CS 229r Information Theory in Computer Science February 26, 2019

Lecture 9

Instructor: Madhu Sudan *Scribe: Amal Mattoo*

1 Overview

1.1 Outline for Today

Polar Coding Continued

- Decoding
- Towards Analysis of Polarization
  - Martingales
  - Local vs. Global Polarization

1.2 Reminders

- Reminder: PS Due today at 8pm
- Madhu Office Hours after lecture
- Mitali Office Hours 4:30 pm

2 Review of Polarization

We built a code by building a tree of polarization. Take n bits i.i.d by Bern(p).

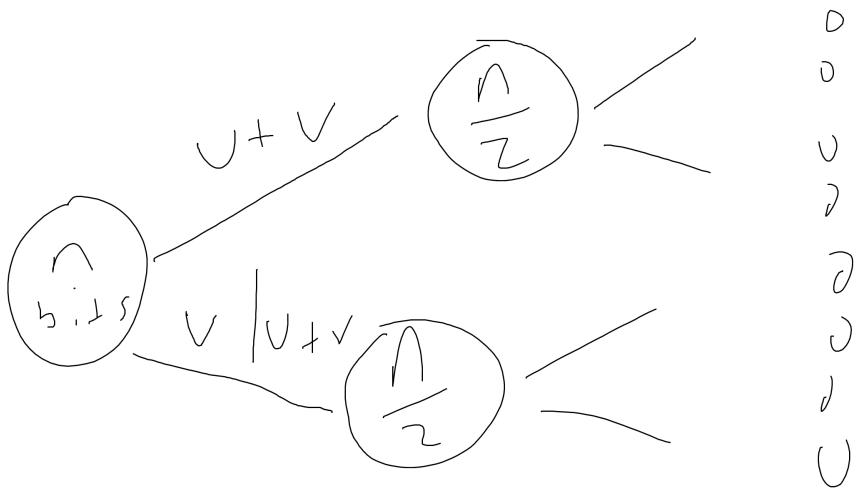

Figure 1: Building the tree of polarization

We claim that they all have very high or very small entropy.

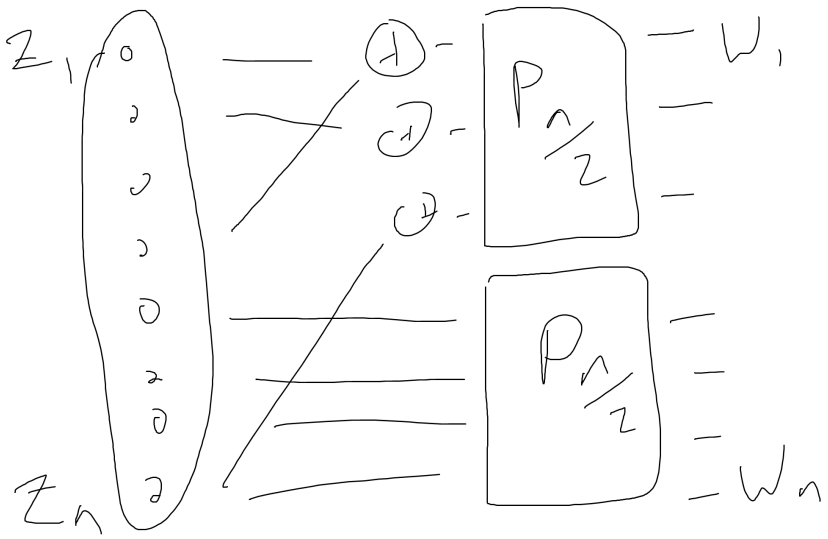

Figure 2: Redrawing the Tree

Theorem 1. *(key theorem to be proved)* $\forall p, c, \exists \beta < 1$ *s.t.* $\forall t$*:*

$$Pr_{j \in [2^t]} \left[H(W_j | W_{cj}) \in (c^{-t}, 1 - c^{-t}) \right] \leq O(\beta^t)$$

3 Decoding

Let

$$S := \{j | H(W_j | W_{<j}) \geq c^{-t}\}$$

As seen in the figure, starting with $z_1, .., z_n$ we obtain $w_1, ..., w_n = W$. We want to compute W_S, that is, W restricted to those indices in S. Let $W = P_n(U, V) = \left(P_{\frac{n}{2}}(U + V), P_{\frac{n}{2}}(V)\right), U, V \in \mathbf{F}_2^{\frac{n}{2}}, P_1(z) = z$.

