CS 229r Essential Coding Theory April 6, 2020

Lecture 18: Polar Codes Part II

Instructor: Madhu Sudan *Scribe: Beatrice Nash*

1 Agenda

Analysis of polarization:

- martingales and Arikan martingales
- local polarization of Arikan martingales
- strong vs. local polarization

2 Properties of polarization

Last class we defined a *polarizing map*,

$$P_u(u, v) \triangleq (P_{u/2}(u + v), P_{u/2}(v))$$

shown diagrammatically in Figure 1. For input (u, v), the first column $Z^{(0)} \to Z^{(1)}$ maps $(u, v) \to (u + v, v)$. This transformation is recursively applied until the input can no longer be divided in this way. Now if we define X_j to be the entropy of the jth column given all the previous columns,

$$X_j \triangleq H(Z_i^{(j)} | Z_{<i}^{(j)}),$$

where $i \sim \text{unif}([n])$ is the row index, chosen uniformly from $[n]$. Our goal is to follow this value of X_j as we move along through the polarizer. Our criteria for polarization is that X_j must tend either very close to 0 or to 1, and increasingly so with larger j. That is,

$$\Pr_{i \in [n]} \left[X_j \in \left[\frac{1}{n^{100}}, 1 - \frac{1}{n^{100}} \right] \right] \leq \frac{1}{n^{0.001}}.$$

In this lecture, we aim to analyze what distributions $\{X_j\}_{j \in [N]}$ satisfy this property of polarization. To do so, we first show that $\{X_j\}_{j \in [N]}$ is a martingale, defined as follows:

Definition 1. $\{X_j\}_{j \in [N]}$ is a *martingale* iff for all j, $\mathbb{E}[X_j | X_0 = a_0, ..., X_{j-1} = a_{j-1}] = a_{j-1}$.

That is, conditioned on past outcomes, future expectation does not change. Later we will show that if $\{X_j\}_{j \in [N]}$ has this property, it will demonstrate *local polarization* which in turn implies *strong polarization*, the definitions of both to follow. Now, we show how to demonstrate whether $\{X_j\}_{j \in [N]}$ is in fact a martingale.

2.1 Arikan martingales

The analysis of the connection between the polarizing behavior of the entropy history and martingales is first discussed in [Ari08].

Claim: $\{X_j\}_{j \in [N]}$ is a martingale.

Proof:

In order to show that $\{X_j\}_{j \in [N]}\}$ is a martingale, we must show that the expected entropy at any step of the polarization is unchanged. Consider a single step j of the polarizer, show in Figure 2, which polarizes

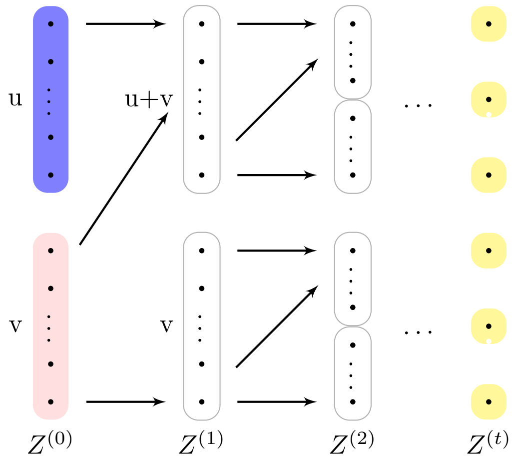

Figure 1: Diagram of polarizing channel.

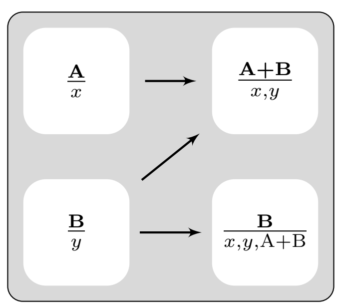

Figure 2: Polarization at time j. $\mathbf{A}$ = r.v. representing single bit, while $\mathbf{B}$ represents a collection of bits. Note that this is equivalent to the left side conditioned on x, y jointly, since $\mathbf{A}$ is independent of y and $\mathbf{B}$ is independent of x.

$(\mathbf{A},\mathbf{B})$ given (x,y). $\mathbf{A},\mathbf{B}$ are random variables representing a single bit, while x,y are collection of bits on which they are dependent. $\mathbf{A},x$ and $\mathbf{B},y$ are independent and thus indistinguishable, so $H(\mathbf{A}|x) = H(\mathbf{B}|y)$

Since the overall entropy is unchanging,

$$\begin{aligned} H(\mathbf{A}|x) + H(\mathbf{B}|y) &= H(\mathbf{A}|x,y) + H(\mathbf{B}|x,y) \\ &= H(\mathbf{A}{+}\mathbf{B}|x,y) + H(\mathbf{B}|x,y,\mathbf{A}{+}\mathbf{B}) \\ \implies \mathbb{E}[X_j|x,y] &= \mathbb{E}[X_{j+1}|x,y] \end{aligned}$$

Since its expected value remains equal to its prior value given the entropy history, X_j is a martingale. Next, we discuss how this helps us with polarization.

