Today

- Polar Codes I.

CS3.318.1
Coding Theory
Lecture 21 (2022-11-14)
Instructor: Prahladh Harsha.

---

Recalling Shannon's Thm for BSC.


Code up w/ $E, D$

$$\Pr[m' \neq m] = o(1).$$

Shannon's Thm:

$\forall\, p \in (0,1)$, $\forall\, \varepsilon \in (0, \frac{1}{2} - p)$

$\forall$ sufficiently large $n$

$\exists\, k$; $k \geq \lfloor (1 - H(p+\varepsilon))n \rfloor$ ; $\delta > 0$

s.t. $E: \{0,1\}^k \to \{0,1\}^n$ ; $D: \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$

$$\Pr_{m,\eta}[D(E(m)+\eta) \neq m] \leq 2^{-\delta n}$$

Non-constructive:

Forney : Via Concatenation.

"Constructive" Version of Shannon's Th

C Construction: Generator matrix : $poly(n) + 2^{O(1/\varepsilon^5)}$

E : Encoding time $O(n^2)$

D : Decoding time : $poly(n) + n \cdot 2^{O(1/\varepsilon^5)}$

Dependence on $\varepsilon$ - exp in decoding time.

Qn: Can we get a truly poly version of Shannon's Thm

History:

1995: Luby, Mitzenmacher, Shokrollahi, Spielman (posed this question)

2008: Arikan (discovered Polar Codes)

2013: Guruswami - Xia
Hassani, Alishahi, Urbanke } Analysis of Polar codes

2018 : Blasiok, Guruswami, Nakkiran, Rudra + Sudan (cleaner analysis)

Thm There exists an explicit polynomial $n_0(\cdot)$

$\forall \varepsilon$. $\exists\, n \in (1/\varepsilon, n_0(1/\varepsilon))$

$k \geq \lfloor (1 - H(p+\varepsilon))n \rfloor$

$\exists\, E, D$ - $O(n \log n)$ ;
failure probability $< \varepsilon$.

## Linear Codes:

Specified $G_y$

$x \in \mathbb{F}_2^k$

$G$ : $n \times k$ matrix

Generator Matrix

$x \mapsto Gx$

or

$H$ : $(n-k) \times n$ matrix, $m = n-k$.

Parity check Matrix

$Gx+\eta \rightsquigarrow$

What happens if we apply $H$ on noisy codeword $Gx+\eta$

$H(Gx+\eta) = H\eta$

## Compression Scheme: for $(\mathrm{Ber}(p))^n$ : w/ error $\varepsilon$

$(H, D)$ : $H \in \{0,1\}^{m \times n}$ (compressing $n$ bits to $m$ bits)

$D : \{0,1\}^m \to \{0,1\}^n \cup \{\perp\}$

$$\Pr_{Z \leftarrow (\mathrm{Ber}(p))^n} [D(HZ) \neq Z] \leq \varepsilon.$$

Claim: $G$ is a "good" generator matrix for $BSC(p)^n$ (error $\leq \varepsilon$)

$\Leftrightarrow$

$H$ is a "good" compressor for $(\mathrm{Ber}(p))^n$ (error $\leq \varepsilon$)

Pf: Decoding Alg: Encoding

Input- $x$ $(= Gx + \eta)$ $x \mapsto Gx$

Algorithm: 1. Apply $H$ on $x$ to obtain $Hx$ $(= H\eta)$

2. Apply $D$ on $Hx$ to obtain $\eta'$

3. Output $x - \eta'$

**Goal: Design a Linear Compression Scheme for $(Ber(p))^n$ $n \to H(p)n$**

**Brief Diversion:**

**Primer on Information Theory:**

1 **Entropy:** $X$ - random variable on finite set $M$

$p_i = Pr[X = i]$ $\forall i \in M$

$$H(X) := \sum_{i \in M} p_i \log \frac{1}{p_i}$$

(informally the number of bits needed to encode a random sample from $X$).

(a) Non-negative quantity (ie $H(X) \geq 0$, $\forall X$)

(b) $H(X) \leq \log_2 M$ if $Supp(X) \subseteq M$.