Given $W \in \{0, 1, ?\}^n$. We want Z s.t. $P_n(Z)_S = W_S$ and has largest probability among all such Zs. We decode recursively as follows:

1. Try to guess $\hat{A}$ (hopefully $U + V$)
2. Try to guess $\hat{B}$ (hopefully V)
3. Combine to get $\hat{U}, \hat{V}$ (hopefully U, V) and this is $\hat{A} - \hat{B}, \hat{B}$.

Deciding the top half of the diagram is obvious since the map is invertible. In the bottom half of the picture, we have erased a lot of information, so we must recover it from the top half.

Assume p (Bernouli probability of the original n bits) is small, around 0.01.

- If $X + Y = 1$, then $\Pr[Y = 1 | X + Y = 1] = \frac{1}{2}$. This is bad, because we can't expect to compress a Bernouli variable with probability $\frac{1}{2}$. Fortunately, it is unlikely.
- If $X + Y = 0$, then $\Pr[Y = 1 | X + Y = 0] \approx 0.001$.

A common strategy when recursion is unsuccessful is to use a strengthened inductive hypothesis. We will now assume not identical distribution, but only independence, proceeding as follows:

To decode $(W \in \{0,1,?\}^n, P_1, ..., P_n)$
$(P_i = P_0[z_i = 1],\ z_1, ..., z_n$ independent)

- Compute the biases $q_1, ..., q_{\frac{n}{2}}$, where $q_i = \Pr[X + Y = 1 | (X,Y) \sim \text{Bern}(P_i) \times \text{Bern}(P_{\frac{n}{2}+i})]$
- $\hat{A} = \text{Decode}(W_1, ..., W_{\frac{n}{2}}; q_1, ..., q_{\frac{n}{2}})$
- Compute $\gamma_1, ..., \gamma_{\frac{n}{2}}$ for $\gamma_i = \Pr[Y = 1 | X + Y = \hat{A}_i, (X,Y) \sim \text{Bern}(P_i) \times \text{Bern}(P_{\frac{n}{2}+i})]$
- $\hat{B} = \text{Decode}(W_{\frac{n}{2}+1}, ..., W_n; \gamma_1, ..., \gamma_{\frac{n}{2}})$
- Output $(\hat{A} - \hat{B}, \hat{B})$,

Note that we have not yet used any of our work with polarization. This arises in the base case:

Base case: $n = 1$. $\text{Decode}(W_1; P_1)$

- If $W_1 \in \{0,1\}$ output (W_1)
- If $W_1 =?$ output 1 if $P_1 \geq \frac{1}{2}$ and 0 otherwise.

If there is an error, there must be a first error, and we sum over the probability of j being the first error. Since our conditional biasing is correct, the probability of getting a bit wrong is the conditional entropy (by problem set 1), and by our polarization the entropy is low.

$$\Pr[\text{making an error}] \leq n\Pr[\text{making an error at location } j] \leq nH(W_j | W_{<j}) \leq nc^{-t}$$

Remark For alphabets of size greater than 2, in general we only apply polar coding for those of finite field size. Just having the additive properties of a group is probably not sufficient.

Exercise 2. *Determine necessary and sufficient conditions on an alphabet for polar coding to be applicable.*

4 Analysis of Polarization

How do we prove that this kind of polarization is even going to happen? We are going to do a random walk on the polarization splitting tree. At each stage, consider the conditional entropy of a bit conditioned on everybody above it. Eventually, we get to the conditional entropy $H(W_j | W_{<j})$.

Label the 2^k clumps at the k^{th} step as $Z_1^k, ..., Z_{2^k}^k$. Letting our random walk be $X_0, ..., X_t$. We have $j_0, ..., j_t$ with $j_i = 2j_{i-1} - b_i$ for $b_i \in \{0,1\}$. We follow the conditional entropy $X_i := H(Z_{j_i}^i | Z_{<j_i}^i)$.

Theorem 3. *(equivalent to previous theorem)* $\forall p, c, \exists \beta < 1$ *s.t* $\forall t$

$$Pr_{X_0,...,X_t}\left[X_t \in \left(c^{-t}, 1 - c^{-t}\right)\right] \leq O(\beta^t)$$

4.1 Martingales

We have $X_t \in [0,1]$. Note that $\forall i$, that $\mathbf{E}_i\left[X_i | X_i, ..., X_{i-1} = \alpha\right] = \alpha$ by invertibility of $(U,V) \to (U+V, V)$. This kind of variable is a *Martingale*. They are well studied in computer science, but we are interested in the less common bounded Martingale. We want it to converge to 1 or to 0.

