = w]$ (from Bayes).
   ML (maximum likelihood) :
   $g^*(y) = \arg\max_{w \in \boldsymbol{\mathcal{W}}} \mathbb{P}[Y = y | W = w]$.

This also shows that the decoder should be deterministic (randomness reduces performance w.r.t. the average error probability).
What about the encoder ?

Is deterministic encoding optimal ?

For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, the encoder that minimizes P_e, is deterministic.

Is deterministic encoding optimal ?

For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, the encoder that minimizes P_e, is deterministic.

Proof : select a random encoder f built as a set of deterministic encoders, with a randomness parameter U :

$$f(w) = \{\tilde{f}(w, u)\}.$$

Then one have

$$P_e = \mathbb{P}\left[\hat{W} \neq W\right] = \mathbb{E}_U\left[\mathbb{P}\left[\hat{W} \neq W\right] | U\right].$$

meaning that $P_e = \mathbb{E}_U[P_e(U)]$, then there exists a realization u such that $P_e(u) \leq P_e$.

Synthesis

1. The ML estimate is the optimal decoder for a given encoder.
2. Deterministic coding is optimal.
3. What do we need ?
   - An efficient way to approximate the ML solution or its performance.
   - A method to select the optimal encoder.

Tips : if an achievability can be proved with random coding, then a deterministic pair encoder/decoder exists that outperforms the random encoder.

Weak converse bound

Theorem 5 (Weak converse bound).

Any $M-$code for $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$ satisfies :

$$\log M \leq \frac{\sup_{P_X} I(X;Y) + h(P_e)}{1 - P_e}.$$

The proof relies on :

- Markov chain : $W \to X \to Y \to \hat{W}$.
- Fano's theorem
- remark : random coding cannot be used for a converse.

Proof

Notes on information density

From the definition of information density, we can write :

$$\imath_{X;Y}(x;y) = \log\left(\frac{dP_{Y|X}}{dP_Y}(x,y)\right) = \log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right).$$

note : we define $\imath_{X;Y}(x;y) = +\infty$, if $P_{Y|X=x}$ is not abs. continuous w.r.t. P_Y.

Notes on information density

From the definition of information density, we can write :

$$\imath_{X;Y}(x;y) = \log\left(\frac{dP_{Y|X}}{dP_Y}(x,y)\right) = \log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right).$$

note : we define $\imath_{X;Y}(x;y) = +\infty$, if $P_{Y|X=x}$ is not abs. continuous w.r.t. P_Y.
Remark : compare this r.v. with a log likelihood ratio test (hyp testing) :

$$\mathrm{LLR}(y) = \log\left(\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_2)}\right),$$

thus the information density is $\approx$ LLR for one hypothesis against all other assumptions.

ML and information density

1. Consider a deterministic codebook : $f(w) = x_w$, where the x_w are the codewords.
2. ML estimate : $\hat{x} = g^*(y) = \arg\max_{x \in \mathcal{C}} \left[P_{Y|X}(y|x) \right]$.
3. Remind : $\imath_{X;Y}(x;y) = \log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right)$.
4. Then : the ML estimate takes x that maximizes the information density.

Properties of information density

These properties are useful for the theorem to be derived :

$$\mathbb{E}_{XY}\left[\imath_{X;Y}(X;Y)\right] = I(X;Y) \tag{1}$$

$$\mathbb{E}_{Y}\left[f(Y)\right] = \mathbb{E}_{Y}\left[e^{-\imath_{X;Y}(x;Y)} f(Y)\right] ; \forall x \tag{2}$$

$$\mathbb{E}_{\bar{X}Y}\left[f(\bar{X},Y)\right] = \mathbb{E}_{XY}\left[e^{-\imath_{X;Y}(X;Y)} f(X,Y)\right] ; \forall x \tag{3}$$

notes : need $f(y) = 0$ and $f(x,y) = 0$ for $\imath_{X;Y}(x;y) = -\infty$.

proofs : develop the expectation for a discrete variable.

Properties of information density (cont')

Then one can also write

1. For any x :
$$\mathbb{P}\left[\imath_{X;Y}(x;Y) > t\right] \leq e^{-t} \tag{4}$$
2. For any $\bar{X} \sim P_X$, s.t.$P_{YX} = P_Y P_X$:
$$\mathbb{P}\left[\imath_{X;Y}(\bar{X};Y) > t\right] \leq e^{-t}. \tag{5}$$

proofs : apply the former results with $f(Y) = \mathbb{1}_{\{\imath_{X;Y}(x;Y) \geq t\}}$.

