α-Capacity of Communication Channels with Feedback: Theoretical Overview

STW 2022, Shenzhen, China, Sept. 28th, 2022

Olivier Rioul
Télécom Paris
Institut Polytechnique de Paris, France

<olivier.rioul@telecom-paris.fr>

Outline

Outline

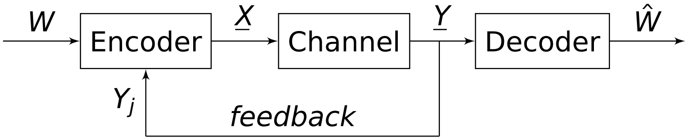
Introduction

Ingredients

Inequalities

Main Result

Point-to-Point Communication Channel With Perfect Feedback



- (n, M) block code
- M-ary information source W
- memoryless channel $\underline{X} = (X_1, \ldots, X_n) \to \underline{Y} = (Y_1, \ldots, Y_n)$
- $X_j = f(W, Y_1, \ldots, Y_j)$ at each time instant j.
- probability of decoding error $\mathbb{P}_e = 1 - \mathbb{P}_s = \mathbb{P}(\hat{W} \neq W)$
- Shannon capacity $C = \max_{p_X} I(X;Y)$ (not increased by feedback)

Non-Asymptotic Converse Theorems

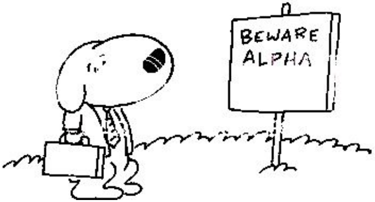
- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR

Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)

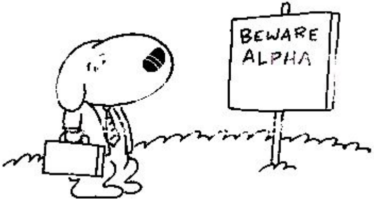
Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)
- using α-information theory:



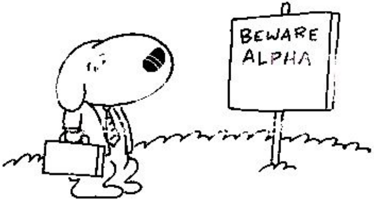
Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)
- using α-information theory:
  - α-divergence $D_\alpha(p\|q)$



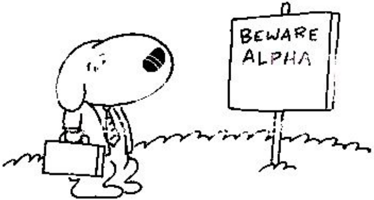
Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)
- using α-information theory:
    - α-divergence $D_\alpha(p\|q)$
    - α-information $I_\alpha(X;Y)$



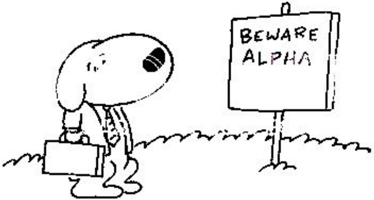
Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)
- using α-information theory:
  - α-divergence $D_\alpha(p\|q)$
  - α-information $I_\alpha(X;Y)$
  - α-capacity C_α



Non-Asymptotic Converse Theorems

- lower bounds on $\mathbb{P}_e$ vs. coding rate $R = \frac{\log_2 M}{n}$ or vs. SNR
- for any (n, M) code (without requiring $n \to +\infty$)
- using α-information theory:
  - α-divergence $D_\alpha(p\|q)$
  - α-information $I_\alpha(X; Y)$
  - α-capacity C_α
- illustration: binary-input symmetric channels: AWGN with or without output quantization



Outline

Basic Notations

- all probability distributions are dominated by some σ-finite measure μ
- any random variable X admits a probability density p_X w.r.t. μ
- α-quantities defined below are independent of the choice of μ
- discrete or continuous:
  - $\mu =$ Lebesgue measure; $p_X =$ p.d.f.; $\int_x p_X(x) = 1$
  - $\mu =$ counting measure: $p_X =$ p.m.f.; $\sum p_X(x)\,\mathrm{d}x = 1$
  - unifying notation $\sum\!\!\!\!\!\!\int_x p_X(x) = 1.$

Basic Notations

- all probability distributions are dominated by some σ-finite measure μ
- any random variable X admits a probability density p_X w.r.t. μ
- α-quantities defined below are independent of the choice of μ
- discrete or continuous:
  - $\mu =$ Lebesgue measure; $p_X =$ p.d.f.; $\int_x p_X(x) = 1$
  - $\mu =$ counting measure: $p_X =$ p.m.f.; $\sum p_X(x)\,\mathrm{d}x = 1$
  - unifying notation $\displaystyle\sum\!\!\!\!\!\!\int_x p_X(x) = 1$.
- order $\alpha > 0$ (either $\alpha < 1$ or $\alpha > 1$); limiting case $\alpha = 1$ (Shannon)
- α-product: Hellinger integral or Bhattacharyya coefficient of two distributions p, q:

$$\boxed{(p\|q)_\alpha \triangleq \Big(\sum\!\!\!\!\!\!\int p^\alpha q^{1-\alpha}\Big)^{1/\alpha}}$$

