CS 229r Essential Coding Theory, Lecture 16 Mar 24, 2017

Lecture 16 Polar Codes II

Instructor: Madhu Sudan *Scribes: Sai Qian Zhang*

1 Administrivia

No Office hour today

2 Agenda

Today: Polar codes II

- Erratum
- Review
- Decoding
- Polarization Speed

3 Recap

In the last lecture, we talked about the polar code, which has a linear compression scheme. We further prove the following theorem:

Claim 1. *Suppose* $f_H : \{0,1\}^n \longrightarrow \{0,1\}^m$ *and* $f_H^{-1} : \{0,1\}^m \longrightarrow \{0,1\}^n$ *are such that*

- f_H *is linear, i.e.* $f_H(x) = xH$
- *For* $x \sim Bern(p)^n$, $f_H^{-1}(f_H(x)) = x$

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

Remember, given $x \sim Bern(p)^n$, $Y = f(x) = xP$,P is invertible, we want make $Y = \{Y_A, Y_B\}$ such that Y_A contains the bits which have high entropy, and Y_B contains the bits which have low entropy, which means that we can take a subset of the output Y_A, and infer the rest (Y_B) from Y_A. If this is true, we can build matrix H by permute the columns of P such that xH corresponds to the bits in Y_A. And we discussed the procedures for polarization.

This lecture, we finish the proof left from the last lecture and analyze the speed of polarization.

4 Erratum from last lecture

In the last lecture, we talked about the procedure of polarization, which is shown in Figure 1. However, the correct procedure of polarization is shown in Figure 2. Consider the one step of polarization shown in Figure 3. In the intermediate stage of calculation, we have variable A, conditioned on some C, where C is a linear combination of the input. And similarly, we have variable B, conditioned on some D, which is another linear combination of the input. Since entropy is preserved, we have the following equality:

$$H(A \oplus B|C, D) + H(B|A \oplus B, C, D) = H(A|C) + H(B|D) \tag{1}$$

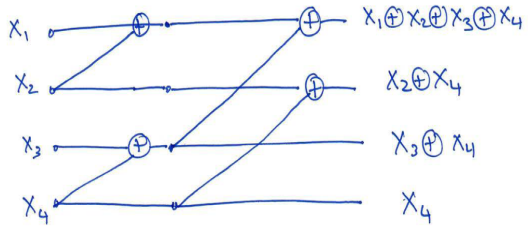

Figure 1: Wrong Polarization procedure

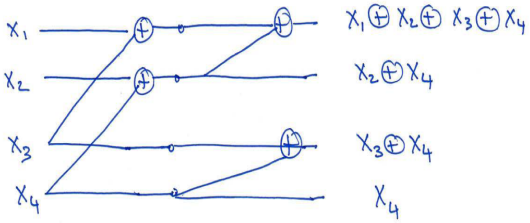

Figure 2: Correct Polarization procedure

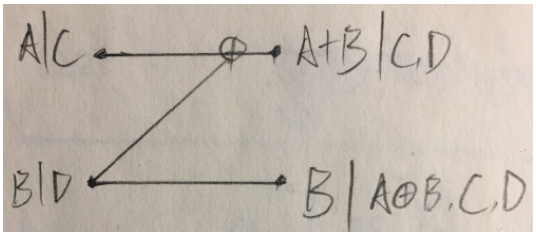

Figure 3: Example

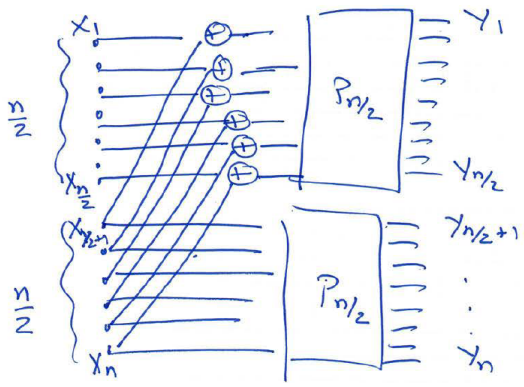

Figure 4: General Polarization procedure

If we use the polarization procedure shown in Figure 2, we will have the $D \perp (A, C)$ and $C \perp (B, D)$, hence we have $H(A|C) = H(A|C, D)$, $H(B|D) = H(B|C, D)$, and $H(A \oplus B|C, D) + H(B|A \oplus B, C, D) = H(A|C, D) + H(B|C, D)$, which makes the chain rule hold.