3 - P2P discrete channels

Key elements

Shannon's method relies on the following key elements to approach the ML estimate :

1. Use random coding to derive asymptotic bounds.
2. Lowerbound the ML decision by a simpler decoder, but asymptotically efficient.

Shannon's achievability bound

Theorem 6 (Shannon's achievability).

For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, for all P_X, for all $\tau > 0$:
$\exists (M, \varepsilon)$-*code, s.t.* $\varepsilon \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) \leq \log(M) + \tau\right] + e^{-\tau}$

Further, the minimal error is :

$$\varepsilon^* \leq \inf_{\tau} \inf_{P_X} \left(\mathbb{P}\left[\imath_{X;Y}(X;Y) \leq \log(M) + \tau\right] + e^{-\tau}\right)$$

Proof (threshold decoder)

- Fix an encoder function, then P_X is fixed.
- The optimal decoder is the ML $g^*(y) = \arg\max_{w \in \mathcal{W}} \imath_{XY}(c_w; y)$.
- The proposed sub-optimal decoder is defined by :
  - Fix a threshold : $\log(M) + \tau$.
  - the decoder is

$$g(y) = \begin{cases} w & \exists! c_w \quad \text{s.t.} \imath_{X;Y}(c_w; y) \geq \log(M) + \tau \\ e & \text{otherwise} \end{cases}$$

Tips

$$\imath_{XY}(c_w; y) \geq \log(M) + \tau \iff P_{X|Y}(c_w|y) \geq e^{\tau}$$

proof : develop $\imath_{X;Y}(c_w; y)$.

Proof (threshold decoder)

The proof relies on standard tools in IT

- Symmetry
- Outage versus confusion probabilities,
- Union bound,
- Random codebook,
- Use Eq.(5).

Dependence testing bound

Theorem 7 (DT bound).
For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$*, for all* P_X *:*
$\exists(M,\varepsilon)$*-code, s.t.* $\varepsilon \leq \mathbb{E}_{XY}\left[\exp\left\{-\left(\imath_{X;Y}(X;Y) - \log\frac{M-1}{2}\right)^{+}\right\}\right]$.

Further, the minimal error is :

$$\varepsilon^* \leq \inf_{P_X}\left(\mathbb{E}_{XY}\left[\exp\left\{-\left(\imath_{X;Y}(X;Y) - \log\frac{M-1}{2}\right)^{+}\right\}\right]\right)$$

Proof (improved threshold decoder)

- Fix an encoder function, then P_X is fixed.
- The second proposed sub-optimal decoder is defined by :
  - Fix a threshold : γ (show below that $\gamma = \log \frac{M-1}{2}$ is optimal).
  - the decoder is

$$g(y) = \begin{cases} w & \min\{w\} \quad \text{s.t.} \imath_{X;Y}(c_w; y) \geq \gamma \\ e & \text{otherwise} \end{cases}$$

The unique difference with Shannon's scheme is that when several codes are above the threshold, one is chosen as the message.

Proof (improved threshold decoder)

The proof again relies on standard tools in IT

- Outage versus confusion probabilities,
- Union bound,
- Random codebook,
- Use Eq.(3).
- Hyp. Testing analogy.

Feinstein's Lemma

Theorem 8 (Feinstein's lemma).

For a given channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, *for all* P_X, $\gamma > 0$, $\varepsilon \in (0,1)$ *:* $\exists (M, \varepsilon)_{max}$-*code, s.t.* $M \geq \gamma(\varepsilon - \mathbb{P}\left[\imath_{X;Y}(X;Y) < \log\gamma\right])$.

Tips :

- If one take $\log(\gamma) = \log(M) + \tau$, the Shannon's bound is found but for error max.
- Feinstein's bound uses a greedy construction.