Example 4. *Suppose* $X_0, ..., X_t$ *has* $X_0 = \frac{1}{2}$, $X_1 = \frac{3}{4}$ *with probability* $\frac{1}{2}$ *and* $x_1 = \frac{1}{4}$ *with probability* $\frac{1}{2}$, *and*

$$x_i = x_{i-1} + 2^{-(i+1)}$$

with probability $\frac{1}{2}$ *and*

$$x_i = x_{i-1} - 2^{-(i+1)}$$

with probability $\frac{1}{2}$.

Remark This Martingale is smeared over the whole interval $[0, 1]$. But as time progresses, the step sizes are getting smaller and smaller. Our Martingale does not have this feature.

Example 5. *Suppose that if* $X_{i-1} \leq \frac{1}{2}$ *then* $X_i = \frac{X_{i-1}}{2}$ *with probability* $\frac{1}{2}$ *and* $X_i = \frac{3X_{i-1}}{2}$ *with probability* $\frac{1}{2}$, *and if* $X_{i-1} > \frac{1}{2}$ *then* X_i *is symmetric in* $(1 - X_i)$.

Remark The probability that this approaches 0 is very low. Our Martingale must do the opposite.

4.2 Local and Global Polarization

Note that both conditions are local, depending on just one time step, leading to the following concept.

The sequence $X_0, ..., X_t$ *polarizes locally* if

1. Variance in the middle: If x_{i-1} is close to $\frac{1}{2}$, the variance of x_i is high. Precisely, $\forall \tau, \exists \sigma$ s.t. $\forall i$, if $X_{i-1} \in (\tau, 1 - \tau)$ then $\mathrm{Var}(X_i | X_{i-1}) \geq \sigma$. This is the best we can hope for from a bounded Martingale that we want to polarize.
2. Suction at ends: (low end) $\exists \Theta > 0, \forall c < \infty, \exists \tau > 0$ s.t. $\forall i$, if $X_{i-1} \leq \tau$ then $\Pr\left[X_i < \frac{X_{i-1}}{c}\right] \geq \Theta$.

(In our examples Θ above will be $1/2$ since given X_{i-1}, X_i takes one of two values each with probability $1/2$.)

Theorem 6. *(Theorem 1) Our Martingale polarizes locally.*

Theorem 7. *(Theorem 2) Local Polarization implies Strong Polarization which is what we required above (with the* β^{-t} *bound).*

Proof. (Sketch)

1. Let $\phi_t = \min\{\sqrt{X_t}, \sqrt{1 - X_t}\}$ be a potential indicating how polarized X_t is already. Then $\exists \beta' < 1$ such that

   $$\mathbf{E}\left[\phi_{t+1} | \phi_t\right] \leq \beta' \phi_t$$

   which means that

   $$\mathbf{E}[\phi_t] \leq (\beta')^t$$

   and by Markov's inequality we have

   $$\Pr\left[\phi_t > (\beta')^{\frac{t}{2}}\right] \leq (\beta')^{\frac{t}{2}}$$

   which implies

   $$\Pr\left[X_t > (\beta')^t \text{ and } X_t < 1 - (\beta')^t\right] \leq (\beta')^{\frac{t}{2}}$$

2. In t more steps, we can push X_t as close to the endpoints as we really want.

$$\Pr\left[X_{t'+1} < \frac{X_{t'}}{100}\right] \geq \frac{1}{2}, \forall t' \in \{t, ..., 2t\}$$

$$\Pr\left[X'_t < X_{t'+1} < 2X_{t'}\right] \leq \frac{1}{2}$$

Note that $\log X_{t'}$ drops dramatically with probability $\frac{1}{2}$ and increases by 1 with probability $\frac{1}{2}$, so in expectation (using Chebushev's bounds) we can make it shrink as fast as we want.

□

Proposition 8. *(Doob's Martingale Inequality)*

$$Pr_{x_0,...,x_t}\left[\exists t \ s.t \ x_t \geq kx_0\right] \leq \frac{1}{k}$$

Exercise 9. *Prove Doob's Martingale Inequality.*

Next time we will show why our Martingale polarizes locally, and that will be the end of polarization and channel coding.