CS229r - Lecture 15

3/21/2017

Today: Polar Codes [Arikan '09]

<u>Context</u>: Fix $0 < p < \frac{1}{2}$. Want a code C_ϵ for every $\epsilon > 0$

$$n(C_\epsilon) \leq \mathrm{poly}(1/\epsilon)$$

$$\mathrm{Rate}(C_\epsilon) \geq 1 - H(p) - \epsilon$$

C_ϵ corrects p-fraction random errors w.p. $1-\epsilon$ in time $\mathrm{poly}(1/\epsilon)$.

- Forney / Concatenation: time $\geq 2^{1/\epsilon^2}$
- LDPC (Low Density Parity Check): Not known to work for every $\epsilon > 0$.
- Today + Next Lecture: Achieve this with "Information-Theoretic Codes" called "Polar Codes"

Step 1: Linear Compression + Efficient Decompression $\Rightarrow$ Linear Coding + Efficient Decoding

Claim:

Suppose $f_H : \{0,1\}^n \longrightarrow \{0,1\}^{\cancel{H(p)\cdot n}\ m}$

~~is such~~

& $f_H^{-1} : \{0,1\}^{\cancel{1-H(p)\cdot n}\ m} \longrightarrow \{0,1\}^n$

are such that

① f_H is linear ie., $f_H(x) = x \cdot H$

② w.h.p. for $X \sim \text{Bern}(p)^n$

$$f_H^{-1}(f^H(X)) = X$$

then H is parity check matrix for code correcting p-fraction random errors.

Proof: $C = \{ x \mid xH = 0 \}$

Transmit $x \longrightarrow$ recieve $z = x + y$; $y \sim \text{Bern}(p)^n$

$(x+y)H = yH$; $f^{-1}((x+y)H) = y$ whp

Rest of lectures

Constructing Linear Compressor By Inf. Theory

1 Information Theory Basics

1.1 Centropy of X distributed on [N] with

$$\Pr[x=i] = p_i$$

$$H(X) = \sum p_i \log_2 \frac{1}{p_i} = \mathop{E}_{X\sim p}\left[\log_2 \frac{1}{\Pr(x)}\right]$$

1.2 Conditional Entropy

for jointly distributed random variables (X,Y)

$$H(X|Y) \triangleq \cancel{H} \mathop{E}_{Y}\left[H(X|Y)\right]$$

$$\triangleq \sum_j \Pr[Y=j] \cdot H(X \mid Y=j)$$

Axioms

1. $H(X,Y) = H(X) + H(Y|X)$
2. $H(X) \geq H(X|Y)$ [
3. $H(X) \leq \log_2 N \quad (X \in [N])$

THE POLARIZATION APPROACH

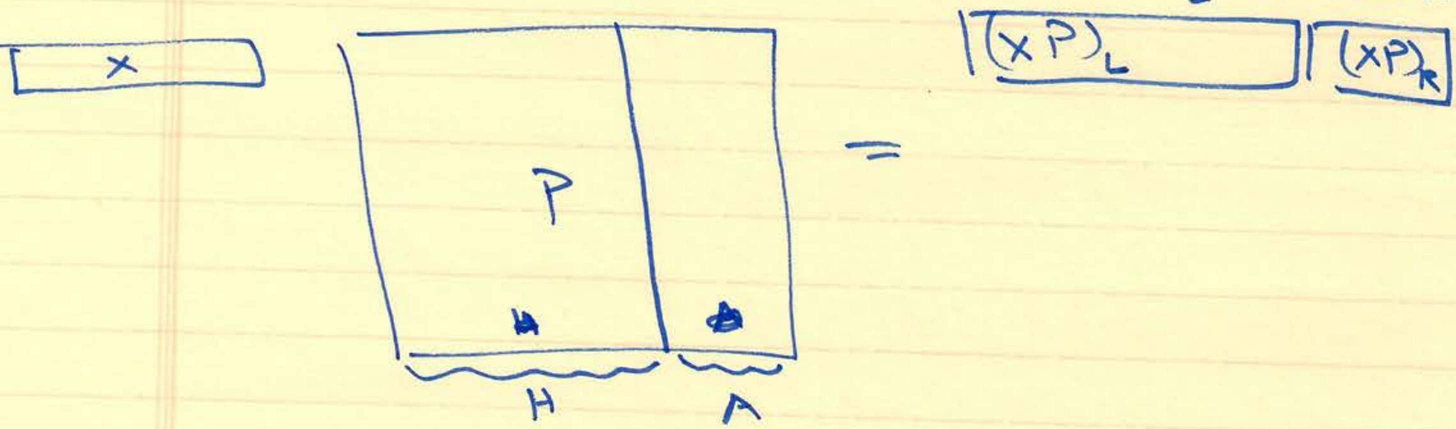
Build $n \times n$ invertible matrix P s.t.