Feinstein's Lemma proof

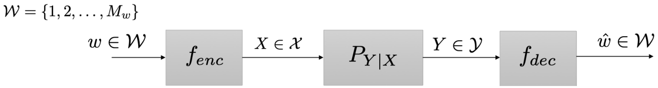
Homework

Propose a comparison of the three achievabilities : Shannon's bound, DT bound and Feinstein's lemma

- You can compare their relative position (strength of the bound), formally.
- You can choose a channel (BSC, Gaussian channel, ...) and illustrate these bounds. The channel may be built as a DMC channel (multiple channel uses).

3 - P2P discrete channels

n-length Channels



Definition 9 (P2P channel).

A point to point (P2P) multiple use communication channel is defined by the tuple $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, where :

- The input domain can be discrete or absolutely continuous : measurable space $(\boldsymbol{\mathcal{X}} = \mathcal{X}^n, \mathcal{F})$.
- The output domain can be discrete or absolutely continuous : measurable space $(\boldsymbol{\mathcal{Y}} = \mathcal{Y}^n, \mathcal{G})$.
- The conditional probability measure, $P_{\mathbf{Y}|\mathbf{X}} = \prod_{k=1}^n P_{Y_k|X_k}$.

Multi-letter information capacity

Definition 10 (Multi-letter information capacity).

The information capacity of a multi channel $(n, \mathcal{X}, \mathcal{Y}, P_{\boldsymbol{Y}|\boldsymbol{X}})$ is :

$$C_i = \lim_{n\to\infty} \inf \frac{1}{n} \sup_{P_{\boldsymbol{X}}} I(\boldsymbol{X}; \boldsymbol{Y}).$$

For a sationnary memoryless channel, the single-letterization is possible $(I(\boldsymbol{X}; \boldsymbol{Y}) = \sum_{k=1}^{n} I(X_k; Y_k) = nI(X; Y))$:

$$C_i = \sup_{P_X} I(X; Y).$$

Model with n channel uses

Definition 11 ((n, M, ε)-code).

For any channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, an (n, M, ε)-code is a (M, ε)-code for the n^{th} transformation $P_{\mathbf{Y}|\mathbf{X}} = \prod_{k=1}^{n} P_{Y|X}$.

A similar definition holds for the max error.

Model with n channel uses

Definition 11 ((n, M, ε)-code).

For any channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, an (n, M, ε)-code is a (M, ε)-code for the n^{th} transformation $P_{\mathbf{Y}|\mathbf{X}} = \prod_{k=1}^{n} P_{Y|X}$.

A similar definition holds for the max error.

The maximal input alphabet (FBL regime) is given by

$$M^*(n, \varepsilon) = \max \{M; \exists (n, M, \varepsilon)\text{-code}\} \quad (6)$$

$$M^*_{max}(n, \varepsilon) = \max \{M; \exists (n, M, \varepsilon)_{max}\text{-code}\} \quad (7)$$

The maximal rate of this code is

$$R^* = \frac{\log(M^*)}{n}$$

Operational metrics

Definition 12 (operational capacity).

The asymptotic capacity is defined by

1. $\varepsilon-$capacity : $C_\varepsilon = \lim_{n\to\infty} \inf \frac{1}{n} \log M^*(n,\varepsilon)$.
2. Shannon noisy channel capacity : $C = \lim_{\varepsilon\to O^+} C_\varepsilon$.

- This metric is operational in the sens it relies on physical properties.

Noisy channel theorem

Theorem 13 (Shannon's noisy channel capacity).

For a stationary memoryless channel, the asymptotic capacity is given

$$C_\varepsilon = C_i = \sup_{P_X} I(X;Y)$$

This theorem is perhaps the most significant result in information theory.
The following proves this result.

Proof : converse

Theorem 14 (Upper bound on C_ε).

For any channel, $\forall \varepsilon \in [0; 1)$

$$C_\varepsilon \leq \frac{C_i}{1-\varepsilon}$$

and

$$C \leq C_i$$

Proof :

1. Start from the generalized weak converse Th.5
2. Integrate C_i definition acc. to Def.10.
3. Takes the limit when $n \to \infty$.

Proof : achievability

Theorem 15 (Lower bound on C_ε).

For any channel, $\forall \varepsilon \in [0;1)$

$$C_\varepsilon \geq C_i$$

The limit does not depend on ε ! !