(Rényi) α-Divergence [Rényi'61]

$$D_\alpha(p\|q) \triangleq \frac{1}{\alpha-1}\log \sum\!\!\!\!\!\!\int p^\alpha q^{1-\alpha} = \frac{\alpha}{\alpha-1}\log(p\|q)_\alpha$$

- $D_\alpha(p\|q) \geqslant 0$ with equality $D_\alpha(p\|q) = 0 \iff p \equiv q$
- binary case:

$$d_\alpha(p\|q) \triangleq \tfrac{1}{\alpha-1}\log\big((1-p)^\alpha(1-q)^{1-\alpha} + p^\alpha q^{1-\alpha}\big).$$

(Rényi) α-Divergence [Rényi'61]

$$D_\alpha(p\|q) \triangleq \frac{1}{\alpha-1}\log \sum\!\!\!\!\!\!\int p^\alpha q^{1-\alpha} = \frac{\alpha}{\alpha-1}\log(p\|q)_\alpha$$

- $D_\alpha(p\|q) \geqslant 0$ with equality $D_\alpha(p\|q)=0 \iff p\equiv q$
- binary case:

$$d_\alpha(p\|q) \triangleq \tfrac{1}{\alpha-1}\log\big((1-p)^\alpha(1-q)^{1-\alpha}+p^\alpha q^{1-\alpha}\big).$$

- $D_\alpha(p,q)$ is nondecreasing in α. Limits $\alpha\to 0, 1, \infty$:
  - $D_0(p\|q) = -\log \sum\!\!\!\!\!\!\int_{p>0} q$
  - $D_1(p\|q) = D(p\|q) = \sum\!\!\!\!\!\!\int p\log\dfrac{p}{q}$ (Kullback-Leibler)
  - $D_\infty(p\|q) = \log \sup_q \dfrac{p}{q}$
- $D_\alpha(p,q)$ is lower semi-continuous in (p,q)

(Sibson) α-information [Sibson'69]

$$I_\alpha(X;Y) \triangleq \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha.$$

- it's $D_\alpha(p_{X|y} \| p_X) = \frac{\alpha}{\alpha-1} \log(p_{X|y} \| p_X)_\alpha$ averaged over Y *inside the logarithm*
- $I_\alpha(X;Y) \geqslant 0$ with equality $I_\alpha(X;Y) = 0$ if and only if X and Y are independent

(Sibson) α-information [Sibson'69]

$$I_\alpha(X;Y) \triangleq \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha.$$

- it's $D_\alpha(p_{X|y} \| p_X) = \frac{\alpha}{\alpha-1} \log (p_{X|y} \| p_X)_\alpha$ averaged over Y *inside the logarithm*
- $I_\alpha(X;Y) \geqslant 0$ with equality $I_\alpha(X;Y) = 0$ if and only if X and Y are independent
- alternative expression $I_\alpha(X;Y) = \dfrac{\alpha}{\alpha-1} \log \displaystyle\sum\!\!\!\!\!\!\int_y \Big(\sum\!\!\!\!\!\!\int_x p_X p_{Y|X}^\alpha \Big)^{1/\alpha}$

  φ-concave in p_X for fixed channel $p_{Y|X}$ (for some increasing φ).

(Sibson) α-information [Sibson'69]

$$I_\alpha(X;Y) \triangleq \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha.$$

- it's $D_\alpha(p_{X|y} \| p_X) = \frac{\alpha}{\alpha-1} \log (p_{X|y} \| p_X)_\alpha$ averaged over Y *inside the logarithm*
- $I_\alpha(X;Y) \geqslant 0$ with equality $I_\alpha(X;Y) = 0$ if and only if X and Y are independent
- alternative expression $I_\alpha(X;Y) = \dfrac{\alpha}{\alpha-1} \log \sum\!\!\!\!\!\!\int_y \big(\sum\!\!\!\!\!\!\int_x p_X p_{Y|X}^\alpha \big)^{1/\alpha}$

  φ-concave in p_X for fixed channel $p_{Y|X}$ (for some increasing φ).
- $I_\alpha(X;Y)$ is non decreasing in α. Limits $\alpha \to 0, 1, \infty$:
  - $I_0(X;Y) = -\log \sup_y \sum\!\!\!\!\!\!\int_{p_{y|x}>0} p_X$
  - $I_1(X;Y) = I(X;Y)$ (Shannon's mutual information)
  - $I_\infty(X;Y) = \log \sum\!\!\!\!\!\!\int_y \sup_{p_X(x)>0} p_{Y|x}$

