EE276: Homework #6 Solutions

Due on Friday Feb 27, 11:59pm - Gradescope entry code: E6VP4X

1. **Polar codes encoding and decoding.**
In this problem, we try to understand the polar code encoding and decoding procedure through small examples with block size $N = 4$. You will work with the circuit shown in Figure 1. You may want to take a look at these polar code decoding slides on the course website to learn more about *successive cancellation decoding*: `https://web.stanford.edu/class/ee276/files/lectures/polar_decoding.pdf`.

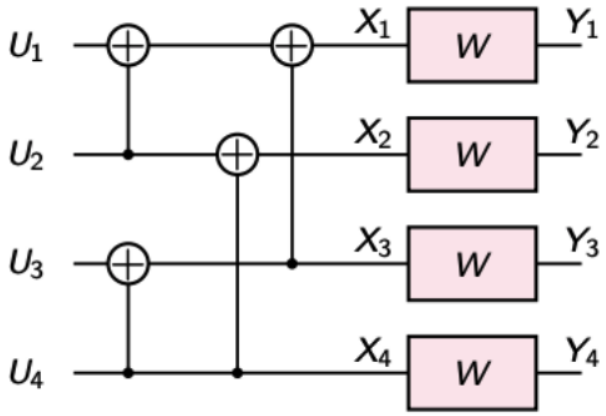

Figure 1: Polar code encoder with $N = 4$. W is a BEC, and X_i's, Y_i's and U_i's are binary.

For parts (a) to (c), assume that U_1 and U_2 are both frozen to 0, while U_3 and U_4 are the message bits.

(a) What is the rate of the code?

(b) Perform encoding for input message $(U_3, U_4) = (1, 1)$ and find the codeword (X_1, X_2, X_3, X_4).

(c) Perform successive cancellation decoding for received vector $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 1)$. Does the decoding succeed, and if yes, what is the decoded message (U_3, U_4)?

Now we try to understand how the choice of the frozen bits impacts the decoding. We also look at the suboptimality of successive cancellation decoding. For parts (d) and (e), assume that U_2 and U_3 are both frozen to 0, while U_1 and U_4 are the message bits. Recall that when a message bit is being decoded using output bits and previously decoded message bits, all other message bits are assumed to be random.

(d) Perform successive cancellation decoding for received vector $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 0)$ and verify that the decoding fails when decoding U_1.

(e) Perform optimal maximum likelihood decoding for the same received as part (d), i.e., $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 0)$. In this case, you can perform maximum likelihood decoding by

- First computing the codeword (X_1, X_2, X_3, X_4) for all 4 possible input messages (U_1, U_4).
- Then finding the input message(s) (U_1, U_4) for which you could receive $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 0)$. If more than one such message exists, declare failure.

Does the decoding succeed and, if yes, what is the decoded message (U_1, U_4)?

Solution:

(a) The rate is $1/2$ since we transmit 2 message bits (U_3, U_4) over 4 channel transmissions (X_1, X_2, X_3, X_4).

(b) $(X_1, X_2, X_3, X_4) = (0, 1, 0, 1)$

(c) Decoding succeeds and we get $(U_3, U_4) = (0, 1)$.

(d) See Figure 2: decoding fails since both outputs of the top left 2x2 block are erasures. Note that when doing successive cancellation decoding, we assume U_2 and U_3 are random while decoding U_1 even though U_2, U_3 are in fact frozen to 0.

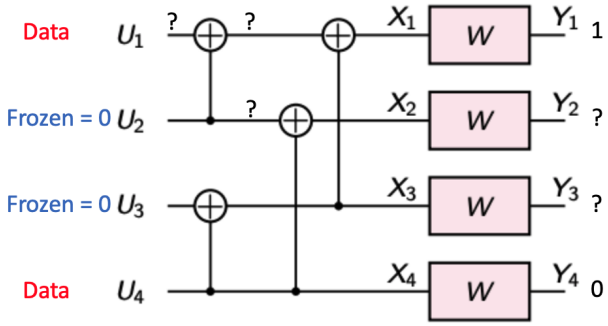

Figure 2: SC decoding for part (d).

(e)

$$(U_1, U_4) = (0, 0) : (X_1, X_2, X_3, X_4) = (0, 0, 0, 0)$$
$$(U_1, U_4) = (0, 1) : (X_1, X_2, X_3, X_4) = (1, 1, 1, 1)$$
$$(U_1, U_4) = (1, 0) : (X_1, X_2, X_3, X_4) = (1, 0, 0, 0)$$
$$(U_1, U_4) = (1, 1) : (X_1, X_2, X_3, X_4) = (0, 1, 1, 1)$$

As you can see, the only input that could produce $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 0)$ is $(U_1, U_4) = (1, 0)$ (so maximum likelihood decoding succeeds in this case).

