April 30, 2019

REMINDERS

(*) Mini project video due by 11:59pm tonight
↳ If >1 group member, everyone should submit PDF with Youtube video link individually

(*) Guest lecture this Th: Sai

(*) 3rd proof reading schedule changed (if you were assigned this Th)

(*) Please fill in course evaluations!

RECAP:

Thm'': If ∃ an efficient linear compression for $\mathrm{Bern}(p)^n$ $\Rightarrow$ efficient coding scheme for BSC_p

Linear compression scheme: (H, D) is a τ-error linear compression:

(•) $H \in \mathbb{F}_2^{n \times m}$ + full rank ($m \le n$)

(•) $D : \mathbb{F}_2^m \to \mathbb{F}_2^n$

rate $= \frac{m}{n}$

(•) $\Pr_{\bar{Z} \in \mathrm{Bern}(p)^n} [D(\bar{Z} \cdot H) \neq \bar{Z}] \le \tau$

Additional matrices for reduction:

(•) $G \in \mathbb{F}_2^{k \times n}$; $G^* \in \mathbb{F}_2^{n \times k}$ s.t. $E(\bar{X}) = \bar{X} \cdot G$

$D'(Y) = (\bar{Y} - D(\bar{Y} \cdot H)) \cdot G^*$

(i) $G \cdot H = 0$ (ii) $G \cdot G^* = I_k$

(G: generator matrix; H: transpose of parity check matrix)

Complexity of enc/dec: Matrix vect mult for G, H, G^*

GOAL: Design ε-error linear compression w/ rate $H(p) + \varepsilon$ (+ G, H, G^* mat vect mult in $O(n \log n)$ time

Polar codes [Arikan, '08]

INFORMATION THEORY! r.v. X ; $H(X) = \sum_{x \in \mathrm{supp}(X)} \Pr(x) \log_2 \frac{1}{\Pr(x)}$

→ $H(X) \le \log_2 |\mathrm{supp}(X)|$ (= for uniform distribution)

Conditional entropy: $H(Y|X) = \mathop{\mathbb{E}}_{x} H(Y|X=x)$

Chain rule: $H(X,Y) = H(X) + H(Y|X)$ $\xrightarrow{X, Y \text{ indep}}$ $H(Y|X) = H(Y)$

TODAY (Proof readers: Amy, Anna, Yichen)

→ Bit more about entropy

→ Start w/ polarization / polar codes (to be continued next week)

3 other properties of entropy:

(•) $H(Y|X) \le H(Y)$ {conditioning cannot increase entropy}

(•) $f: \text{supp}(X) \to \text{supp}(X)$ is a bijection

$\Rightarrow H(f(X)) = H(X)$

Cor: $\bar{X} \in \{0,1\}^n$, $M \in \mathbb{F}_2^{n \times n}$ full rank matrix

$H(\bar{X} \cdot M) = H(\bar{X})$

(•) (Proved in Sec 11.6.1 in book)

Prop 1: (X,Y) be jointly distributed s.t $H(X|Y) \le \alpha$ (<<1)

$\forall y \in \text{supp}(Y)$ $A(y) = \arg\max_x \Pr[X = x | Y = y]$

$\Rightarrow \Pr_{(X,Y)}[X \neq A(Y)] \le \alpha.$

Polarization phenomenon (... Polar codes)

$n = 2$ $\quad \bar{Z} = (Z_1, Z_2) \in \text{Bern}(p)^2$

$\to H(Z_1) = H(p) = H(Z_2)$

$\to H(\bar{Z}) \overset{\text{chain rule}}{=} H(Z_1) + H(Z_2|Z_1) \overset{Z_1, Z_2 \text{ indep}}{=} H(Z_1) + H(Z_2)$

$= 2H(p)$

$M \in \mathbb{F}_2^{2\times 2}$ full rank. $\quad H(\bar{Z} \cdot M) = H(\bar{Z}) = 2H(p)$

Case 1: $M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad \bar{W} = \bar{Z} \cdot M$ $\quad [(Z_1\ Z_2) M = (W_1\ W_2)]$

$W_1 = Z_1$, $W_2 = Z_2$

$H(W_1) = H(p)$

$H(W_2|W_1) = H(p)$

Case 2: $M_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ $\quad W_1 = Z_1$, $W_2 = Z_1 + Z_2$

$H(W_1) = H(p)$

$H(W_2|W_1) = H(W_2, W_1) - H(W_1) = H(p)$

Case 3: $M_3 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ $\quad W_1 = Z_1 + Z_2$, $W_2 = Z_2$

$H(W_1) = H(Z_1 + Z_2) = H(2p(1-p)) > p$ for $0 < p < \frac{1}{2}$

$H(W_2|W_1) = H(W_1, W_2) - H(W_1) > H(p)$

$< H(p)$

⇒ (i) entropies polarize
(ii) The order of taking sum & conditioning matters.