α-Response

For any p_X, define its α-response of the channel $X \to Y$ by

$$q_{Y,p_X} \triangleq \frac{(p_{X|Y}\|p_X)_\alpha \, p_Y}{\mathbb{E}_Y(p_{X|Y}\|p_X)_\alpha} = \frac{\left(\sum\!\!\!\!\!\!\int_x p_X p_{Y|X}^\alpha\right)^{1/\alpha}}{\sum\!\!\!\!\!\!\int_y \left(\sum\!\!\!\!\!\!\int_x p_X p_{Y|X}^\alpha\right)^{1/\alpha}}.$$

- by chain rule for α-product : $(p_{XY}\|q_{XY})_\alpha = \big((p_{X|Y}\|q_{X|Y})_\alpha \, p_Y\|q_Y\big)_\alpha$, $(p_{XY}\|p_X q_Y)_\alpha = \big((p_{X|Y}\|p_X)_\alpha \, p_Y\|q_Y\big)_\alpha = (q_{Y,p_X}\|q_Y)_\alpha \cdot \mathbb{E}_Y(p_{X|Y}\|p_X)_\alpha$ gives:

α-Response

For any p_X, define its α-response of the channel $X \to Y$ by

$$q_{Y,p_X} \triangleq \frac{(p_{X|Y} \| p_X)_\alpha \, p_Y}{\mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha} = \frac{\left(\sum\!\!\!\!\!\int_x p_X p^\alpha_{Y|X} \right)^{1/\alpha}}{\sum\!\!\!\!\!\int_y \left(\sum\!\!\!\!\!\int_x p_X p^\alpha_{Y|X} \right)^{1/\alpha}}.$$

- by chain rule for α-product : $(p_{XY} \| q_{XY})_\alpha = \big((p_{X|Y} \| q_{X|Y})_\alpha \, p_Y \| q_Y \big)_\alpha$, $(p_{XY} \| p_X q_Y)_\alpha = \big((p_{X|Y} \| p_X)_\alpha \, p_Y \| q_Y \big)_\alpha = (q_{Y,p_X} \| q_Y)_\alpha \cdot \mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha$ gives:
- Sibson's identity: For any q_Y,

$$D_\alpha(p_{XY} \| p_X q_Y) = D_\alpha(q_{Y,p_X} \| q_Y) + I_\alpha(X;Y).$$

α-Response

For any p_X, define its α-response of the channel $X \to Y$ by

$$q_{Y,p_X} \triangleq \frac{(p_{X|Y} \| p_X)_\alpha \, p_Y}{\mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha} = \frac{\left(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\right)^{1/\alpha}}{\sum\!\!\!\!\!\!\int_y \left(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\right)^{1/\alpha}}.$$

- by chain rule for α-product : $(p_{XY} \| q_{XY})_\alpha = \big((p_{X|Y} \| q_{X|Y})_\alpha \, p_Y \| q_Y\big)_\alpha$, $(p_{XY} \| p_X q_Y)_\alpha = \big((p_{X|Y} \| p_X)_\alpha \, p_Y \| q_Y\big)_\alpha = (q_{Y,p_X} \| q_Y)_\alpha \cdot \mathbb{E}_Y (p_{X|Y} \| p_X)_\alpha$ gives:
- Sibson's identity: For any q_Y,

$$D_\alpha(p_{XY} \| p_X q_Y) = D_\alpha(q_{Y,p_X} \| q_Y) + I_\alpha(X;Y).$$

- in particular

$$I_\alpha(X;Y) = \min_{q_Y} D_\alpha(p_{XY} \| p_X q_Y) = D_\alpha(p_{XY} \| p_X q_{Y,p_X})$$

  where the α-response q_{Y,p_X} is the unique distribution achieving the minimum.

α-Capacity

By analogy with Shannon's formula $C = \max_{p_X} I(X;Y)$,

$$\boxed{C_\alpha \triangleq \max_{p_X} I_\alpha(X;Y)}.$$

α-Capacity

By analogy with Shannon's formula $C = \max_{p_X} I(X;Y)$,

$$\boxed{C_\alpha \triangleq \max_{p_X} I_\alpha(X;Y)}.$$

Theorem (Characterization of α-Capacity [Csiszar'95,CaiVerdu'19])

For discrete X,

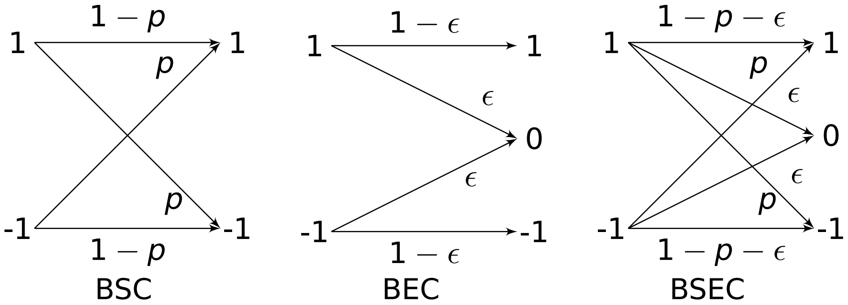
$$C_\alpha = \min_{q_Y} \max_x D_\alpha(p_{Y|x} \| q_Y) = \max_x D_\alpha(p_{Y|x} \| q_{Y,p_X^*})$$

where q_{Y,p_X^*} is the α-response of the distribution p_X^* achieving the maximum of $I_\alpha(X;Y)$.

