Information Theory & Coding

Dec 7, 2020

Wrapping up polar codes

- Seen: given a channel W, the polar transform, applied t-times:
  - synthesizes 2^t channels $\{W^s : s \in \{+,-\}^t\}$ from 2^t uses of the "real" channel W.
  - for any fixed $\varepsilon > 0$,

    fraction of ε-good channels $\xrightarrow[t \to \infty]{} I(W)$

- to convert this to a code seems reasonable, but we need to show something stronger about the ε-good channels: i.e., most of these ε-good channels should also be ε_t-good with $\varepsilon_t \ll \frac{1}{2^t} \Rightarrow P_e(\text{construction}) \xrightarrow[t \to \infty]{} 0$.

to this end, we defined the notion of ε-untainted, & ε-lucky channels.

ε-untainted: W^{s^t} is ε-tainted if

$\exists\ i \in [\sqrt{t}, t)$ s.t. W^{s^i} is ε-ugly

we showed fraction of ε-tainted channels among

$\{W^s : s \in \{+,-\}^t\} \to 0$ as $t \nearrow \infty$

ε-lucky: W^{s^t} is ε-lucky if

$\#$ of $-$ signs in $s^t < (\frac{1}{2}+\varepsilon)t$

$\Rightarrow$ fraction of unlucky channels is $\to 0$ as $t \nearrow \infty$

$\Rightarrow$ the fraction of ε-good, ε-lucky, ε-untainted channels

among $\{W^{s^t} : s^t \in \{+,-\}^t\} \to I(W)$ as $t \nearrow \infty$

Idea: use these channels to communicate data, on the others, send fixed known bits.

We will now show (for the BEC case) that

ε-good, lucky, untainted channels are in fact

ε_t good with $\varepsilon_t \ll 1/2^t$.

$$p(s^{i+1}) = \begin{cases} p(s^i)^2 & s_{i+1} = + \\ 2p(s^i) - p(s^i)^2 & s_{i+1} = - \end{cases}$$

$$\leq \begin{cases} p(s^i)^2 & + \\ 2p(s^i) & - \end{cases}$$

$$z = \varepsilon^{x}$$
$$\log z = x \log \varepsilon$$

$$p(s^{i+1}) \leq \begin{cases} p(s^{i})^{2} & + \\ p(s^{i})\, \varepsilon^{1/\log \varepsilon} & - \end{cases}$$

$$\leq \begin{cases} p(s^{i})^{2} & + \\ p(s^{i})\, p(s^{i})^{1/\log \varepsilon} & - \end{cases}$$

w^{st} is ε-good
- lucky
- untruthful

$$p(s^{i}) \leq \varepsilon$$

$$\log p(s^{i+1}) \leq \log p(s^{i}) \times \begin{cases} 2 & + \\ 1 + \underbrace{\frac{1}{\log \varepsilon}}_{1-\varepsilon'} & - \end{cases}$$

$$\log\log \frac{1}{p(s^{i+1})} \geq \log\log \frac{1}{p(s^{i})} + \begin{cases} 1 & + \\ -\varepsilon'' & - \end{cases}$$

$$-\varepsilon'' = \log(1-\varepsilon')$$

$$\log\log \frac{1}{p(s^{t})} \geq \log\log \frac{1}{p(s^{\leq t})} + \#\{+\text{'s in } (s_{\leq t} \dots s_t)\}$$
$$- \varepsilon'' \{\# -\text{'s} \quad " \quad \}$$

Since w^{s^t} is ε-lucky

the # of $-$ signs in $s_{\sqrt{t}} \ldots s_t$ is $\leq (1+\varepsilon)\frac{t}{2}$