Proof : the proof starts from the Shannon's bound, with $\tau = \delta n$, $\log(M) = n(I(X;Y) - 2\delta)$, and $\delta > 0$, arbitrarily small :

$$\varepsilon \leq \mathbb{P}\left[\sum_{k=1}^{n} \imath_{X;Y}(X_k;Y_k) \leq nI(X;Y) - \delta n\right] + \exp(-\delta n) \qquad (8)$$

Then there exists a sequence of $(n, M, \varepsilon)\text{-code}$ with $\varepsilon \to 0$ and :

$$\log(M(n)) = n(I(X;Y) - 2\delta)$$

that proves : $C_\varepsilon \geq I(X;Y) - 2\delta$.

Second order approximation

Let be defined

Definition 16 (Dispersion).

$$V = \min_{P_X; I(X;Y)=C} \mathrm{var}[\imath_{X;Y}(X;Y)].$$

Then one have

$$\log(M^*(n,\varepsilon)) \geq nC - \sqrt{nV}Q^{-1}(\varepsilon) + o(\sqrt{n})$$

Second order approximation : proof

Proof starts from the Feinstein's bound, and uses WLLN.

Table of contents

ide//

4 - P2P continuous channels with constraint

Tips on continuous variables

Capacity with input constraints

Differential entropy

Remind that for an absolutely continuous r.v. X on $\boldsymbol{\mathcal{X}}$:

Definition 17 (Differential Entropy).

$$h(X) = -\int_{\boldsymbol{\mathcal{X}}} f_X(x) \cdot \log(f_X(x)) \cdot dx.$$

Anbiguity of an *information* variable :

$$\imath_X(x) \stackrel{?}{=} \log\left(\frac{1}{f_X(x)}\right)$$

When one value is drawn from a continuous distribution, what is the information you get ? infinite.
Problem : the differential entropy changes whith scaling.

Mutual information of cont. variables

Remind :

Definition 18 (Mutual information).

$$I(X;Y) = D(P_{XY}||P_X P_Y)$$

valid as well for discrete and absolutely continuous variables.

Mutual information of cont. variables

Remind :

Definition 18 (Mutual information).

$$I(X;Y) = D(P_{XY}||P_X P_Y)$$

valid as well for discrete and absolutely continuous variables. Then, the information density exists for continuous variables :

$$\imath_{X;Y}(x,y) = log \frac{f_{X;Y}(x;y)}{f_Y(y) \cdot f_X(x)}$$

This is why the former results relative to the channel capacity are also valid for continuous channels.

Mutual information of cont. variables

Remind :

Definition 18 (Mutual information).

$$I(X;Y) = D(P_{XY}||P_X P_Y)$$

valid as well for discrete and absolutely continuous variables. Then, the information density exists for continuous variables :

$$\imath_{X;Y}(x,y) = log \frac{f_{X;Y}(x;y)}{f_Y(y) \cdot f_X(x)}$$

This is why the former results relative to the channel capacity are also valid for continuous channels.
It is remarquable that the mutual information for c.v. relies on differential entropy with

$$I(X;Y) = h(X) + h(Y) - h(X,Y)$$

Gaussian random variable

Given a Gaussian random variable $X \sim \mathcal{N}\left((,0), \sigma^2\right)$. Its entropy is given by :

$$h(X) = \frac{1}{2}\log(2\pi e\sigma^2)$$

If Y is an absolutely continuous random variable of variance σ^2, then its entropy is upper bounded :

$$h(Y) \leq \frac{1}{2}\log(2\pi e\sigma^2)$$

Tips : The Gaussian random variable achieves the maximal entropy under variance constraint.

4 - P2P continuous channels with constraint

Tips on continuous variables

Capacity with input constraints

Infinite capacity of a continuous channel

Consider the AWGN channel

$$Y = X + Z$$

What is the capacity of this channel ?

Infinite capacity of a continuous channel

Consider the AWGN channel

$$Y = X + Z$$

What is the capacity of this channel ?
Under a given second order constraint $\mathbb{E}[X^2] = P$, the capacity is achieved with $X \sim \mathcal{N}(0, P)$.
Then taking $P \to \infty$ means an infinite capacity.
Need to constrain the source to fit with some reasonable properties.
Objective : how theory can deal with input constraints ?

Cost constraints

Given a channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X})$, an input constraint is defined by a subset $\mathcal{F}^n \subset \mathcal{X}^n$ of feasible codes. Typically, the input mapping cannot be done outside $\mathcal{F}^n$.