Proof.

Simple proof in [RioulNguyen'22] (ICCE'22). □

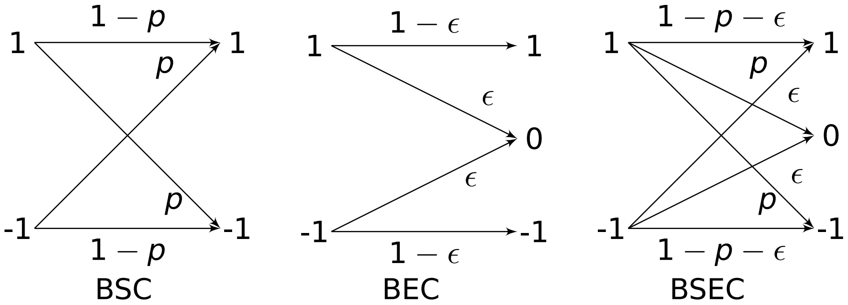
Illustration: Binary-Input Symmetric Channels



binary-input symmetric channels: arise from AWGN with or without output quantization (energy per bit $E_b = 1$ for input $X \in \{\pm 1\}$ and noise variance $\sigma^2 = N_0/2$)

- binary symmetric channel BSC(p), $p = Q(\sqrt{\frac{2E_b}{N_0}}) = Q(\frac{1}{\sigma})$

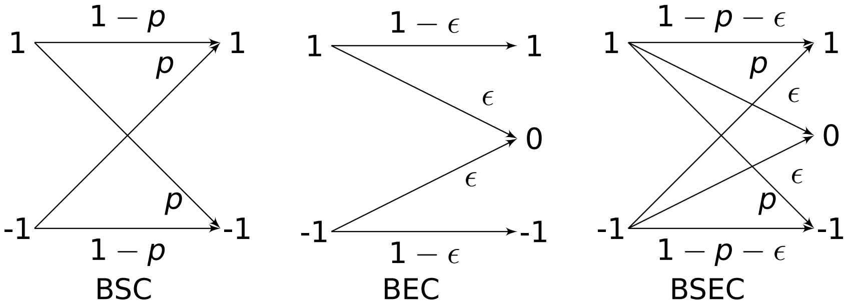
Illustration: Binary-Input Symmetric Channels



binary-input symmetric channels: arise from AWGN with or without output quantization (energy per bit $E_b = 1$ for input $X \in \{\pm 1\}$ and noise variance $\sigma^2 = N_0/2$)

- binary symmetric channel BSC(p), $p = Q(\sqrt{\frac{2E_b}{N_0}}) = Q(\frac{1}{\sigma})$
- binary erasure channel BEC(ϵ), $\epsilon = Q(\frac{1}{2\sigma})$

Illustration: Binary-Input Symmetric Channels



binary-input symmetric channels: arise from AWGN with or without output quantization (energy per bit $E_b = 1$ for input $X \in \{\pm 1\}$ and noise variance $\sigma^2 = N_0/2$)

- binary symmetric channel BSC(p), $p = Q(\sqrt{\frac{2E_b}{N_0}}) = Q(\frac{1}{\sigma})$
- binary erasure channel BEC(ϵ), $\epsilon = Q(\frac{1}{2\sigma})$
- binary symmetric erasure and error channel BSEC(p, ϵ) $p = Q(\frac{3}{2\sigma})$ and $p + \epsilon = Q(\frac{1}{2\sigma})$

Binary-Input Symmetric Channel (General Case)

Theorem (α-Capacity of a binary-input symmetric channel)

$$\boxed{C_\alpha = 1 - \frac{\alpha}{1-\alpha} \log \sum\!\!\!\!\!\!\int \tfrac{1}{2}\bigl(p_{Y|1}^\alpha + p_{-Y|1}^\alpha\bigr)^{1/\alpha}.}$$

Binary-Input Symmetric Channel (General Case)

Theorem (α-Capacity of a binary-input symmetric channel)

$$\boxed{C_\alpha = 1 - \frac{\alpha}{1-\alpha} \log \sum\!\!\!\!\!\!\int \tfrac{1}{2}\bigl(p_{Y|1}^\alpha + p_{-Y|1}^\alpha\bigr)^{1/\alpha}.}$$