(c) $f: M \to M$ Bijection

$H(f(X)) = H(X)$

eg: $(U, V)$ – Circle

$H(U, V) = H(U+V, V)$

$X, Y$ – joint pair of random variables

(a) $X, Y$ – independent $(\forall i \in [M], j \in [N]$

$P[X=i, Y=j] = P[X=i] \cdot P[Y=j]$

$p_{ij} = p_i \cdot q_j$

$H(X, Y) = H(X) + H(Y)$

eg: $X \sim Ber(p)$

$H(X) = H(p)$

$X_1 \ldots X_n \sim (Ber(p))^n$

$H(X_1 \ldots X_n) = n H(p)$

## 2. Conditional Entropy:

$X, Y$ – joint pair of r.v.s

$H(X)$, $H(Y)$, $H(X, Y)$

$H(Y/X) := H(X, Y) - H(X)$

$= \mathbb{E}_{x \leftarrow X}[H(Y/X=x)]$

(a) Chain Rule: $H(X, Y) = H(X) + H(Y/X)$

$H(X_1 \ldots X_n) = H(X_1) + H(X_2/X_1) + H(X_3/X_1 X_2)$

$+ \ldots + H(X_n/X_1 \ldots X_{n-1})$

(b) $H(Y/X)$ vs $H(Y)$

Conditioning reduces Entropy: $H(Y/X) \leq H(Y)$

**Proposition:**

(1) $X$ is a r.v. on supp $M$ s.t. $H(X) \leq \alpha$ for $\alpha \in (0,1)$

then $\exists\, x \in M$; $\Pr[X \neq x] \leq \alpha$.

(2). $(X, Y)$ - joint r.v. on $M \times N$ s.t. $H(Y|X) \leq \alpha$ for some $\alpha \in (0,1)$

If $A(x) = \underset{y \in N}{\operatorname{argmax}} \{ \Pr[Y = y \mid X = x] \}$

then $\Pr_{x,y}[Y \neq A(x)] \leq \alpha$.

---

**Back to linear compression scheme:**

$P \in \{0,1\}^{n \times n}$ - invertible matrix.

$Z = (Z_1 \ldots Z_n) \sim (\mathrm{Ber}(p))^n$

$H(Z) = H(p) \cdot n$

$H(PZ) = H(Z) = H(p) n.$

$H(Z) = H(Z_1) + H(Z_2) + \ldots + H(Z_n)$

(since $Z_i$'s are independent)

However, this is no longer true for $PZ$

$W = PZ$

$H(W) \neq H(W_1) + H(W_2) + \ldots + H(W_n)$

$$H(W) = H(W_1) + H(W_2|W_1) + \dots + H(W_n|W_1 \dots W_{n-1})$$

$$= \sum_{i=1}^{n} H(W_i|W_{<i})$$

$\tau \in (0,1)$ : $S_\tau \triangleq \{ i \in [n] \mid H(W_i|W_{<i}) \geq \tau \}$

$S_\tau$ – set of "unpredictable" bits in $W$

$$H(W) = \sum_{i=1}^{n} H(W_i|W_{<i})$$

$$= \sum_{i \in S_\tau} H(W_i|W_{<i}) + \sum_{i \notin S_\tau} H(W_i|W_{<i})$$

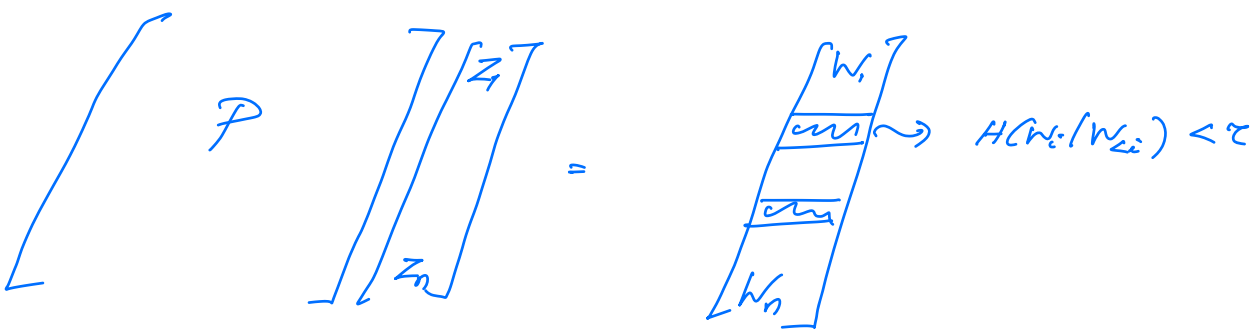
$$\leq \sum_{i \in S_\tau} 1 + \sum_{i \notin S_\tau} \tau$$

$$= |S_\tau| + (n - |S_\tau|) \cdot \tau$$

$$\leq |S_\tau| + n \cdot \tau$$

$$|S_\tau| \geq n \cdot H(P) - n \cdot \tau$$

$P \in \{0,1\}^{n \times n}$ is an $(\varepsilon, \tau)$-polarizing matrix if $|S_\tau| \leq (H(P) + \varepsilon) n$



**Compression:**

Input: $Z$

1. Apply $P$ to $Z$
2. Output $(PZ)|_{S_\varepsilon}$

**Decompression:**

Input: $P$, $S$, $W : S \to \{0,1\}$

(Typically, $W = (PZ)|_S$)

Performance Parameter: $\varepsilon$.