2. **Proving polarization for the BEC.** In polar coding, we preprocess the input so that the n identical uses of a symmetric memoryless channel become n synthetic channel uses with very different capacities. We state a polarization theorem, which says that as $n \to \infty$, the fraction of almost noiseless channels approaches C and the fraction

of almost useless channels approaches $1 - C$, where C is the capacity of the original channel. In this question, we consider the binary erasure channel (BEC) with erasure probability p and prove the polarization theorem rigorously. For the BEC $\mathcal{W}$ with erasure probability p, define $\mathcal{M}(\mathcal{W}) = \sqrt{p(1-p)}$ as its *mediocrity*.

(a) When is the mediocrity of a channel 0?

(b) Consider the polarized channels $\mathcal{W}^+$ and $\mathcal{W}^-$ we have seen in the class for $m = 1$. Are they also BECs? If so, what are $\mathcal{M}(\mathcal{W}^+)$ and $\mathcal{M}(\mathcal{W}^-)$?

(c) Recall the tree of channel capacities obtained by recursively applying the polarization formula to the BECs. Suppose an ant walks on the tree of channel capacities starting at $\mathcal{W}$ and choosing $\mathcal{W}^+$ and $\mathcal{W}^-$ with equal probability $1/2$. Upon reaching each channel $\tilde{\mathcal{W}}$ (e.g., $\tilde{\mathcal{W}} = \mathcal{W}^+$), it chooses $\tilde{\mathcal{W}}^+$ ($\mathcal{W}^{++}$) and $\tilde{\mathcal{W}}^-$ ($\mathcal{W}^{+-}$) with equal probability $1/2$. Let F_m denote the distribution of the erasure probabilities for $n = 2^m$ and let $F_0 = p$ (with probability 1). What are the distributions F_1 and F_2?

(d) Let $\mathcal{M}_m$ denote the average mediocrity of the channels for the distribution F_m. For instance $\mathcal{M}_0 = \sqrt{p(1-p)}$. What is $\mathcal{M}_1$? Prove that $\mathcal{M}_1 \leq \sqrt{\frac{3}{4}}\mathcal{M}_0$.

(e) Let $\rho = \sqrt{\frac{3}{4}}$. Prove that $\mathcal{M}_m \leq \rho^m$.

(f) Let $\mathsf{mediocre}(m, \epsilon)$ denote the fraction of the $n = 2^m$ channnels with mediocrity strictly larger than $\sqrt{\epsilon(1-\epsilon)}$. Observe that these channels have erasure probability in $(\epsilon, 1-\epsilon)$. Show that $\mathsf{mediocre}(m, \epsilon) \to 0$ as $m \to \infty$.

(g) Let $\mathsf{good}(m, \epsilon)$ and $\mathsf{bad}(m, \epsilon)$ denote the fraction of the channels with erasure probability strictly smaller than ϵ (i.e., good channels) and strictly larger than $1 - \epsilon$ (i.e., bad channels) respectively. Show that

$$p \geq \mathsf{bad}(m, \epsilon)(1 - \epsilon).$$

(Hint: recall that the expected erasure probability under F_m is the same for all m and equal to p.)

(h) Define $g(\epsilon) := \lim_{m\to\infty} \mathsf{good}(m, \epsilon)$. Argue that $\mathsf{good}(m, \epsilon) \geq \mathsf{good}(m, \delta)$ for any $\epsilon \geq \delta$. Conclude that $g(\epsilon) \geq g(\delta)$ for any $\epsilon \geq \delta$.

(i) Prove that $g(\epsilon) \geq 1 - p$. Thus, for any given $\epsilon \in (0, 1)$, the fraction $g(\epsilon)$ of good channels becomes at least $C = 1 - p$ as $m \to \infty$.

Solution:

(a) Mediocrity, $\mathcal{M}(\mathcal{W}) = \sqrt{p(1-p)} = 0$, when $p = 0$ or $p = 1$.

(b) The channels $\mathcal{W}^+$ and $\mathcal{W}^-$ are also BEC, with erasure probabilities p^2 and $1-(1-p)(1-p) = 2p - p^2$. Thus, $\mathcal{M}(\mathcal{W}^+) = \sqrt{p^2(1-p^2)} = p\sqrt{1-p^2}$ and $\mathcal{M}(\mathcal{W}^-) = \sqrt{(2p-p^2)(1-2p+p^2)} = (1-p)\sqrt{2p-p^2}$.