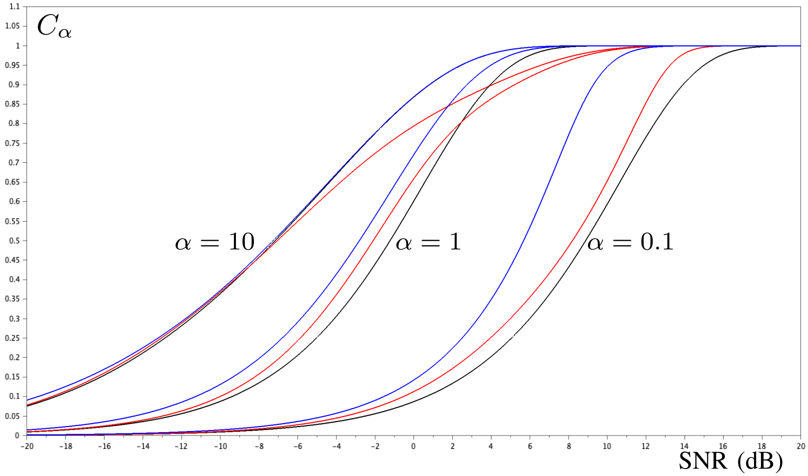
- $C_\alpha \leqslant 1$ bit
- C_α is nondecreasing in α
- $\alpha \mapsto C_\alpha$ is continuous in α. Limits $\alpha \to 0, 1/2, 1, \infty$:
  - $C_0 =$ *feedback zero-error capacity*
  - $C_{1/2} = R_0 = 1 - \log\bigl(1 + \sum\!\!\!\!\!\!\int \sqrt{p_{Y|1}\, p_{Y|-1}}\bigr)$ *cut-off rate* [Massey'74]
  - $C_1 = C =$ Shannon capacity
  - $C_\infty = 1 + \log \sum\!\!\!\!\!\!\int \frac{1}{2} \max(p_{Y|1}, p_{Y|-1}) = \log \frac{\mathbb{P}_s(X|Y)}{\mathbb{P}_s(X)}$ (MAP)

Some α-Capacities of Binary-Input Memoryless Channels

	C_α	cut-off $C_{1/2}$	usual capacity $C=C_1$	C_∞
BSC	$1-\frac{1}{1-\alpha}\log(p^\alpha+(1-p)^\alpha)$	$1-\log(1+2\sqrt{p(1-p)})$	$1-h(p)$	$1-\log\frac{1}{1-p}$
BEC	$1-\frac{\alpha}{1-\alpha}\log(1-\epsilon+2^{\frac{1-\alpha}{\alpha}}\epsilon)$	$1-\log(1+\epsilon)$	$1-\epsilon$	$1-\log\frac{1}{1-\epsilon/2}$
BSEC	$1-\frac{\alpha}{1-\alpha}\log\big((p^\alpha+(1-p-\epsilon)^\alpha)^{\frac{1}{\alpha}}+2^{\frac{1-\alpha}{\alpha}}\epsilon\big)$	$1-\log(1+\epsilon+2\sqrt{p(1-p-\epsilon)})$	$(1-\epsilon)(1-h(\frac{p}{1-\epsilon}))$	$1-\log\frac{1}{1-p-\epsilon/2}$
AWGN	$1-\frac{\alpha}{1-\alpha}\log\int_{-\infty}^{\infty}\frac{e^{-(y^2+1)/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \times\frac{1}{2}(e^{y\alpha/\sigma^2}+e^{-y\alpha/\sigma^2})^{1/\alpha}\,\mathrm{d}y$	$1-\log(1+e^{-1/2\sigma^2})$	$1-\int_{-\infty}^{\infty}\frac{e^{-(y-1)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \times\log(1+e^{-2y/\sigma^2})\,\mathrm{d}y$	$1-\log\frac{1}{1-Q(1/\sigma)}$

where $h(p)=-p\log p-(1-p)\log(1-p)$ is the binary entropy function.

α-capacities: BSC, BSEC, and AWGN



α-capacities of binary-input BSC (black), BSEC (red) and AWGN channel (blue) as a function of SNR$= 1/(2\sigma^2)$ per transmitted bit.

Outline

α-Data Processing Inequality (DPI)

Theorem (DPI for α-Divergence)

When a given channel $p_{Y|X}$ responds to two different inputs: $\begin{cases} p_X \to \boxed{p_{Y|X}} \to p_Y \\ q_X \to \boxed{p_{Y|X}} \to q_Y, \end{cases}$ *then*

$$\boxed{D_\alpha(p_Y \| q_Y) \leqslant D_\alpha(p_X \| q_X)}.$$

α-Data Processing Inequality (DPI)

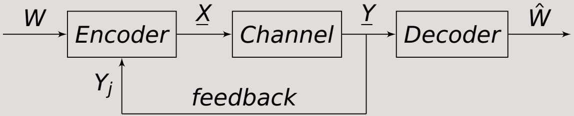
Theorem (DPI for α-Divergence)

When a given channel $p_{Y|X}$ responds to two different inputs: $\begin{cases} p_X \to \boxed{p_{Y|X}} \to p_Y \\ q_X \to \boxed{p_{Y|X}} \to q_Y, \end{cases}$ *then*

$$\boxed{D_\alpha(p_Y \| q_Y) \leqslant D_\alpha(p_X \| q_X)}.$$

Theorem (DPI for α-Information)

If $W - Y - \hat{W}$ forms a Markov chain:



then $\boxed{I_\alpha(W;Y) \geqslant I_\alpha(W;\hat{W})}$.