$+$ " " " $\geq t - \sqrt{t} - (1+\varepsilon)\frac{t}{2}$

$$\log\log\frac{1}{p(s^t)} \geq \log\log\frac{1}{p(s^{\sqrt{t}})} + t - \sqrt{t} - (1+\varepsilon)\frac{t}{2} - \varepsilon''(1+\varepsilon)\frac{t}{2}$$

$$= \frac{t}{2}(1-\varepsilon''')$$

So an $\boxed{\varepsilon\text{-good, lucky, untainted}}$ channel w^s has

$$\log\log\frac{1}{p(s^t)} \geq \log\log\frac{1}{\varepsilon} + \frac{t}{2}(1-\varepsilon''').$$

$$\geq \frac{t}{2}(1-\varepsilon'') \quad \text{if } \varepsilon \text{ small enough}$$

$$p(s^t) \leq \underbrace{2^{-2^{\frac{t}{2}(1-\varepsilon'')}}}_{\triangleq\, \varepsilon_t} \ll 2^{-t}$$

We conclude that the error probability of a code constructed as above

$$\leq 2^t \cdot 2^{-2^{\frac{t}{2}(1-\varepsilon'''')}} = 2^{-\underbrace{2^{\frac{t}{2}(1-\varepsilon'''')}}_{\approx \sqrt{n}}}$$

$2^t = n =$ blocklength.

So: we can design polar codes with

- rates up to $I(W)$
- enc / dec complexity $n \log n$
- $p(\text{error}) \approx 2^{-\sqrt{n}}$

as $t = \log_2 n$ gets large.

We saw the proofs for $\underline{W = BEC()}$, but the conclusion holds for any binary-input channels, can be generalized for non-binary inputs, & non-uniform input distributions.

Back to Source coding : Lossy Compression.

Recall : we had discussed "exact" ≡ "lossless" source coding at the beginning of the semester

$$c : \mathcal{U}^n \to \{0,1\}^*.$$

We may want to give up on "exactness" & tolerate some loss, in the hope of improving efficiency. (i.e. fewer bits / letter).

Question : what are the fundamental limits of the tradeoff between quality (= fidelity) & efficiency (≡ bit-rate.)?

We will see "some" answers to this question.

Let us consider the following set-up :

$U_1, U_2, U_3 \ldots$ — iid $\sim P_U$ source.

a lossy source encoder / decoder :

$$\underbrace{enc : \mathcal{U}^n \to \{1, \ldots, M\}}_{(rate = \frac{1}{n}\log_2 M)} \equiv \log_2 M \text{ bit description of } n \text{ source letters}$$

dec: $\{1 .., M\} \to \mathcal{V}^n$

$\log_2 M$ bits description

$\mathcal{U}$: source alphabet

$\mathcal{V}$: reproduction alphabet.

We also have a "distortion measure" d,

$d: \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ $\quad d(u, v)$ is the cost (distortion) (infidelity) of representing u by v.

$u_1 ... u_n \to$ enc $\to$ integer $\{1 .. M\}$ $\to$ dec $\to v_1 ... v_n$

$$\text{distortion} = \frac{1}{n} \sum_{i=1}^{n} d(u_i, v_i)$$

examples $\quad \mathcal{U} = \mathcal{V} = \{0,1\}$. $\quad d(u,v) = \mathbb{1}\{u \neq v\}$

$\mathcal{U} = \mathcal{V} = \mathbb{R} \qquad d(u,v) = (u-v)^2$

$\mathcal{U} = \{0,1\} \quad \mathcal{V} = \{0, 1, ?\} \qquad d(0,0) = 0 = d(1,1)$

$d(0,1) = 10 = d(1,0)$

$d(0,?) = 1 = d(1,?)$

distortion : $\frac{1}{n} \sum' d(u_i, v_i)$ becomes a Random

variable when the input to the enc is $U_1 \ldots U_n$

we will study the trade-off between

$E[\frac{1}{n} \sum' d(U_i, V_i)]$ & $\text{rate} = \frac{1}{n} \log_2 M$

when $U^n \to \boxed{\text{enc}} \to \text{dec} \to V^n$

$\in \{1, \ldots, M\}$.