2.2 Martingales and polarization

As a reminder, for strong polarization, we want the following to be true:

$$\Pr[x_t \in \left(2^{-100t}, 1-2^{-100t}\right)] \leq 2^{-.001t}.$$

Now, consider a single bit A, a collection of bits B, and sets of the values of previous variables, S,T, on which A,B depend, respectively. A and S are jointly distributed and so too are B,T. Pairs of variables (A,S), (B,T) are i.i.d. Here's what we know:

$$\begin{aligned} \text{if } X_j = \alpha, \implies \exists \text{ dist } (\mathrm{A},\mathrm{S}) &=_d (\mathrm{B},\mathrm{T}) \\ \text{s.t. } H(\mathrm{A}|\mathrm{S}) = H(\mathrm{B}|\mathrm{T}) &= \alpha \\ X_{j+1} &= H(\mathrm{A{+}B}|\mathrm{S,T}) \text{ w.p. } \tfrac{1}{2} \\ &= H(\mathrm{B}|\mathrm{A{+}B,S,T}) \text{ w.p. } \tfrac{1}{2}. \end{aligned}$$

However, this behavior is very much local, and it is not clear how to go from here to strong polarization. Consider the following examples:

1. $x_{t+1} = x_t + 2^{-t+2}$ w.p. $\frac{1}{2}$, $x_t - 2^{-t+2}$ w.p. $\frac{1}{2}$
2. $x_{t+1} = x_t \pm \min\{x_t, x_t/2\}$ each w.p. $\frac{1}{2}$
3. $x_{t+1} = x_t^2$ w.p. $\frac{1}{2}$, $2x_t - x_t^2$ w.p. $\frac{1}{2}$

Example 1 is a martingale, but does not polarize. Instead of going towards an increasingly polar distribution, the distribution of x_t approaches uniform as t increases.

Example 2 is also a martingale, but it does not polarize to the desired extent. At time step t, the probability that x_t is very close to the edges,

$$\Pr[x_t \notin \left(2^{-100t}, 1-2^{-100t}\right)] = 0.$$

This martingale does not have enough "suction at the edges" to satisfy the polarization requirement.

Example 3 is a martingale and it also polarizes. But how is that? First, consider the behavior a martingale should have in order to exhibit polarization.

2.3 Local to strong polarization

We identify two general barriers to polarization: 1) not enough variance as time $t \to \infty$ and 2) weak attraction to the extrema $\{0,1\}$. To formalize this notion, we define *local polarization* as occurring when x_t has the following two properties,

1. Variance in the middle:

$$\forall \tau \; \exists \sigma > 0 \text{ s.t. } \forall t : \mathrm{Var}[x_t | x_{t-1} = a_{t-1} \in (\tau, 1-\tau)] \geq \sigma^2$$

2. Strong and increasing attraction to the extrema:

$$\exists \theta > 0 \; \forall c \; \exists \tau > 0 \text{ s.t. } \Pr[x_t < x_{t-1}/c | x_{t-1} < \tau] \geq \theta.$$

The first property is enough to rule out Example 1 from the previous subsection, and the second property rules out Example 2. We now prove the following, which allows us to conclude that Example 3 is strongly polarizing.

Theorem 2. *If a martingale $X_0, ..., X_t$ is locally polarizing, then $X_0, ..., X_t$ is also strongly polarizing.*

Here we sketch the proof of this result; for a more detailed treatment, see the original paper, [BGN+18]. Say we have a locally polarizing martingale, $x_0, ..., x_t$. Define

$$\phi_t \triangleq \min\{\sqrt{x_t}, \sqrt{1-x_t}\}$$

Then

$$\exists \beta < 1 \text{ s.t. } \forall t, x_{t-1} \;\; \mathbb{E}[\phi_t | x_{t-1}] \leq \beta \cdot \phi_{t-1}$$

must hold true. This implies the following,

$$\begin{aligned} \mathbb{E}[\phi_t] &\leq \beta^t \cdot x_0 \\ \implies \Pr[\phi_t > \beta^{t/2} x_0] &< \beta^{t/2} \\ \implies \Pr[H(w_i | w_{<i}) \in \left(2^{-0.001t}, 1 - 2^{-0.001t}\right)] &\leq 2^{-0.001t} \end{aligned}$$

Where the second inequality follows from Markov's inequality. This result is almost what we are looking for in that the probability that x_t does not polarize is very small, but we want x_t to polarize to a greater extent. To do this, we consider a two-step approach.