α-Fano Inequality

Theorem (Fano Inequality for α-Information [Rioul-GSI'21])

$$\boxed{I_\alpha(W;Y) \geqslant d_\alpha\big(\mathbb{P}_s(W|Y)\big\|\mathbb{P}_s(W)\big)}$$

where $d_\alpha(p\|q)$ is the binary α-divergence and

$$\begin{cases} \mathbb{P}_s(W|Y) \triangleq \max\limits_{W-Y-\hat{W}} \mathbb{P}(\hat{W}=W) = \mathbb{E}_Y\big(\max_w p_{W|Y}(w|Y)\big) \\ \mathbb{P}_s(W) \triangleq \max\limits_{w} p_W(w) \end{cases}$$

achieved by the MAP rule yielding minimum probability of error $\mathbb{P}_e$ upon observing channel output Y or not.

α-Fano Inequality

Theorem (Fano Inequality for α-Information [Rioul-GSI'21])

$$\boxed{I_\alpha(W;Y) \geqslant d_\alpha\big(\mathbb{P}_s(W|Y)\big\|\mathbb{P}_s(W)\big)}$$

where $d_\alpha(p\|q)$ is the binary α-divergence and

$$\begin{cases} \mathbb{P}_s(W|Y) \triangleq \max\limits_{W-Y-\hat{W}} \mathbb{P}(\hat{W}=W) = \mathbb{E}_Y\big(\max_w p_{W|Y}(w|Y)\big) \\ \mathbb{P}_s(W) \triangleq \max\limits_w p_W(w) \end{cases}$$

achieved by the MAP rule yielding minimum probability of error $\mathbb{P}_e$ upon observing channel output Y or not.

- for *equiprobable* M-ary source W: $I_\alpha(W;Y) \geqslant d_\alpha\big(\mathbb{P}_s\big\|\frac{1}{M}\big)$.
- $\alpha \to 1$ one recovers the classical Fano inequality [Fano'52]

Memoryless Channel With (or Without) Perfect Feedback

In this case

$$p_{\underline{Y}|W} = \prod_{j=1}^{n} p_{Y_j|W,Y_1,\ldots,Y_{j-1}} = \prod_{j=1}^{n} p_{Y_j|X_j}$$

where $X_j = f(W, Y_1, \ldots, Y_{j-1})$ for $j = 1, \ldots, n$.

Memoryless Channel With (or Without) Perfect Feedback

In this case

$$p_{\underline{Y}|W} = \prod_{j=1}^{n} p_{Y_j|W,Y_1,\dots,Y_{j-1}} = \prod_{j=1}^{n} p_{Y_j|X_j}$$

where $X_j = f(W, Y_1, \dots, Y_{j-1})$ for $j = 1, \dots, n$.

Theorem ([PolyanskiyVerdu10])

$$\boxed{I_\alpha(W, \underline{Y}) \leqslant n \cdot C_\alpha}$$

Proof.

Simple proof in [RioulNguyen'22] with the inequality

$$D_\alpha(p_{XY} \| q_X q_Y) \leqslant D_\alpha(p_X \| q_X) + \max_x D_\alpha(p_{Y|x} \| q_Y).$$

Outline

Main (α-Converse) Theorem

By combining the above inequalities:

Theorem

For any $\alpha \in [0, +\infty]$ and any block code (n, M) with rate $R = \frac{\log M}{n}$ and decoding error probability $\mathbb{P}_e = 1 - \mathbb{P}_s$ on a memoryless channel (with or without perfect feedback) of α-capacity C_α,

$$\boxed{d_\alpha(\mathbb{P}_s \| \mathbb{P}'_s) \leqslant n \cdot C_\alpha}$$

where $\mathbb{P}'_s = \max_w p_W(w) \leqslant \mathbb{P}_s$

- In particular, $\mathbb{P}'_s = \frac{1}{M}$ for equiprobable messages W.

Main (α-Converse) Theorem

By combining the above inequalities:

Theorem

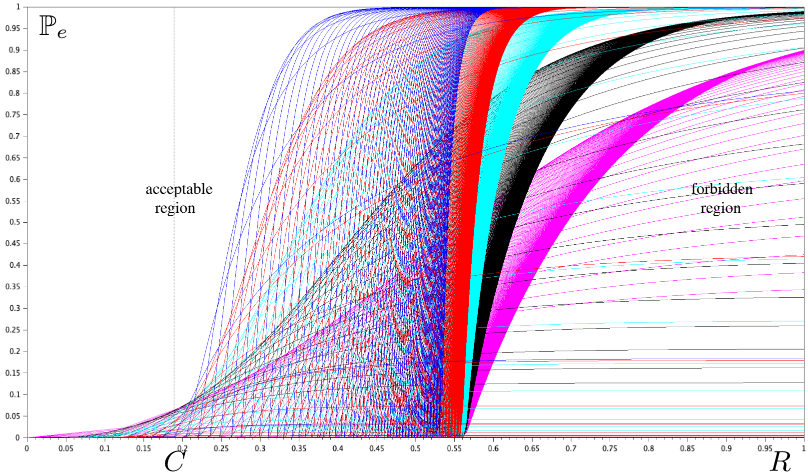
For any $\alpha \in [0, +\infty]$ and any block code (n, M) with rate $R = \frac{\log M}{n}$ and decoding error probability $\mathbb{P}_e = 1 - \mathbb{P}_s$ on a memoryless channel (with or without perfect feedback) of α-capacity C_α,

