COMPSCI 690T Coding Theory and Applications Feb 13, 2017

Lecture 7

Instructor: Arya Mazumdar *Scribe: Names Redacted*

For a linear code, the Weight Enumerator Polynomial is defined as below.

$$A(x,y) = \sum_{w=0}^{n} A_w x^{(m-w)} y^w,$$

where A_w is the number of codewords of weight w.

1 Bhattacharyya Bound

In the Binary Symmetric Channel (BSC), each bit has a probability p to go wrong. The Bhattacharyya bound is a bound on the probability of error over BSC.

Suppose we transmit all zero codeword $\underline{0}$ and receive y, the probability that y is not correctly decoded to $\underline{0}$, P_e, is bounded as the following expression. Below, for a codeword c, let $D(c)$ be the Voronoi Region of c.

$$\begin{aligned}
P_e &= \sum_{c \in C \setminus \{\underline{0}\}} P(y \to c) \\
&= \sum_{c \in C \setminus \{\underline{0}\}} \sum_{y \in D(c)} P(y|c) \\
&\leq \sum_{c \in C \setminus \{\underline{0}\}} \sum_{y \in D(c)} \sqrt{P(y|0)P(y|c)} \\
&= \sum_{c \in C \setminus \{\underline{0}\}} \sum_{y \in D(c)} \prod_{i=1}^{n} \sqrt{P(y_i|0)P(y_i|c)} \\
&\leq \sum_{c \in C \setminus \{\underline{0}\}} \prod_{i=1}^{n} \sum_{y_i=0}^{1} \sqrt{P(y_i|0)P(y_i|c_i)} \\
&= \sum_{w=1}^{n} A_w (2\sqrt{p(1-p)})^w \\
&= A(1, 2\sqrt{p(1-p)}) - A_0 \\
&= A(1, 2\sqrt{p(1-p)}) - 1
\end{aligned}$$

Therefore,

$$P_e \leq A(1, 2\sqrt{p(1-p)}) - 1.$$

2 Shannon Bound

For a Random Linear Codes (RLC), let A_w be the random number of codewords with weight w, then the expected value of A_w is as follows.

$$\mathbb{E}[A_w] = \frac{\binom{n}{w}}{2^{n-k}}$$

According to Markov inequality, there exist a random linear code such that

$$P[A_w \leq n^2 \frac{\binom{n}{w}}{2^{n-k}}] \leq 1 - \frac{1}{n},\ \forall w.$$

To analyze the performance of a RLC on a BSC, we need to fix a decoding method. Assume that y is received. The decoding method contains 2 steps:

1. Find the nearest codeword x of y.
2. If $|d(x,y) - np| \leq \epsilon n$, output x. Otherwise, declare error.

Let the zero codeword $\bar{0}$ be sent and y be received. The error contains two events. 1. $wt(y)$ deviates from np. 2. $wt(y)$ is close to np and there exists another codeword other than 0 that is also close to np.

$$P_e \leq P[|wt(y) - np| > \epsilon n] + P[\exists c \in C \backslash \{0\}, |d(c,y) - np| \leq \epsilon n, |wt(y) - np| \leq \epsilon n]$$

By Chernoff bound for i.i.d. Bernoulli random variables, the first portion of error is bounded by

$$P[|wt(y) - np| > \epsilon n] \leq 2e^{-\frac{n\epsilon^2}{3}}.$$

The first term decrease exponentially with n. We only need to study the second term. Specifically, since $\{|wt(y) - np| \leq \epsilon n\}$ happens w.h.p., we can look at the probability

$$\begin{aligned}
&P[\exists c \in C \backslash \{0\}, d(c,y) \approx np, wt(y) \approx np] \\
&\leq \sum_{w=1}^{n} A_w P[d(c,y) \approx np, wt(y) \approx np | wt(c) = w] \\
&= \sum_{w=1}^{n} A_w \frac{\binom{w}{w/2}\binom{n-w}{np-w/2}}{\binom{n}{np}}.
\end{aligned}$$

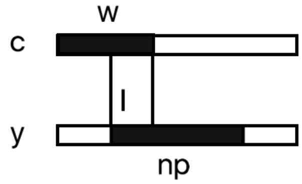

Figure 1: Calculation of the probability of $\{d(c,y) \approx np, wt(y) \approx np | wt(c) = w\}$

The last equality can be derived from Figure 2. Assume that there are l non-zero entries of c and y that are overlapped. Since $d(c,y) = np$, we have

$$\begin{aligned}
w - l + np - l &= np \\
l &= w/2.
\end{aligned}$$

Then it follows from the counting of y that are of distance np from a fixed c.

So,

$$P_e \leq 2e^{-\frac{n\epsilon^2}{3}} + \sum_{w=1}^{n} A_w \frac{\binom{w}{w/2}\binom{n-w}{np-w/2}}{\binom{n}{np}}. \tag{1}$$

Equation 1 is a version of the Shannon bound, which holds for general linear codes. For random linear code we can exploit the property of A_w being binomial-like:

$$\begin{aligned}
P_e &\leq \sum_{w=1}^{n} n^2 \frac{\binom{n}{w}\binom{w}{w/2}\binom{n-w}{np-w/2}}{2^{n-k}\binom{n}{np}} \\
&=\leq \frac{n^2}{2^{n-k}} \sum_{w=1}^{n} \frac{n!w!(n-w)!np!(n-np)!}{w!(n-w)!(w/2!)^2(np-w/2)!(n-w/2-np)!n!} \\
&=\leq \frac{n^2}{2^{n-k}} \sum_{w=1}^{n} \frac{np!(n-np)!}{(w/2!)^2(np-w/2)!(n-w/2-np)!} \\
&= \frac{n^2}{2^{n-k}} \sum_{w=1}^{n} \binom{np}{w/2}\binom{n-np}{w/2} \\
&\leq \frac{n^3}{2^{n-k}} \binom{n}{np(1-p)}\binom{n-np}{np(1-p)} \\
&\leq \frac{n^3}{2^{n-k}} \binom{n}{np} \\
&\leq n^3 2^{k-n+nh(p)}.
\end{aligned}$$

Setting $k = n(1-h(p)-\epsilon)$, we have $P_e \to 0$. In that case the rate,

$$R(p) = \lim_{n\to\infty} \frac{k}{n} = 1 - h(p).$$

So as long as $k < n - nh(p)$, we have $R = k/n < 1 - h(p)$. This is called the **Capacity** of the BSC channel.

For now, we know that there exists:

1. a code with rate $1 - h(2p)$ to correct pn adversarial error, and
2. a code with rate $1 - h(p)$ to correct pn random error w.h.p.

Note that for random error it achieves the sphere packing bound.

3 Erasure correction for linear code

Now consider the erasure correction with linear code. Given a parity check matrix H and a received code word x

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} 1 & ? & ? & 0 & 1 & 0 & ? \end{bmatrix}$$

We can decode by solving $Hx = 0$. However, the complexity of solving linear system is $O(n^{2.376})$ to $O(n^3)$. A faster decoding method is available by using *Factor Graph*. First we use the parity check matrix as the adjacency matrix of the bipartite graph. The nodes corresponding to the $n-k$ rows are called check nodes, and the nodes corresponding to the n columns are called variable nodes. Then, we find the check node that has exactly one erased neighbor. The value for this node is the sum of the other neighbors of this check node. In the example, if we label the rows of H as a, b, c and the column as

$[1, ..., 7]$, the first check node we find is a, and the value for node 7 is 0. We repeat this procedure until all the erasures are corrected. The complexity of this method is linear to the size of the parity check matrix.