Let be define a cost function $c(\boldsymbol{X}) : \boldsymbol{\mathcal{X}} \to \mathbb{R}^+$ and a constraint such that a code is feasible if $c(\boldsymbol{X}) \leq P_{th}$.

Definition 19 (Channel with separable cost constraint).

A channel with separable cost constraint noted $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{th})$ is specified by

- $\boldsymbol{\mathcal{X}} = \mathcal{X}^n, \boldsymbol{\mathcal{Y}} = \mathcal{Y}^n$, input/output spaces.
- $P_{\boldsymbol{YX}} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{Y}}$.
- cost c : $\mathcal{X} \to \mathbb{R}^+$
- a threshold P_{th}

and verifies : $c(\boldsymbol{x}) = \frac{1}{n} \sum_{k=1}^{n} c(x_k) \leq P_{th}$, où $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Example of usual constraints

Average power constraint : $\frac{1}{n}\sum_{i=1}^{n}|x_i|^2 \leq P_{th}$.

Max power constraint : $\max_{1\leq k\leq n}|x_k| \leq A$.

Cost constrained code

Definition 20 ((n, M, ε, P)-code).

For any channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, an (n, M, ε, P)-code is an (n, M, ε)-code which satisfies the input cost constraint

$$\mathcal{F}^n \triangleq \{\boldsymbol{x}; \frac{1}{n} \sum_k c(x_k) \leq P\}.$$

The input codes have to be selected in $\mathcal{F}^n$.

A similar definition holds for the max error.

Cost constrained code

Definition 20 ((n, M, ε, P)-code).

For any channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, an $(n, M, \varepsilon, P)\text{-code}$ is an $(n, M, \varepsilon)\text{-code}$ which satisfies the input cost constraint

$$\mathcal{F}^n \triangleq \{\boldsymbol{x}; \frac{1}{n}\sum_k c(x_k) \leq P\}.$$

The input codes have to be selected in $\mathcal{F}^n$.

A similar definition holds for the max error.
The maximal input alphabet (FBL regime) is given by

$$M^*(n, \varepsilon, P) = \max\{M; \exists (n, M, \varepsilon, P)\text{-code}\} \quad (9)$$

$$M^*_{max}(n, \varepsilon, P) = \max\{M; \exists (n, M, \varepsilon, P)_{max}\text{-code}\} \quad (10)$$

Capacity under constraints

Definition 21 (Multi-letter information capacity under P constraint).

The information capacity of a multi channel $(n, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{\boldsymbol{Y}|\boldsymbol{X}}, P)$ is :

$$C_i = \lim_{n\to\infty} \inf \frac{1}{n} \sup_{P_{\boldsymbol{X}};\mathbb{E}[c(\boldsymbol{X})]\leq nP} I(\boldsymbol{X};\boldsymbol{Y}).$$

Note : the constraint should be an admissible constraint. I.e. there exists at least one code.

Definition 22 (Operational capacity).

The asymptotic capacity under constraint is defined by

1. $\varepsilon-$capacity : $C_\varepsilon(P) = \lim_{n\to\infty} \inf \frac{1}{n} \log M^*(n, \varepsilon, P)$.
2. Shannon noisy channel capacity : $C(P) = \lim_{\varepsilon\to O^+} C_\varepsilon(P)$.

Questions

The key question is now to show how the operational capacity and the information capacity remain connected to each other.

As a side question, is the single-letterization for stationary memoryless channels still valid ?

Single Letterization

Theorem 23 (Information capacity of sationary memoryless channel with cost).

The information capacity of a stationary memoryless channel with separable cost is

$$C_i(P) = \sup_{P_X; \mathbb{E}[c(X)] \leq P} I(X;Y)$$

The proof uses converse/achievability
Let be denoted :

$$c_0(P) = \sup_{P_X; \mathbb{E}[c(X)] \leq P} I(X;Y)$$

Single Letterization : converse

Proove : $C_i(P) \geq c_0(P)$

This comes by taking $P_{\boldsymbol{X}} = P_X^n$ (iid).

Then : $I(\boldsymbol{X}; \boldsymbol{Y}) = nI(X; Y)$.