xP splits into a left ~~half~~ part $(xP)_L$

& right part $(xP)_R$

s.t. $H((xP)_R \mid (xP)_L)$ is tiny (think zero).

$[x \sim \mathrm{Bern}(p)^n]$



if so then Information-theoretically $(xP)_L$ (almost always) specifies $(xP)_R$

$$\Rightarrow \quad (xP)_L \rightsquigarrow (xP)_L (xP)_R \underset{P^{-1}}{\rightsquigarrow} x$$

[gives decompression].

- But Entropy of x is $n \cdot H(p)$.
- $H(xP) = n \cdot H(p) \Rightarrow H((xP)_L) \approx n \cdot H(p)$

Can we squeeze all entropy into left part?

Some Calculations

if $H((xP)_R \mid (xP)_L) \leq \delta$

then $\exists f$ s.t. $\Pr_x\left[(xP)_R \neq f((xP)_L)\right] \leq 2\sqrt{\delta}$

Proof: Let $q_a \triangleq \Pr[(xP)_L = a]$

$$q_{b|a} \triangleq \Pr\left[(xP)_R = b \;\middle|\; (xP)_L = a\right]$$

Then

$$H((xP)_R \mid (xP)_L) = \sum_a q_a \cdot H(q_{\cdot|a}) \leq \delta$$

$$\Rightarrow \Pr\left[H(q_{\cdot|a}) > \sqrt{\delta}\right] \leq \sqrt{\delta}$$

When $H(q_{\cdot|a}) \leq \sqrt{\delta}$ we have

$\exists\, b$ s.t. $q_{b|a} \geq 1 - \sqrt{\delta}$

follows from
$[\,H(p) \geq p\,]$

[So decoding works with some uncertainty in $(xP)_R \mid (xP)_L$] $\boxtimes$

Polarization Phenomenon ("XOR Lemma")

① if (X, Y) are independent Bern(p) variables $\left[0 < p < \frac{1}{2}\right]$

then $H(X \oplus Y) > \max\{H(X), H(Y)\}$

② So the map

$$(X, Y) \longrightarrow (X \oplus Y,\ Y)$$

leads to a "polarized bit" $X \oplus Y$

& an conditionally "less random bit" $Y \mid X \oplus Y$.

Proof of ①: Let $X, Y \in \{-1, 1\}$ (w.p. p, w.p. $1-p$)

then $E[X] = E[Y] = 1 - 2p$

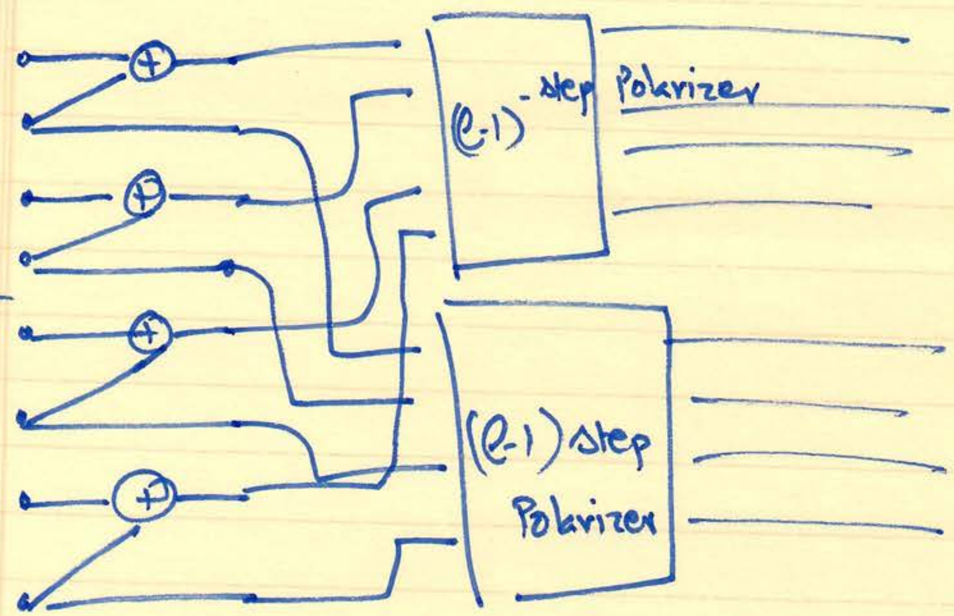
$$E[\text{"}X \oplus Y\text{"}] = E[X \cdot Y] = E[X] \cdot E[Y]$$