The prototypical result we will prove is of the following kind:

Thm: given $U_1, U_2 \ldots,$ iid $\sim P_U$

given a reproduction alphabet $\mathcal{V}$ and distortion measure $d: \mathcal{U} \times \mathcal{V} \to \mathbb{R}$.

Compute $\min\limits_{\substack{P_{V|U} \\ E[d(U,V)] = D}} I(U;V) \triangleq R(D)$

$\min\limits_{P_{V|U}} E[d(U,V)] \leq D \leq \max\limits_{P_{V|U}} E[d(U,V)]$

Then any system (i.e. enc/dec)

$$U^n \longrightarrow \boxed{enc} \longrightarrow \boxed{dec} \longrightarrow V^n$$ with

$\underline{E(\text{distortion}) = D}$ has $\underline{\text{rate} \geq R(D)}$.

Before the proof some examples of computation:

$\underline{Ex}$: $U \in \{0,1\}$ $\text{Bern}(\frac{1}{2})$ $p(u=0) = \frac{1}{2}$
$p(u=1) = \frac{1}{2}$

$\mathcal{V} = \{0,1\}$ $d(u,v) = \mathbb{1}\{u \neq v\}$.

given $0 \leq D \leq 1$, how do we compute

$$\min \left\{ I(U;V) : \underline{P_{V|U} \text{ with } E(d(U,V)) = D} \right\}.$$

to this end: for such a V:

$$I(U;V) = H(U) - H(U|V)$$
$$= 1 - H(U \oplus V | V)$$
$$\geq 1 - H(U \oplus V)$$

$$\begin{cases} 0 & \text{if } U = V \\ 1 & \text{if } U \neq V \end{cases}$$

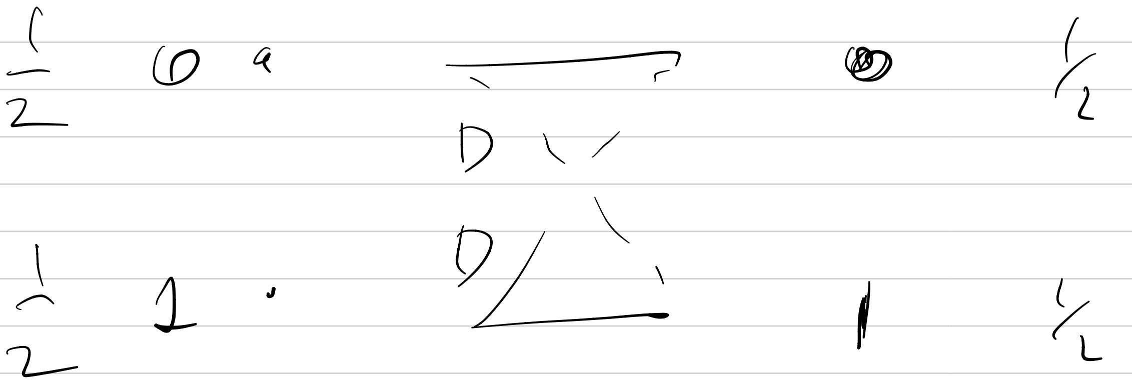
now $U \oplus V = \begin{cases} 1 & \text{w.p. } D \\ 0 & \text{w.p. } 1-D \end{cases}$

$\Rightarrow H(U \oplus V) = h_2(D)$

$I(U;V) \geq \underline{1 - h_2(D)}$.