(c) Using the recursive formula $C^+ = 2C - C^2$ and $C^- = C^2$, we can write

$$F_1 = \begin{cases} 2p - p^2 & \text{with probability } \frac{1}{2} \\ p^2 & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$F_2 = \begin{cases} (2p - p^2)(2 - 2p + p^2) & \text{with probability } \frac{1}{4} \\ (2p - p^2)^2 & \text{with probability } \frac{1}{4} \\ 2p^2 - p^4 & \text{with probability } \frac{1}{4} \\ p^4 & \text{with probability } \frac{1}{4} \end{cases}$$

(d) $\mathcal{M}_1 = \frac{1}{2}\mathcal{M}(\mathcal{W}^+) + \frac{1}{2}\mathcal{M}(\mathcal{W}^-) = \frac{1}{2}(p\sqrt{1-p^2} + (1-p)\sqrt{2p-p^2})$, where we have calculated $\mathcal{M}(\mathcal{W}^+)$ and $\mathcal{M}(\mathcal{W}^-)$ in part (b). To prove $M_1 \leq \sqrt{\frac{3}{4}}\mathcal{M}_0$, we use Jensen's inequality. Since $\sqrt{\cdot}$ is concave, it holds that

$$\begin{aligned} &\frac{1}{2}(p\sqrt{1-p^2} + (1-p)\sqrt{2p-p^2}) \\ &\leq \frac{1}{2}\sqrt{p(1-p^2) + (1-p)(2p-p^2)} \\ &= \frac{1}{2}\sqrt{p(1-p)}\sqrt{1+p+2-p} \\ &= \sqrt{\frac{3}{4}p(1-p)}, \end{aligned}$$

(e) We observe that $\mathcal{M}_{m+1} \leq \rho\mathcal{M}_m$ for all m by applying the observation in part (c) to all new branches. Thus, $\mathcal{M}_m \leq \rho^m\mathcal{M}_0 \leq \rho^m$ since $\mathcal{M}_0 \leq 1$ for all $p \in (0,1)$.

(f) Note that by definition, $\mathsf{mediocre}(m,\epsilon)\sqrt{\epsilon(1-\epsilon)} \leq \mathcal{M}_m$; as $\mathcal{M}_m$ also includes the contributions of the channels with mediocrity in $(0, \sqrt{\epsilon(1-\epsilon)})$ and the mediocrity of the channels counted for $\mathsf{mediocre}(m,\epsilon)$ can be larger than $\sqrt{\epsilon(1-\epsilon)}$. Since $\mathcal{M}_m \leq \rho^m$ by part (e), $\mathcal{M}_m \geq 0$ and $\rho \in (0,1)$, as $m \to \infty$, $\mathsf{mediocre}(m,\epsilon) \leq \mathcal{M}_m \to 0$. This implies $\mathsf{mediocre}(m,\epsilon) \to 0$ as $m \to \infty$.

(g) As the average erasure probability of F_m is p for all m, it holds that $p \geq \mathsf{bad}(m,\epsilon)(1-\epsilon)$, as $\mathcal{M}_m$ also includes the contributions of the good and mediocre channels, i.e., $\mathsf{good}(m,\epsilon)$ and $\mathsf{mediocre}(m,\epsilon)$, and the erasure probability of the channels counted for $\mathsf{bad}(m,\epsilon)$ can be larger than $1-\epsilon$. Hence, $p \geq \mathsf{bad}(m,\epsilon)(1-\epsilon)$.

(h) Since $\mathsf{good}(m,\epsilon)$ is defined as the fraction of the channels with erasure probability $< \epsilon$, and any channel with erasure probability $< \delta$ is also a channel with erasure probability $< \epsilon$, $\mathsf{good}(m,\epsilon) \geq \mathsf{good}(m,\delta)$ for any $\epsilon \geq \delta$ and all m. Hence $g(\epsilon) = \lim_{m\to\infty}\mathsf{good}(m,\epsilon) \geq lim_{m\to\infty}\mathsf{good}(m,\delta)g(\delta)$ for any $\epsilon \geq \delta$.