And the constraints are linked with

$$(\mathbb{E}[c(X)] \leq P) \Rightarrow (\mathbb{E}[c(\boldsymbol{X})] \leq nP)$$

If a distribution P_X achieves $c_0(P)$, P_X^n is valid (constraint) and achieves $c_0(P)$. $c_0(P)$ is achievable.

Single Letterization : achievability

Proove : $C_i(P) \leq c_0(P)$

$$I(\boldsymbol{X};\boldsymbol{Y}) \leq \sum_{k=1}^{n} I(X_k;Y_k)$$

$$\leq \sum_{k=1}^{n} c_0\left(\mathbb{E}[c(X_k)]\right)$$

$$\leq nc_0\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[c(X_k)]\right)$$

$$= nc_0(P)$$

The last inequality comes from the concavity of $c_0(P)$.

Capacity under constraint

Theorem 24 (Noisy channel with cost constraint).
The operational capacity and the information capacity for a staitonary memoryless channel with cost constraints are equal :

$$C(P) = C_i(P)$$

Proof :

1. General weak converse : $C_\varepsilon(P) \leq \frac{C_i(P)}{1-\varepsilon}$.
2. For any stationary memoryless channel with input constraints, $C(P) \leq C_i(P)$.

Weak converse

The weak converse follows the same reasoning (Fano, ...)
Taking one (n, M, ε, P)-code one can follow the Markov chain $W \to \boldsymbol{X} \to \boldsymbol{Y} \to \hat{W}$:

$$\begin{aligned} -h(\varepsilon) + (1-\varepsilon)\log(M) &\le I(W;\hat{W}) \\ &\le I(\boldsymbol{X};\boldsymbol{Y}) \\ &\le \sup_{P_X;\mathbb{E}[c(\boldsymbol{X})]\le P} I(\boldsymbol{X};\boldsymbol{Y}) \\ &\le nc_0(P) = nC_i(P). \end{aligned}$$

Achievability

First proove the following

Theorem 25 (Extended Feinstein's lemma).

For a channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X}, P)$, where $\mathcal{F} \subset \boldsymbol{\mathcal{X}}$ is the set of feasible codes, the error probability is bounded by :

$$\varepsilon \cdot P_X(\mathcal{F}) \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) < \log(\gamma)\right] + \frac{M}{\gamma}$$

Achievability

First proove the following

Theorem 25 (Extended Feinstein's lemma).

For a channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X}, P)$, where $\mathcal{F} \subset \boldsymbol{\mathcal{X}}$ is the set of feasible codes, the error probability is bounded by :

$$\varepsilon \cdot P_X(\mathcal{F}) \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) < \log(\gamma)\right] + \frac{M}{\gamma}$$

Then apply this theorem with $\log(M) = n(I(X;Y) - 2\delta)$ and $\log(\gamma) = n(I(X;Y) - \delta)$, leading to :

$$\varepsilon P_X(\mathcal{F}) \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) < n(I(X;Y) - \delta)\right] + e^{-n\delta}$$

Achievability

First proove the following

Theorem 25 (Extended Feinstein's lemma).

For a channel $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, P_{Y|X}, P)$, where $\mathcal{F} \subset \boldsymbol{\mathcal{X}}$ is the set of feasible codes, the error probability is bounded by :

$$\varepsilon \cdot P_X(\mathcal{F}) \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) < \log(\gamma)\right] + \frac{M}{\gamma}$$

Then apply this theorem with $\log(M) = n(I(X;Y) - 2\delta)$ and $\log(\gamma) = n(I(X;Y) - \delta)$, leading to :

$$\varepsilon P_X(\mathcal{F}) \leq \mathbb{P}\left[\imath_{X;Y}(X;Y) < n(I(X;Y) - \delta)\right] + e^{-n\delta}$$

The result is obtained with the WLLN.