$\Rightarrow R(D) \geq 1 - h_2(D)$.

now observe:



$U \qquad BSC(D) \qquad V$

$$p(u,v) = \begin{cases} \frac{1}{2} D & \text{if } u \neq v \\ \frac{1}{2}(1-D) & \text{if } u = v. \end{cases}$$

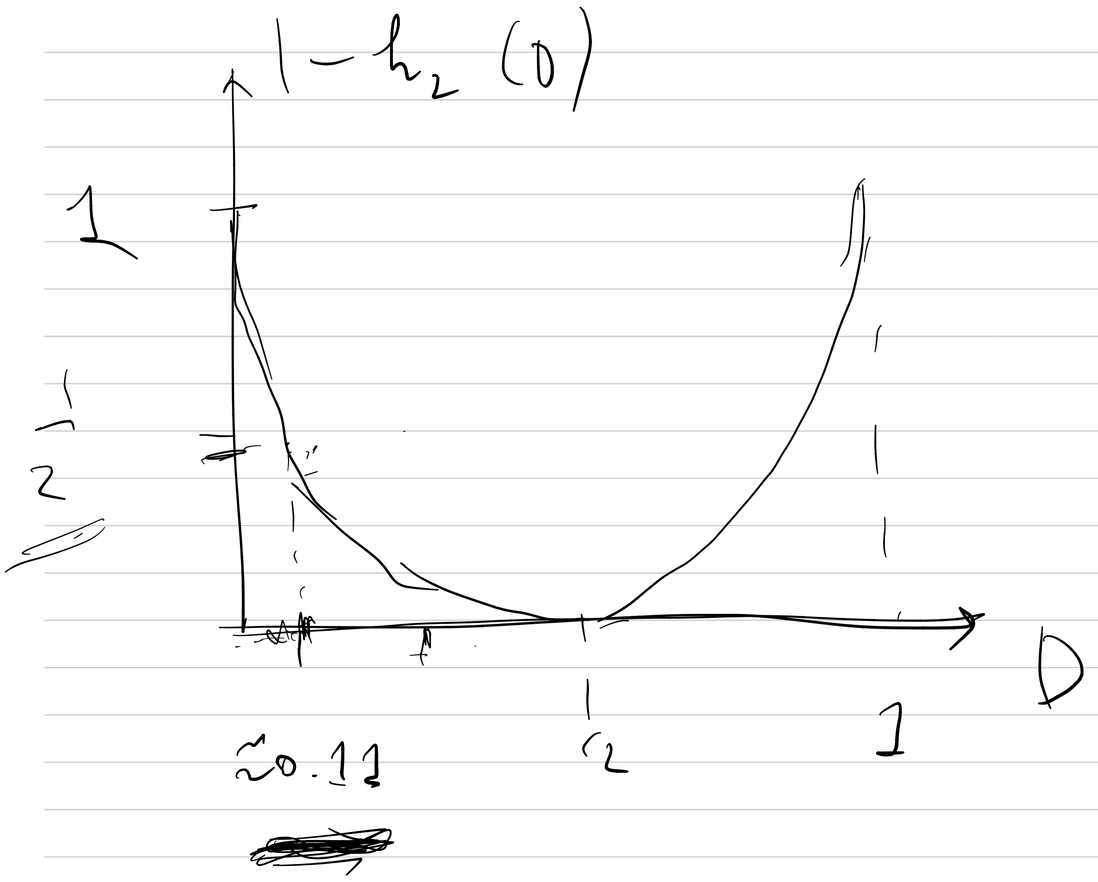
$\underline{U \oplus V}$ = indicator of "flip". $\perp\!\!\!\perp$ of $\underline{U}$.

Because of the symmetry of $p(u,v)$ in $u \leftrightarrow v$.

$U \oplus V$ is indep of V.

$\Rightarrow H(U \oplus V | V) = H(U \oplus V)$, which makes the lower bound to be equality in $I(U;V) \geq 1 - h_2(D)$.

$\Rightarrow R(D) = 1 - h_2(D)$.



We will also show:

Thm: given P_u, iid $U_1, U_2, \ldots$, $\mathcal{V}$, $d(u,v)$, ... $R(D)$ just as in the theorem above, given $\varepsilon > 0$. There exists enc, dec. s.t.

$$U^n \to \text{enc} \to \text{dec} \to V^n$$

1. rate $\leq R(D) + \varepsilon$.
2. $|E(\text{distortion}) - D| < \varepsilon$.

(Tomorrow: proof of these two complementary theorems.)