Usual Instructions — See Pset 2.

Problems

1. **Channel capacity.** Suppose we are given finite alphabets $\Omega_X, \Omega_Y$, and a channel $P(Y|X)$. Recall that the channel capacity is defined as $\max_{P_X} I(X;Y)$, where the maximum is taken over all distributions $P_X$ on $\Omega_X$. Construct a channel $P(Y|X)$ such that the uniform distribution on $\Omega_X$ does not achieve the channel capacity.

2. **Random linear codes.** Recall that a linear code of block length $n$ and dimension $k$ (where $k \leq n$) is specified by a matrix $G \in \mathbb{F}_2^{n \times k}$ with rank $k$. The encoding of a codeword $a \in \mathbb{F}_2^k$ is given by $E(a) = Ga \in \mathbb{F}_2^n$.

   (a) (Not to be turned in.) Show that there exists a matrix $H \in \mathbb{F}_2^{(n-k) \times n}$ of rank $n-k$ such that $H \cdot G = 0$. Conclude that the code $C := \{Ga : a \in \mathbb{F}_2^k\}$ is equivalently given by $\{x \in \mathbb{F}_2^n : Hx = 0\}$. The matrix $H$ is known as the parity check matrix corresponding to the code $C$.

   (b) (Not to be turned in.) Given $G$, find an algorithm (as efficient as you can) that computes any parity check matrix $H$ as described in part (a). What is the runtime?

   (c) Apply your solution from parts (a) and (b) to give the matrix $H$ corresponding to

   $$G^T = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

   (d) Show that a random linear encoder achieves capacity over the binary symmetric channel with parameter $p$. Specifically show that for every $\varepsilon > 0$ if $n$ is large enough, then a random encoder given by a uniformly random matrix $G \in \mathbb{F}_2^{n \times k}$ has a decoder that correctly decodes the code (with error probability going to zero exponentially fast) for some $k \geq (1 - h(p) - \varepsilon)n$, over the worst case choice of the message!

3. **Polarization Preprocessing:** Recall that the algorithm for compression based on polarization started with $Z \in \text{Bern}(p)^n$ and compressed it using the recursive expression $P_n(U,V) = (P_{n/2}(U + V), P_{n/2}(V))$, where $Z = (U,V)$ for $U,V \in \mathbb{F}_2^{n/2}$. The final compressed output is $P_n(Z)_S$ for some set $S \subseteq [n]$, where $S = \{j \in [n] | H(P_n(Z)_j | P_n(Z)_{<j}) \geq \tau\}$. Give a poly$(n/\tau)$ time randomized algorithm to compute the set $S$. (The input to your algorithm is $n, p, \tau$ and the output should be a set $S$ as above.)
Hint: Roughly our task is to sample, given an index \( j \), a random sequence of the form \( w_{<j} \) according to the distribution of \( P_n(Z)_{<j} \) and then to compute the distribution of \( P_n(Z)_{<j} = w_{<j} \). Show how such a sampling + distribution estimation algorithm follows from the Polar decoder.

4. **Local Polarization**: The goal of this question is to show that if \((U,A)\) and \((V,B)\) are identically and independently distributed random variables with \( U,V \in \mathbb{F}_2 \) then for every \( \tau > 0 \) there exists \( \sigma > 0 \) such that if \( H(U|A) = H(V|B) \in (\tau, 1 - \tau) \) then \( H(U + V|A, B) \geq H(U|A) + \sigma \).

(a) **(Not to be turned in.)** In what step in the analysis of Polar coding is such a statement useful? What choices of \( U, V, A \) and \( B \) do we use?

(b) Prove that if \( X, Y \) are independent random variables with \( X \sim \text{Bern}(p_1) \) and \( Y \sim \text{Bern}(p_2) \) then \( H(X + Y) \geq \max\{H(X), H(Y)\} \). Furthermore if \( H(X), H(Y) \in (\tau/2, 1 - \tau/2) \) prove that there is some \( \varepsilon > 0 \) (depending only on \( \tau \)) such that \( H(X + Y) \geq \max\{H(X), H(Y)\} + \varepsilon \).

(c) **(Not to be turned in.)** Use case analysis on \( \text{Pr}_a[H(U|A = a) \in (\tau/2, 1 - \tau/2)] \) to show that \( H(U + V|A, B) \geq H(U|A) + \sigma \). (Roughly, if \( H(U|A = a) \) is in the interval with noticeable probability then Part (b) yields the entropy increase. And if this probability is tiny then with noticeable probability \( H(U|a) \) is low, \( H(V|b) \) is high and so \( H(U + V|a, b) \) is high compared to \( H(U|a) \) and this leads to an entropy increase in \( H(U + V|A, B) \) compared to \( H(U|A) \).

5. **Generalizations of local polarization.** Our construction of polar codes was based off of the primitive polarizing matrix \( P_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \) corresponding to the transformation \((U, V) \mapsto (U + V, V)\). We then defined a basic polarizing matrix \( P_n \), for \( n = 2^t \), recursively, via \( P_n((U, V)) = P_{n/2}(U + V) + P_{n/2}(V) \) for \( U, V \in \mathbb{F}_2^{n/2} \). This construction can be generalized to the setting where we replace \( P_2 \) by an arbitrary square matrix. For instance, consider the basic polarizing matrix

\[
P_4 := \begin{bmatrix}
1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix},
\]

which leads to the recursion, for \( n = 4^t \), \( P_n((U, V, W, X)) = (P_{n/4}(U + V + W), P_{n/4}(W + X), P_{n/4}(V), P_{n/4}(U)) \), where \( U, V, W, X \in \mathbb{F}_2^{n/4} \).

Will this lead to a Polar code that gets \( \varepsilon \) close to capacity at polynomial block lengths? If so, what aspects of the analysis need to change to show this? If not, what aspects of the analysis break down?

---

\(^1\)The matrix \( P_n \) is actually the log \( n \)th tensor power of \( P_2 \).