$$\boxed{d_\alpha(\mathbb{P}_s \| \mathbb{P}'_s) \leqslant n \cdot C_\alpha}$$

where $\mathbb{P}'_s = \max_w p_W(w) \leqslant \mathbb{P}_s$

- In particular, $\mathbb{P}'_s = \frac{1}{M}$ for equiprobable messages W.
- For varying $\alpha \in [0, +\infty]$, gives non-asymptotic lower bounds on $\mathbb{P}_e$ (upper bounds on $\mathbb{P}_s$) for
  - any particular choice of block code parameters (n, M)
  - any choice of code length n with varying coding rate $R = \frac{\log M}{n}$

Lower bounds on error probability $\mathbb{P}_e$ vs. coding rate R



Lower bounds on error probability $\mathbb{P}_e$ vs. coding rate R
on a BSC(.25) for $n = 8$ (magenta), 16 (black), 32 (cyan), 64 (red), 128 (blue).

Theoretical Applications of the Main Theorem

Zero-Error Problem If one requires strictly $\mathbb{P}_e = 0$, the α-converse is optimal for $\alpha \to 0$, which gives

$$R \leqslant C_0 = \max_{p_X} I_0(X;Y) = \max_{p_X} \inf_y \log \frac{1}{\sum\limits_{p_{y|x}>0} p_X}$$

C_0 is the zero-error capacity with feedback (when not all inputs pairs can cause the same output [Shannon'56]).

Theoretical Applications of the Main Theorem

Zero-Error Problem If one requires strictly $\mathbb{P}_e = 0$, the α-converse is optimal for $\alpha \to 0$, which gives

$$R \leqslant C_0 = \max_{p_X} I_0(X;Y) = \max_{p_X} \inf_y \log \frac{1}{\sum\limits_{p_{y|x}>0} p_X}$$

C_0 is the zero-error capacity with feedback (when not all inputs pairs can cause the same output [Shannon'56]).

Strong Converse For $\alpha > 1$, $\frac{1}{\alpha-1} \log(\mathbb{P}_s^{\alpha} \frac{1}{M^{1-\alpha}}) < d_\alpha(\mathbb{P}_s \| \frac{1}{M}) \leqslant nC_\alpha$. Simplifying gives $\mathbb{P}_s < 2^{-n(R-C_\alpha)\frac{\alpha-1}{\alpha}}$. $R > C$ implies $R > C_\alpha + \epsilon$ for some $\alpha > 1$ and $\epsilon > 0$, and $\mathbb{P}_s < 2^{-n\epsilon\frac{\alpha-1}{\alpha}} \to 0$ exponentially.

$$R > C \implies \mathbb{P}_e \text{ tends exponentially to 1 as } n \to +\infty.$$

Arimoto's converse bound [Arimoto'75] can be recovered from this result.

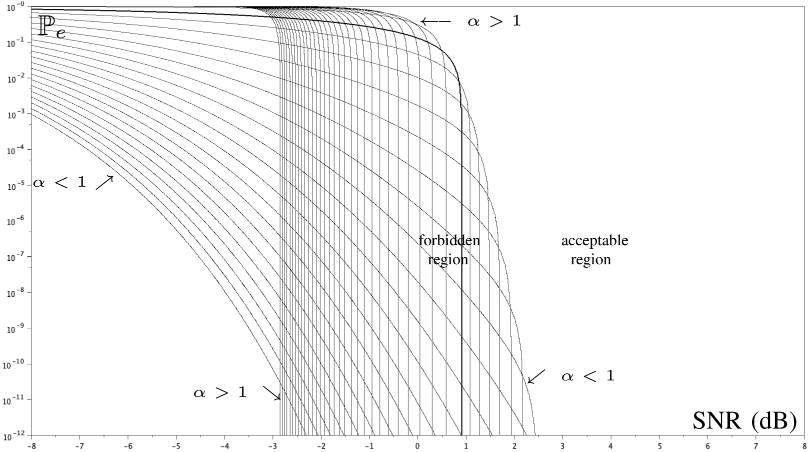
Application: Lower Bound on the SNR

- C_α is expressed in terms of $\frac{1}{\sigma^2} = 2R\cdot\text{SNR}$ per coded bit sent on the channel where $E_b/N_0 = \text{SNR}$ per (information) bit
- since C_α is increasing in SNR, the α-converse theorem gives a lower bound on the feasible SNR for a given performance level $(\mathbb{P}_e, R)$ over a given channel.