Example : Capacity of an AWGN channel

Table of contents

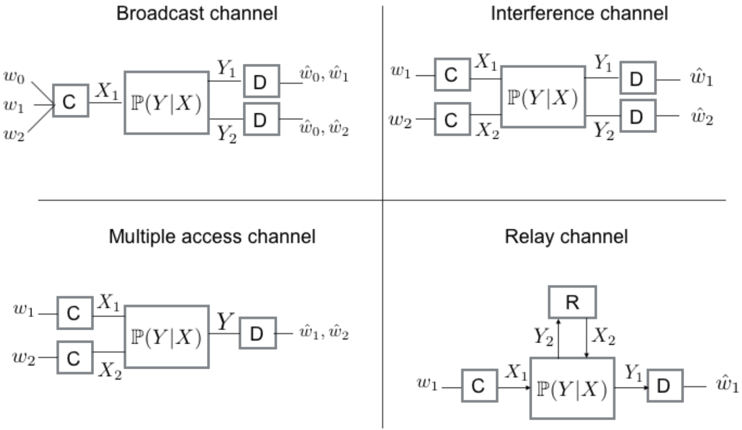
ide//

5 - Multi-user channels

General properties

Details of the MAC channel

Multi-user scenarios



- Focus on Multiple Access Channel

Other relevant scenarios

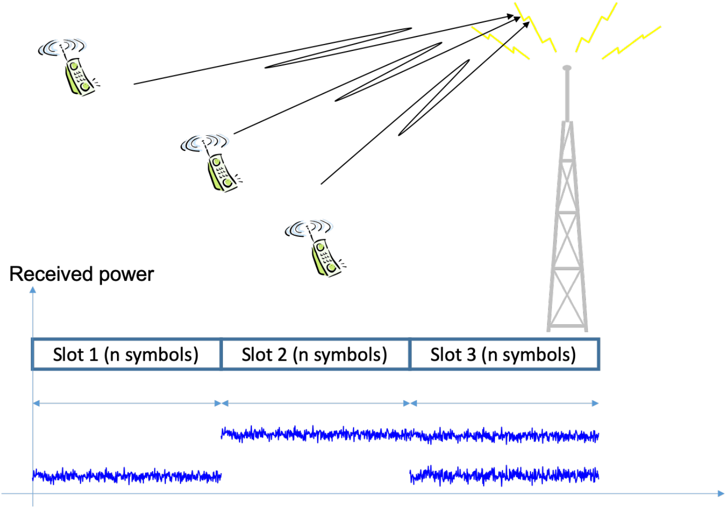
Some other models have been studied in the litterature : channel with states, graphical networks, large scale Gaussian networks, Wiretap channel.

5 - Multi-user channels

General properties

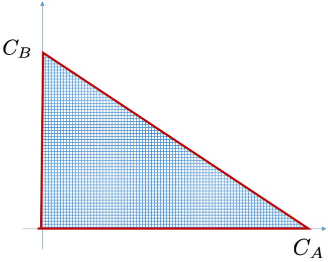
Details of the MAC channel

MAC : problem positionning



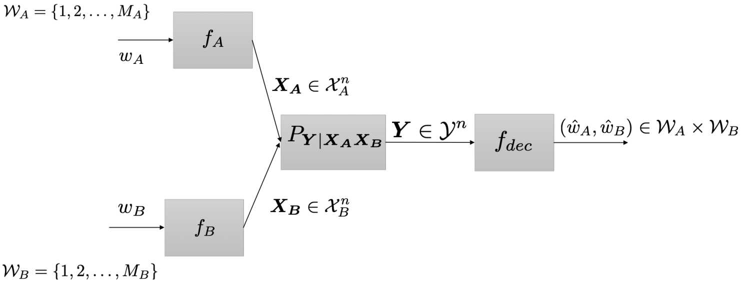
Natural solution : time sharing

Time-sharing means alternative transmission : i.e. two P2P independant transmissions.



Could we achieve a higher rate region ?

The IT MAC model



The IT 2-user MAC model (cont)

Definition 26.

$(n, M_A, M_B, \varepsilon)-$code for the multiple access channel For a MAC channel $(n, \boldsymbol{\mathcal{X}}_A, \boldsymbol{\mathcal{X}}_B, \boldsymbol{\mathcal{Y}}, P_{\mathbf{Y}|\mathbf{X}_A\mathbf{X}_B})$, a $(n, M_A, M_B, \varepsilon)-$code is defined by

- $f_A : \boldsymbol{\mathcal{W}}_A \to \boldsymbol{\mathcal{X}}_A$
- $f_B : \boldsymbol{\mathcal{W}}_B \to \boldsymbol{\mathcal{X}}_B$
- $g : \boldsymbol{\mathcal{Y}} \to \boldsymbol{\mathcal{W}}_A \times \boldsymbol{\mathcal{W}}_B$

such that $\mathbb{P}\left[\left\{W_A \neq \hat{W}_A\right\} \cup \left\{W_B \neq \hat{W}_B\right\}\right] \leq \varepsilon$

