Lecture 16: Shannon's Coding Theorem

Binary Symmetric Channel

- Recall that a $B(1, p)$ is a distribution over the sample space $\{0, 1\}$ such that $B(1, p)$ outputs 1 with probability p

Definition (Binary Symmetric Channel)

For $\varepsilon \in (0, 1/2)$, an ε-binary symmetric channel, represented as ε-BSC, is a noisy channel that takes as input a bit b and outputs a bit $\widetilde{b} := b + B(1, \varepsilon)$.

- Intuitively, the channel flips each input bit independently with probability ε
- If an n-bit string c is passed through the channel, then the output string is expected to have $n\varepsilon$ errors
- By concentration inequalities, if an n-bit string c is passed through the channel, then the output string has at most $(\varepsilon + \delta)n$ errors with probability $\leqslant \exp(-2\delta^2 n/\varepsilon)$.

Original Motivation for Error-correcting Codes

- Intuitively: Our goal is to "reliably transmit" messages over ε-BSC with minimum "per-bit overhead"
- Formalization:
  - A sender wants to reliably send a message $m \in \{0,1\}^k$ to a receiver
  - The sender encodes m into a codeword $c \in \{0,1\}^n$ and sends c over the ε-BSC
  - The receiver obtains the erroneous string $\widetilde{c}$, finds the closest codeword c' to $\widetilde{c}$, and outputs the message m' corresponding to c'
  - We want $\mathbb{P}\left[m = m'\right] \geqslant 1 - 2^{-\lambda n}$ while minimizing n/k
- Intuitively, the overhead of reliably transmitting a k-bit messages is $(n-k)$ bits. So, we the "per-bit overhead" is $(n-k)/k$. Or, equivalently, we minimize n/k

(A very special form of) Shannon's Coding Theorem

Definition (Rate of a Code)

An $[n, k]_2$ code has rate k/n.

- For every channel, there exists a number called *its capacity* $C \in (0, 1)$ that measures the reliability of the channel
- For ε-BSC, we have $C = 1 - h_2(\varepsilon)$

Theorem (Shannon's Theorem)

For every channel and threshold τ, there exists a code with rate $R \geqslant C - \tau$ that reliably transmits over this channel, where C is the capacity of the channel. Such a code is referred to as capacity achieving*.*

- The capacity achieving code for a channel need not be linear
- The capacity achieving code for ε-BSC *happens* to be linear
- In general, the best rate of linear codes to reliably transmit over a channel can be significantly smaller than its capacity

What we will prove

We will show the following.

- For all ε, we can construct a random binary linear code (with probability $1 - 2^{-\alpha n}$) that has rate $R = 1 - h_2(\varepsilon) - \tau$ and reliably transmits messages over ε-BSC correctly with probability $1 - 2^{-\lambda n}$

You have already proven this in your homework problem! We will provide an alternate proof.

Randomized Construction

For an ε-BSC, we choose the following parameters.

- Let δ be such that $1 - \exp(-2\delta^2 n/\varepsilon) \geqslant 1 - 2^{-\lambda n}$
- Let $d = 2(\varepsilon + \delta)n + 1$
- τ is a parameter that is chosen based on d and α that will be explained later
- We choose $k/n = R = 1 - h_2(\varepsilon) - \tau$

Randomized Construction.

- Generate a random $P \in \{0,1\}^{k \times (n-k)}$ matrix and output the code generated by $G = \left[T_{k \times k} \| P\right]$

Proof

- Note that the code is always an $[n, k]_2$ code with rate $R = 1 - h_2(\varepsilon) - \tau$
- Note that the channel introduces at most $(\varepsilon + \delta)n$ errors with probability $\geqslant 1 - 2^{-\lambda n}$
- Conditioned on the introduction of at most $(\varepsilon + \delta)n$ errors by the channel, we can always correctly recover the transmitted message with probability 1, if the distance of the code is $d \geqslant 2(\varepsilon + \delta)n + 1$
- So, all that remains to argue is the following. The code generated by G has distance $\geqslant 2(\varepsilon + \delta)n + 1$ with probability $1 - 2^{-\alpha n}$

Proof

- Let $\mathcal{C}$ be the code generated by the matrix G
- Let $H = \left[-P^\top \| I_{n-k \times n-k} \right]$ be the generator matrix of the dual code of $\mathcal{C}$
- Suppose there exists a weight w codeword in $\mathcal{C}$. Suppose the codeword is c and it has 1 only at positions $i_1 < i_2 < \cdots < i_w$.
- This implies that the sum of the columns $\{i_1, \ldots, i_w\}$ of H is the 0-column
- The probability of these w columns adding up to the 0-column is $\leqslant 2^{-(n-k)}$

- The probability that some $\leqslant w$ columns of H add up to 0-column is at most (by union bound)

$$\sum_{i=0}^{w} \binom{n}{i} 2^{-(n-k)} = \mathrm{Vol}_2(w,n) 2^{-(n-k)} \leqslant 2^{h_2(w/n)n} \cdot 2^{-(n-k)}$$

- The probability that some $\leqslant (\varepsilon + \delta)n$ columns of H add up to 0-column is

$$\leqslant 2^{-(\,1 - R - h_2(\varepsilon+\delta)\,)n}$$

- Recall, we have set $R = 1 - h_2(\varepsilon) - \tau$ and τ is a parameter we need to choose
- Suppose we choose τ such that

$$2^{-(\,1 - R - h_2(\varepsilon+\delta)\,)n} \leqslant 2^{-\alpha n}$$

then we will done

So, we choose τ such that

$$\begin{aligned} & 1 - R - h_2(\varepsilon + \delta) \geqslant \alpha \\ \iff \quad & h_2(\varepsilon) + \tau - h_2(\varepsilon + \delta) \geqslant \alpha \\ \iff \quad & \tau \geqslant \alpha + \Big(h_2(\varepsilon + \delta) - h_2(\varepsilon) \Big) \end{aligned}$$