Algorithm:

For $i \leftarrow 1$ to $n$.

if $i \in S_\varepsilon$

$\tilde{W}_i \leftarrow W_i$

else $\tilde{W}_i \leftarrow \operatorname{argmax}_b \{ P_{\mathrm{Ber}}[W_i = b \mid W_{\alpha} = \tilde{W}_{\alpha}] \}$

Output $\tilde{W}$.

**Proposition:** If $P$ is an $(\varepsilon, \tau)$-polarizing matrix then the above compression scheme has error at most $\tau n$.

$$\Pr_{Z \leftarrow \mathrm{Ber}(p)^n} \left[ D\left((PZ)|_S\right) \neq Z \right] \leq \tau n.$$

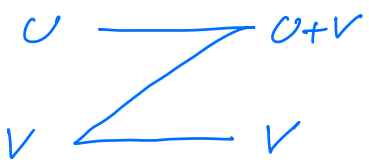
Qn: Does there exist a polarizing matrix?

To begin.

$n = 2$

$Z = (U, V) \longrightarrow (W_1, W_2)$

$(U+V, V)$




$H(U+V, V) = H(U) + H(V)$

$U, V \sim Ber(0.01)$

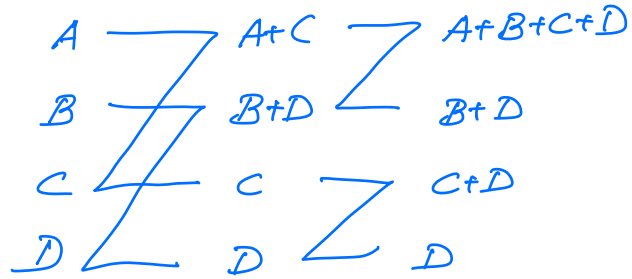
$U+V \sim Ber(0.0199\ldots$
$= Ber(0.02)$

$H(U+V) > H(U), H(V)$

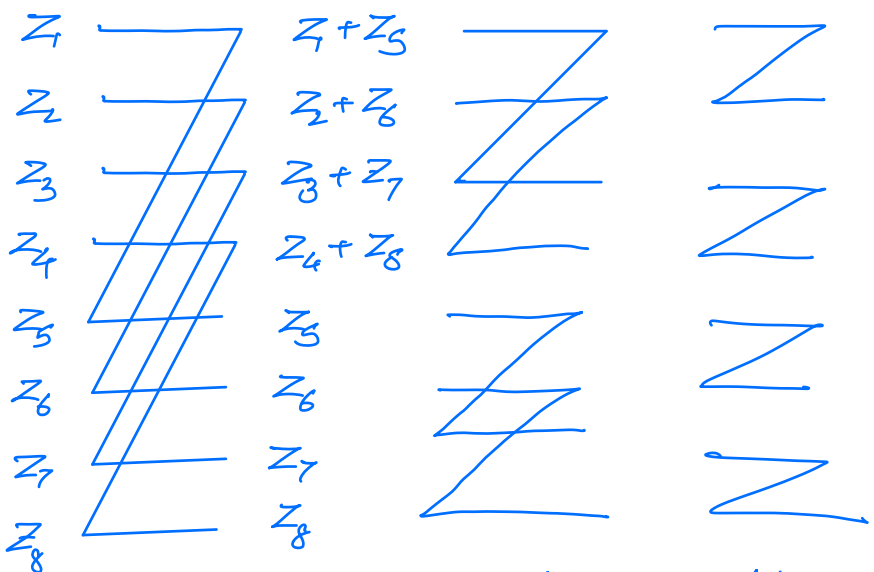
$H(V/U+V) < H(U), H(V)$

Для каждого $n$, кроме:
$n = 2^t$

$n=4$




$n=8$




$\{0,1\}^{2^t} \ni Z = (U, V) \in \{0,1\}^{2^{t-1}} \times \{0,1\}^{2^{t-1}}$

$$PZ = P_\varepsilon Z = \left(P_{\varepsilon-1}(C+V),\ P_{\varepsilon-1}(V)\right)$$

Next lecture, $P$ is a packaging matrix.