In general, we have the polarization procedure shown in Figure 4.

5 Decoding Algorithm

The decoding algorithm is described below:

Algorithm 1 Decompression Algorithm

Input: $y_{i \in A}$, where A is the set such that $\forall i \notin A, H(Y_i|Y_1, ..., Y_{i-1}) \leq \sigma$ (σ is small), matrix P, $x_1, ..., x_n \sim Bern(p)^n$

Output: $\hat{y}_{i \notin A}$, which is the most likely y_i for $i \notin A$

```
for i=1,...,n do
    if i ∈ A then
        ŷ_i = y_i
    else
        Compute α_i = P(Y_i = 1|y_0, ..., y_{i-1})
        if α_i > 0.5 then
            ŷ_i = 1
        else
            ŷ_i = 0
        end
    end
end
```

5.1 Accuracy of decoding

Given that $H(Y_i|Y_1, ..., Yi - 1) \leq \sigma$, we have $E_{Y_1,...,Yi-1}(H(Y_i|Y_1, ..., Y_{i-1})) \leq \sigma$, by Markov inequality, we have $P(H(Y_i|y_1, ..., y_{i-1}) > \sqrt{\sigma}) < \sqrt{\sigma} \implies P(Y_i = mode(Y_i)|y_1, ..., y_{i-1}) > 1 - \sqrt{\sigma} \implies P(Y_i \neq \hat{y}_i|y_1, ..., y_{i-1}) < \sqrt{\sigma}$. Therefore the union bound of error is less or equal to $2N\sqrt{\sigma}$, where N is total number of input.

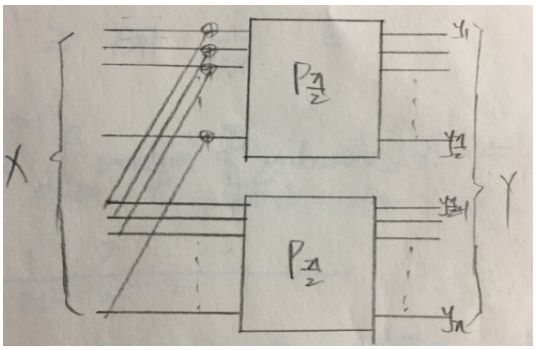

Figure 5: Polarization

5.2 How to efficiently compute $P(Y_i = 1|y_0, ..., y_{i-1})$

we need to find an efficient way to calculate $P(Y_i = 1|y_0, ..., y_{i-1})$, which is used in the decoding algorithm above. Next we propose a recursive algorithm to calculate this probability. Consider the block shown in Figure 5, let $x_1, ..., x_n \in Bern(p_1), ..., Bern(p_n)$ be the input of this block, because this is an intermediate stage of the polarization procedure, each $P(x_i = 1)$ differs for each i. This recursion is described in Algorithm 2:

Algorithm 2 Decode$(y_A, p_1, ..., p_n)$

Input: $y_A, p_1, ..., p_n$
Output: $P(Y_i = 1|y_1, ..., y_{i-1}), for 1 \leq i \leq A$
Let $A+ = A \cap [\frac{n}{2}], A- = A - A+$
for $i = 1, ..., \frac{n}{2}$ **do**
  $q_i = p_i(1 - p_{\frac{n}{2}+i}) + p_{\frac{n}{2}+i}(1 - p_i)$
end
$(z_1, ..., z_{\frac{n}{2}})$=Decode$(y_{A+}, q_1, ..., q_{\frac{n}{2}})$
for $i = 1, ..., \frac{n}{2}$ **do**
  $r_i = P(b = 1|a \sim Bern(p_i), b \sim Bern(p_{i+\frac{n}{2}}), a \oplus b = z_i)$
end
$(y_1, ..., y_{\frac{n}{2}})$=Decode$(y_{A-}, r_1, ..., r_{\frac{n}{2}})$
$Output(z \oplus y, y)$

The complexity of this recursion is $\mathcal{O}(N \log N)$.

6 Speed of Polarization

Theorem 2 provides a description on the degree of polarization and the number of polarization steps.