(i) By part (g), $\mathsf{good}(m,\epsilon) = 1 - \mathsf{mediocre}(m,\epsilon) - \mathsf{bad}(m,\epsilon) \geq 1 - \mathsf{mediocre}(m,\epsilon) - \frac{p}{1-\epsilon} = \frac{1-\epsilon-p}{1-\epsilon} + \mathsf{mediocre}(m,\epsilon)$. Taking $m \to \infty$, we observe that $g(\epsilon) = \lim_{m\to\infty}\mathsf{good}(m,\epsilon) \geq$

$\frac{1-\epsilon-p}{1-\epsilon} - \lim_{m\to\infty} \mathsf{mediocre}(m,\epsilon) = \frac{1-\epsilon-p}{1-\epsilon}$ as $\lim_{m\to\infty} \mathsf{mediocre}(m,\epsilon) = 0$ by part (f). Finally, since $g(\epsilon) \geq g(\delta)$ for any $\epsilon \geq \delta$ by part (h), we can conclude that $g(\epsilon) \geq \lim_{\delta\to 0} g(\delta) = \lim_{\delta\to 0} \frac{1-\delta-p}{1-\delta} = 1-p$.

3. **Prelude to Sanov.** In this problem we compute a standard bound on the tail probability of a sum of i.i.d. Bernoulli random variables, which you may recall from a probability class. With some effort, we see that the tail probabilities decay exponentially. In the lecture you will soon see there is a close connection between the rate of this exponential decay and relative entropy between certain types and the true distribution of the random variable.

   (a) Let $X_1, X_2, \ldots$ be i.i.d. Ber($\frac{1}{2}$). Show that

   $$\Pr\left\{\sum_{i=1}^{n} X_i \geq \frac{2}{3}n\right\} \leq \exp\left(-\frac{2}{3}tn\right)\left(1+\frac{1}{2}(e^t-1)\right)^n$$

   for all $t \geq 0$.

   Hint: Recall the Chernoff bound

   $$\Pr\{Y \geq a\} = \Pr\{\exp(tY) \geq \exp(at)\} \leq e^{-at}\mathbb{E}[\exp(tY)].$$

   (b) What tightest (smallest) bound in the form of $\Pr\left\{\sum_{i=1}^{n} X_i \geq \frac{2}{3}n\right\} \leq e^{-cn}$ for some constant $c > 0$ you can derive from part (a)?

   (c) Which type classes are included in the event $\{\sum_{i=1}^{n} X_i \geq \frac{2}{3}n\}$? Which are most likely? Assume n is divisible by 3.

   (d) Compute $D(P\|\text{Ber}(\frac{1}{2}))$ for each of the type(s) P of the most likely type class(es) from part (c). Compare with the result from part (b).

   (e) Is it possible to improve on the constant c in part (b) using some other method? Note we require the inequality to hold for all n.

Solution:

(a) We have

$$\mathbb{E}[\exp(tX_i)] = \frac{1}{2} + \frac{1}{2}e^t$$

and thus by independence

$$\mathbb{E}[\exp(t\sum_i X_i)] = \left(\frac{1}{2} + \frac{1}{2}e^t\right)^n.$$

Substituting $a = \frac{2}{3}n$ and $Y = \sum_i X_i$ yields the desired result.

(b) Taking the log and dividing out n, the optimal bound has

$$c = \max_{t \geq 0} \frac{2}{3}t - \ln\left(1 + \frac{1}{2}(e^t - 1)\right).$$

Differentiating, we find it is necessary that

$$\begin{aligned} &\frac{2}{3} - \frac{\frac{1}{2}e^t}{1 + \frac{1}{2}(e^t - 1)} = 0 \\ \implies\ &t = \ln 2 = 0.6931\ldots \end{aligned}$$

We can check by second derivative that this is a maximum. We conclude that

$$c = \frac{5}{3}\ln 2 - \ln 3 = 0.056633\ldots$$

is the optimal rate.

(c) The type classes are $T(P)$ for distributions P such that $P(1) \geq \frac{2}{3}$. The type class for the type for $P = \mathrm{Ber}(\frac{2}{3})$ is most likely. One way to see this is to note that the binomial distribution has pmf decreasing on integers above its mean.

(d) Computing, we get $D(\mathrm{Ber}(\frac{2}{3})\|\mathrm{Ber}(\frac{1}{2})) = \frac{5}{3}\ln 2 - \ln 3$. Note this is the same as c in part (b).

(e) We have that

$$\Pr\left\{\sum_{i=1}^{n} X_i \geq \frac{2}{3}n\right\} \geq \Pr\left\{X^n \in T(\mathrm{Ber}(\frac{2}{3}))\right\} \geq \frac{1}{n+1}\exp(-cn) = \exp(-cn - \log(n+1))$$

Since log grows sublinearly, this will eventually be larger than $\exp(-c'n)$ for all $c' > c$, so improving the constant is not possible.