- first $t/2$ steps:

$$\Pr[X_{t/2} \in (\beta^{t/2}, 1 - \beta^{t/2})] < \beta^{t/2}$$

- second $t/2$ steps:

  1. $\Pr[\text{go above } \tau_0 | \text{start} < \beta^{t/2}] < \frac{\beta^{t/2}}{\tau_0}$
  2. if always below τ_0, then: double with probability $< \frac{1}{2}$, or reduce by a factor of c with probability $> \frac{1}{2}$.

The behavior for the first $t/2$ steps can be shown using Markov's inequality, as done previously. The probability that x_t exceeds τ_0 given that $x_{t/2} < \beta^{t/2}$ is given by the following,

Theorem 3 (Doob's inequality [Dur11])**.** *For a martingale $X_0, ..., X_t$ such that $X_i > 0 \; \forall i$, then for all X_0, τ, t, $Pr[\sup_{i \in [t]} X_i > \tau] \leq \mathbb{E}[X_0]/\tau$.*

Doob's inequality is the martingale analogy of Markov's inequality. The second possibility for the final $t/2$ steps is that x_t falls below τ_0, in which case x_t either doubles with probability $< 1/2$ or else is divided by c. This behavior comes from by the suction principle of local polarization: $\log x_t$ has drift towards the lower extremum.

$$\mathbb{E}[\log x_t] \le -t\theta \log c$$
$$\Pr[\log x_t \ge -\frac{t\theta \log c}{2}] \le \exp(-t)$$
$$\implies \Pr[x_t \ge \exp(-t\theta \log c/2)] \le \exp(-t)$$

The case for which $x_{t/2} > 1 - \beta^{t/2}$ is symmetric, except $\log x_{t/2}$ is pulled towards the upper extremum. Therefore with probability $1 - \exp(\Omega(-t))$, x_t is exponentially close to one of two extrema. Conversely, the probability that x_t has not been successfully polarized is exponentially small. The exact derivation of Theorem 2 is presented in detail in [BGN$^+$18].

2.4 Arikan martingales and local polarization

Finally, we can show that Arikan martingales imply local polarization, and from Theorem 2, therefore also imply strong polarization [BGN$^+$18]. We will only sketch the proof at a very high level here, as done in class; for a full treatment, see [BGN$^+$18].

First, the variance of x_t in the middle - that is, in the region $(\tau, 1-\tau)$ - will be approximately large:

$$(H(p), H(p)) \to \left(H(2p - p^2), 2H(p) - H(2p - p^2)\right),$$

where the distance between $p, 2p - p^2$ will be largest when p in the middle region.

Now, consider the suction in the lower region,

$$\begin{aligned}
H(p) &\approx p \log \frac{1}{p} \\
2H(p) - H(2p) &\approx 2p \log \frac{1}{p} - 2p \log \frac{1}{2p} \\
&= 2p \\
&= \frac{H(p)}{\log 1/p} \\
\implies H(2p - p^2) &\approx \frac{H(p)}{\log H(p)} \\
\implies x_{t+1} &\le \frac{x_t}{\log x_t}, \text{ with probability } \tfrac{1}{2}.
\end{aligned}$$

And in the upper region,

$$\begin{aligned}
p &= \frac{1}{2} - \varepsilon \\
H(\frac{1}{2} - \varepsilon) &= 1 - \Theta(\varepsilon^2) \\
2p - p^2 &= \frac{1}{2} - \Theta(\varepsilon^2) \\
H(2p - p^2) &= 1 - \Theta(\varepsilon^4)
\end{aligned}$$

So we have $x_t = (1 + \Theta(\varepsilon))x_{t-1}$ given $x_{t-1} = H(p)$ with probability $\frac{1}{2}$.

3 Summary

To summarize what we've done...

- Arikan martingale $\implies$ locally polarizing
- Local polarization $\implies$ strong polarization
- Strong polarization $\implies$ exponential gap

$$\Pr[H(w_i|w_{<i}) \in (n^{-100}, 1 - n^{-100})] \leq n^{-0.001}$$

- Therefore, polar codes achieve a polynomial gap with respect to capacity.

References

[Ari08] Erdal Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *CoRR*, abs/0807.3917, 2008.

[BGN+18] Jaroslaw Blasiok, Venkatesan Guruswami, Preetum Nakkiran, Atri Rudra, and Madhu Sudan. General strong polarization. *CoRR*, abs/1802.02718, 2018.

[Dur11] Rick Durrett. *Probability: Theory and examples.* 2011.