Operational metrics

1. The fundamental limit is the closure of a capacity region :

$$\mathcal{R}^*(n,\varepsilon) = \left\{(R_A, R_B); \exists(n, 2^{nR_A}, 2^{nR_B}, \epsilon) - \mathrm{code}\right\}.$$

2. Asymptotic regime

$$\mathcal{C}_\epsilon = \mathrm{cl}\left(\lim_{n\to\infty} \inf \mathcal{R}^*(n,\varepsilon)\right).$$

   where cl denotes the closure set.

3. Capacity region :

$$\mathcal{C} = \lim \mathcal{C}_\varepsilon = \bigcap_{\varepsilon>0} \mathcal{C}_\varepsilon.$$

Note :

$$\lim_n \inf \mathcal{A}_n = \{a; a \in \mathcal{A}_n, \forall n > n_0\}$$

Capacity region

Theorem 27 (2-user MAC Capacity region).

The capacity region of the 2-user MAC is :

$$C_\varepsilon = \overline{co} \bigcup_{P_A P_B} Penta(P_A, P_B)$$

$$= \left[\bigcup_{P_U P_{A|U} P_{B|U}} Penta(P_{A|U}, P_{B|U} | P_U) \right]$$

Only the second definition is valid when the problem is with cost constraint.
with

- $\overline{co}$ *: convex hull followed by taking the closure.*

Penta function

- $\mathrm{Penta}(P_A, P_B) = \left\{ (R_A, R_B);\ \begin{array}{c} 0 \leq R_A \leq I(A;Y|B) \\ 0 \leq R_b \leq I(B;Y|A) \\ R_A + R_B \leq I(A,B;Y) \end{array} \right\}.$
- $\mathrm{Penta}(P_{A|U}, P_{B|U}|P_U) =$
  $\left\{ (R_A, R_B);\ \begin{array}{c} 0 \leq R_A \leq I(A;Y|B,U) \\ 0 \leq R_B \leq I(B;Y|A,U) \\ R_A + R_B \leq I(A,B;Y|U) \end{array} \right\}.$

Achievability bound

Theorem 28 (2-user MAC achievability bound).
Give a MAC channel $P_{A,B,Y} = P_A \cdot P_B \cdot P_{Y|AB}$, then $\forall \gamma_A, \gamma_B, \gamma_{AB}$, $\forall M_A, M_B$, there exists an (M_1, M_2, ε)-code, such that :

$$\begin{aligned}\varepsilon \leq & \mathbb{P}\left[\{\imath_{AB;Y}(A,B;Y) \leq \log(\gamma_{AB})\}\right.\\ & \cup \left\{\imath_{A;Y|B}(A;Y) \leq \log(\gamma_A)\right\} \cup \left.\left\{\imath_{B;Y|A}(B;Y) \leq \log(\gamma_B)\right\}\right]\\ & + (M_A - 1)(M_B - 1)e^{-\gamma_{AB}} + (M_A - 1)e^{-\gamma_A} + (M_B - 1)e^{-\gamma_B}\end{aligned}$$

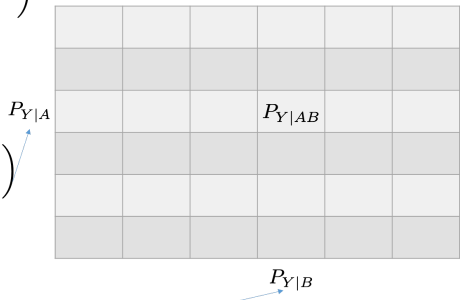
Tips : use random coding, and threshold decoding with three threshold tests on information densities.

Equivalent with LLR tests

$$\imath_{AB;Y}(a,b;y) = \log\left(\frac{P_{Y|AB}}{P_Y}\right)$$

$$\imath_{B;Y|A}(b;y|a) = \log\left(\frac{P_{Y|AB}}{P_{Y|A}}\right)$$

$$\imath_{A;Y|B}(a;y|b) = \log\left(\frac{P_{Y|AB}}{P_{Y|B}}\right)$$



Gaussian MAC

Model and results : toward superposition coding.

Table of contents

ide//