$$= (1-2p)^2 < 1 - 2p.$$

$\Rightarrow$ monotonicity of Entropy $\Rightarrow H(X \oplus Y) > H(X) = H(Y)$

ℓ-step Polarization

$N = 2^\ell$



1. Start with $N = 2^\ell$ bits
2. XOR odd & even bits ; ~~send~~
3. - $(\ell-1)$ - polarize XOR-ed pair
   - $(\ell-1)$ · polarize even bits
4. Output all 2^ℓ outputs in step 3 .

<u>Claim</u> : "<u>Conditional Entropies</u>" of Output getting "<u>Polarized</u>"

Specifically

Input $= (X_1 \dots X_N)$; output $= X \cdot P = (Y_1 \dots Y_N)$

$$\eta_i \triangleq H(Y_i \mid Y_1 \dots Y_{i-1})$$

Claim As $\ell \to \infty$

$$\# \left\{ i \mid \eta_i \in \left(\frac{1}{N^2}, 1 - \frac{1}{N^2} \right) \right\} = o(N)$$

$$\boxed{\text{Proof : Next Lecture}}$$

Assuming Claim: Why does P work for us.

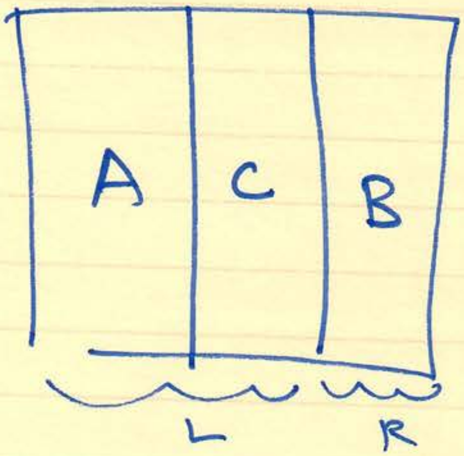
① $A \triangleq \left\{ i \mid \eta_i \geq 1 - \frac{1}{N^2} \right\}$

$B \triangleq \left\{ i \mid \eta_i \leq \frac{1}{N^2} \right\}$

$C \triangleq \left\{ i \mid \eta_i \in \left(\frac{1}{N^2}, 1 - \frac{1}{N^2} \right) \right\}$

$$\left.\begin{aligned} |A| &= ? \\ |B| &= ? \\ |C| &= ? \end{aligned}\right\} \quad \begin{aligned} &\leq H(p) \cdot N \text{ ② (or else we are manufacturing entropy).} \\ &\geq (1 - H(p) - o(1))\, N \text{ ③} \Leftarrow \text{①+②} \\ &\leq o(N) \text{ by Claim ①} \end{aligned}$$

Permute columns of P so that



& we now have $H\left((xP)_R \mid (xP)_L\right)$

$$\leq \frac{1}{N^2} \cdot |R| \leq \frac{1}{N}$$

$\Rightarrow \Pr[\text{decoding failure}] \leq O\left(\frac{1}{\sqrt{N}}\right)$.

So works great information theoretically;

But Algorithm?

Will get algorithm to "guess"

$(xP)_i$ given $(xP)_1^{i-1}$

when $H\left((xP)_i \mid (xP)_1^{i-1}\right) \leq \frac{1}{N^2}$

SUCCESSIVE CANCELLATION DECODER.

(Also Next Lecture)