Application: Lower Bound on the SNR

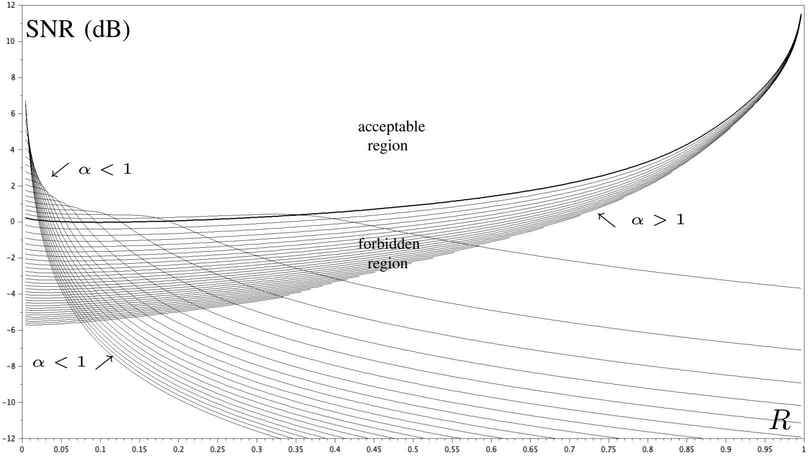
- C_α is expressed in terms of $\frac{1}{\sigma^2} = 2R \cdot$SNR per coded bit sent on the channel where $E_b/N_0 =$ SNR per (information) bit
- since C_α is increasing in SNR, the α-converse theorem gives a lower bound on the feasible SNR for a given performance level $(\mathbb{P}_e, R)$ over a given channel.
- for $n \to +\infty$ and $R \to 0$ we recover the well-known Shannon limits -1.59 dB (binary-input AWGN) and 0.37 dB (BSC)
- non-asymptotic regions for a given choice of code parameters as illustrated below

Lower bounds on error probability $\mathbb{P}_e$ vs. SNR



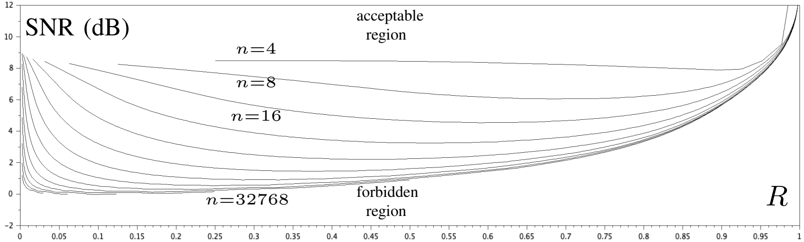
Lower bounds on error probability $\mathbb{P}_e$ vs. SNR
for a $[128, 64]$ code ($n = 128$, $R = 1/2$) on a BSEC. The thick curve is for $\alpha = 1$.

Lower bounds on SNR vs. coding rate



Lower bounds on SNR vs. coding rate for $n = 1024$ on a BSEC.
The thick curve is for $\alpha = 1$.

Max Lower bounds on SNR vs. coding rate



Lower bounds (maximized over α) on SNR vs. coding rate for $n = 4, 8, 16, \ldots, 32768$ on a BSEC.

Conclusions & Perspectives

- α-information theory allows to derive simple non-asymptotic lower bounds on the probability of error for any binary block code used on symmetric memoryless channels with or without feedback [RioulNguyen'22]
- bounds can be rewritten as lower bounds on the SNR for any given code parameters

Conclusions & Perspectives

- α-information theory allows to derive simple non-asymptotic lower bounds on the probability of error for any binary block code used on symmetric memoryless channels with or without feedback [RioulNguyen'22]
- bounds can be rewritten as lower bounds on the SNR for any given code parameters

- since $I_\alpha(X;Y) \neq I_\alpha(Y,X)$, one can also define a "reverse" α-capacity $C'_\alpha = \max_{p_X} I_\alpha(Y;X)$, but [AishwaryaMadiman'20] $C_\alpha \leqslant C'_\alpha$ without feedback.
- compare to other known (finite-length) bounds — sphere packing bounds
- more general types of channels

Conclusions & Perspectives

- α-information theory allows to derive simple non-asymptotic lower bounds on the probability of error for any binary block code used on symmetric memoryless channels with or without feedback [RioulNguyen'22]
- bounds can be rewritten as lower bounds on the SNR for any given code parameters

- since $I_\alpha(X;Y) \neq I_\alpha(Y,X)$, one can also define a "reverse" α-capacity $C'_\alpha = \max_{p_X} I_\alpha(Y;X)$, but [AishwaryaMadiman'20] $C_\alpha \leqslant C'_\alpha$ without feedback.
- compare to other known (finite-length) bounds — sphere packing bounds
- more general types of channels
- other types of problems: α-DPI and α-Fano were recently applied to side-channel analysis [LiuChengGuilleyRioul'21].

α-Capacity of Communication Channels with Feedback: Theoretical Overview

Thank you!

Olivier Rioul
Télécom Paris
Institut Polytechnique de Paris, France

<olivier.rioul@telecom-paris.fr>