General polarization idea:

Design a full rank $P \in \mathbb{F}_2^{n \times n}$ s.t

$\bar{W} = \bar{Z} \cdot P = (W_1, \ldots, W_n)$

$\bar{Z} \in Bern(p)^n$ want $\forall i$ $H(W_i \mid W_1, \ldots, W_{i-1}) \in \{0,1\}$

$S = \{ i \mid H(W_i \mid W_1, \ldots, W_{i-1}) = 1 \}$

Compression: $(\bar{Z} \cdot P)_S$

Q: How large can $|S|$ be?

$H(\bar{Z} \cdot P) = H(\bar{Z}) = n \cdot H(p)$

By chain rule $n \cdot H(p) = H(\bar{W}) = \sum_{i=1}^{n} H(W_i \mid W_1, \ldots, W_{i-1}) = |S|$

We will not get "perfect polarization" $H(W_i \mid W_1, \ldots, W_{i-1}) \le \tau$

Def: $\bar{W}_{<i} = (W_1, \ldots, W_{i-1})$

Call $i \in [n]$ predictable if $H(W_i \mid \bar{W}_{<i}) \le \tau$
o/w unpredictable

$S \stackrel{def}{=} \{ i \mid H(W_i \mid \bar{W}_{<i}) \ge \tau \}$

Q: How small/large can $|S|$ be?

$n H(p) = H(\bar{Z} \cdot P) = H(\bar{W}) = \sum_{i=1}^{n} H(W_i \mid \bar{W}_{<i})$

$\le \sum_{i \in S} H(W_i \mid \bar{W}_{<i}) + \tau n$

$\le \sum_{i \in S} H(W_i) + \tau n$

$\le |S| + \tau n \Rightarrow |S| \ge n H(p) - \tau n$

Def (Polarizing matrix, unpredictable columns of P)

$P \in \mathbb{F}_2^{n \times n}$ invertible. P is (ε, τ) polarizing for $\text{Bern}(p)^n$ if:

(•) $\bar{W} = \bar{Z} \cdot P$

(•) $S = \{ i \mid H(W_i \mid \bar{W}_{<i}) \geq \tau \}$

$\Rightarrow$ $|S| \leq n \cdot H(p) + \varepsilon n$

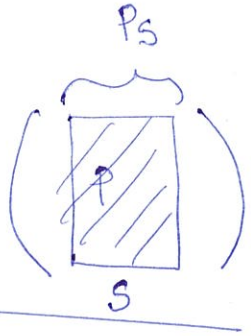
$S \to$ unpredictable columns of P ($\{W_i\}_{i \in S}$ are unpredictable bits)

Next: Polarizing matrices are enough.

Q1: H? $= P_S$

$\bar{W} = \bar{Z} \cdot P$

Compression: $(\bar{Z} \cdot P)_S$ ($\bar{Z}$) $= \bar{Z} \cdot P_S$



Side note: $G = (P^{-1})_{\bar{S}}$ $\quad G^* = P_{\bar{S}}$

($\Rightarrow$ final complexities determined by matrix vect Px & $P^{-1}x$)

Q2: D? Successive Cancellation Decoder (SCD) ($\bar{W} = (\bar{Z} \cdot P)_S$)

i/p: $\bar{W} \in \mathbb{F}_2^S$, $P \in \mathbb{F}_2^{n \times n}$

o/p: $\tilde{Z} \in \mathbb{F}_2^n$ s.t. $(\tilde{Z} \cdot P)_S = \bar{W}$

s.t $\Pr_{\bar{Z} \in \text{Bern}(p)^n} [\tilde{Z} \neq Z]$ is small!

Goal: Let P be an (ε, τ) polarizing

$\Rightarrow \Pr_{\bar{Z} \in \text{Bern}(p)^n} [\bar{Z} \neq \text{SCD}((\bar{Z} \cdot P)_S, P, S)] \leq \tau \cdot n$

$\Rightarrow$ What we want: $(\varepsilon, \frac{\varepsilon}{n})$ polarizing matrix P s.t Px, $P^{-1}x$ can be done in $O(n \log n)$ time

Idea: Construct $\tilde{W} \in \mathbb{F}_2^n$ s.t $\tilde{W}_S = \overline{W}$

"Somehow guess" rest of $\tilde{Z} \cdot P$ s.t $\tilde{W} = \tilde{Z} \cdot P$

$\Rightarrow$ output $\tilde{Z} = \tilde{W} \cdot P^{-1}$

SCD $(\overline{W}, P, S)$

(1) $i = 1 \dots n$

if $i \in S$ $\tilde{W}_i = W_i \leftarrow (i \in S)^{th}$ value in $\overline{W}$

else $\tilde{W}_i = \max_{b \in \mathbb{F}_2} \Pr[W_i = b \mid \overline{W}_{<i} = \tilde{W}_{<i}]$

(2) return $\tilde{Z} \leftarrow \tilde{W} \cdot P^{-1}$