Theorem 2. *$\exists$ polynomial K, s.t. $\forall \epsilon > 0$, $0 < p < \frac{1}{2}$, $N = K(\frac{1}{\epsilon}) = 2^l$, where l is the number of polarization steps. Given $x_1, ..., x_n \sim Bern(p)$ and $[y_1, ..., y_n]^T = P_N[x_1, ..., x_n]^T$, this polarization satisfies $|\{i|H(Y_i|Y_1, ..., Y_{i-1}) \leq \frac{1}{10N^2}\}| \geq (1 - H(p) - \epsilon)N$*

This theorem is proven by using the following two lemmas:

Lemma 3. *(weak, two-side polarization) $\exists$ polynomial K s.t. $\forall \epsilon, p > 0$, if $N = K(\frac{1}{\epsilon})$ and $[y_1, ..., y_n]^T = P_N[x_1, ..., x_n]^T$, $x_1, ..., x_n \sim Bern(p)$, we have $|\{i|H(Y_i|Y_1, ..., Y_{i-1}) \in (\epsilon, 1 - \epsilon)\}| \leq \epsilon N$*

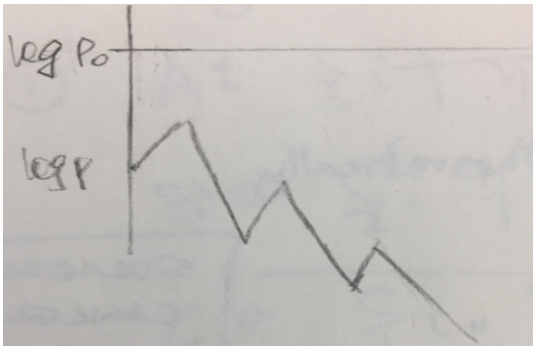

Figure 6: Random Walk

Lemma 3 claims that after l-steps of polarization, the number of entropies which are closed to each other (within the interval $(\epsilon, 1-\epsilon)$) is small ($\leq \epsilon N$).However, this bound is weak because K can be any polynomial that we can not control. When $\epsilon = \frac{1}{N^{0.01}}$, the interval becomes $(\frac{1}{N^{0.01}}, 1 - \frac{1}{N^{0.01}})$, which is very narrow.

Lemma 4. *(strong, one-side polarization)* $\exists$ *polynomial* K_1, K_2 $\forall \epsilon > 0$, *if* $p \leq K_1(\epsilon)$ *and* $N \geq K_2(\frac{1}{\epsilon})$, *then* $|\{i|H(Y_i|Y_1, ..., Y_{i-1}) \leq \frac{1}{N^3}\}| \geq (1 - H(p) - \epsilon)N$

Lemma 4 claims that the number of outputs with small entropy is very large. Lemma 4 is proven by using the following idea, let's go through one step of polarization, assume $x_0, x_1 \sim Bern(p)$, after polarization, we have $x_0 \oplus x_1 \sim Bern(p^+)$ and $x_0|x_0 \oplus x_1 \sim Bern(p^-)$, where $p^+ = 2p(1-p)$ and $p^- = h^{-1}(2h(p) - h(2p(1-p)))$, $h(p) = plog(\frac{1}{p}) + (1-p)log(\frac{1}{1-p})$. if p is very small, we have $h(p) \approx plog(\frac{1}{p}) \Longrightarrow 2h(p) - h(p^+) = h(p^-) \approx 2p \Longrightarrow p^- \approx \frac{p}{log\frac{1}{1-p}}$. Now assume $p < 2^{-10}$. after one step of polarization p^+ will be doubled and p^- will become one tenth of itself, there is a drift towards the negative direction.

This is illustrated in Figure 6, as long as $log(p) < log(p_0)$, each time $log(p)$ will either increase by 2 or decrease by 10. If $p > p_0$, p will escape from the this process and become large. However, the probability that p hits p_0 is approximately equal to $poly(p)$.

Similarly, lemma 3 can be proven by using claim 5, denote $p^+ = 2p(1-p)$ and $p^- = h^{-1}(2h(p) - h(p^+))$ and define potential function to be $\phi(p) = \sqrt{h(p)(1-h(p))}$

Claim 5. $\exists \Lambda < 1$ *s.t.* $\forall 0 < p < \frac{1}{2}$, $\frac{\phi(p^+)+\phi(p^-)}{2} \leq \Lambda(\phi(p))$.

Claim 5 indicates that the expected value of potential decreases exponentially with the number of polarization steps.