COMPSCI 690T Coding Theory and Applications Feb 8, 2017

Lecture 6

Instructor: Arya Mazumdar *Scribe: Names Redacted*

1 Random Linear Codes

1.1 Idea of a Random Construction

It is our goal to construct a code which achieves the Gilbert-Varshamov bound. One idea is to try a random construction and then attempt to derandomize. Here, we construct a random linear code. We will see that with high probability, such a code will achieve the Gilbert-Varshamov bound. Unfortunately, it is an NP-Hard question on determining if our code is good, thus making this construction impractical for use.

1.2 Achieving the Gilbert-Varshamov bound

We create a linear code by making a random Parity Check matrix

$$H = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n-k,1} & x_{n-k,2} & x_{n-k,3} & \cdots & x_{n-k,n} \end{bmatrix}$$

Each x_{ij} is a Bernoulli random variable, with a half probability of being either 1 or 0. We know that the code will be the set

$$C := \{x \in \{0,1\}^n | Hx = 0\}$$

From this construction, we have that $R \geq \frac{k}{n}$.

Definition 1 *A_W is # of random codewords with weight w.*

We have that

$$\mathbf{E}[A_w] = \binom{n}{w}\frac{1}{2^{n-k}} \implies \mathbf{E}[\sum_w A_w] = \frac{\sum_{i=1}^{d-1}\binom{n}{i}}{2^{n-k}}$$

Define X as the number of codewords of weight $\leq d-1$. Then

$$\mathbf{Pr}[X \geq n\sum_{i=1}^{d-1}\binom{n}{i}\frac{1}{2^{n-k}}] \leq \frac{1}{n}$$

by Markov's Inequality. Now if we set

$$n\sum_{i=1}^{d-1}\binom{n}{i}\frac{1}{2^{n-k}} = 1 \implies k = \log\frac{2^n}{n\sum_{i=1}^{d-1}\binom{n}{i}}$$

then $X < 1$ with probability $1 - \frac{1}{n}$.
If we set $d = \delta n$ for asymptotics, we see that

$$R = \frac{k}{n} = 1 - \frac{1}{n}\log(n\sum_{i=1}^{\delta n-1}\binom{n}{i}) \implies R = 1 - h(\delta)$$

where $h(x)$ is the binary entropy function.

2 Recovery in Linear Codes

2.1 Correctable Errors

Let C be a code and let $x \in C$. Then the *Set of Correctable Errors* is the set of vectors e such that

$$d(x, x+e) < d(y, x+e) \qquad \forall y \in C \backslash \{x\}$$

Definition 2 *Voronoi Region of* $x \coloneqq \{x{+}e : e \text{ is correctable with } x\}$

For an example, let $C = \{0000, 1100, 1110\}$. Then the vector $e = 0010$ is correctable for 0000, but not for 1100, 1110. Note that if the code C is linear, then the set of correctable errors is the same for each $x \in C$.

Now suppose C is a linear code. Note that C is a subgroup of $\{0,1\}^n$ under the operation of binary addition. By Lagrange's Theorem, the number of cosets of C in $\{0,1\}^n$ is $\frac{2^n}{|C|} = \frac{2^n}{2^k} = 2^{n-k}$. Therefore, there is a bijection between the cosets and the codewords of the dual code.

We define the *Standard Array* to be the tabular array of cosets. For example, here is the standard array of $C = \{0000, 1100, 0011, 1111\}$.

0000	1100	0011	1111	0000+C
0001	1101	0010	1110	0001+C
0100	1000	0111	1011	0100+C
1001	0101	1010	0110	1001+C

Claim 1 *The set of correctable errors is equal to the set of coset leaders which have the unique minimum weights in their cosets.*

Proof Suppose e is a coset leader. We examine x and $x + e$. We see that

$$d(y, x+e) = d(x+y, e) > d(0, e) = d(x, x+e)$$

■ How about for the Hamming Code? The coset leaders are the weight 1 binary vectors.

2.2 Decoding Linear Codes

Definition 3 *Let $x \in \{0,1\}^n$. The Syndrome of x is equal to Hx^T, where H is the Parity Check matrix.*

Notice that we can redefine cosets here as groups of vectors which have the same syndrome. This is because if x, y are in the same coset, then $x = c_1 + r$ and $y = c_2 + r$, for some vector r and two codewords c_1, c_2.

$$Hx^T = H(c_1 + r)^T = H(c_2 + r)^T = Hr^T = Hy^T$$

We now can develop a general decoding method of errors in linear codes:
Say we have a codeword $x \in C$. After transmission, it becomes $r = x + e$, where e is an error vector. Find the coset which has the syndrome Hr^T. Subtract the coset leader of this coset from r, and we will retrieve x.
In the real domain, this process is known as *Sparse Recovery/Compress Sensing.*

2.3 Dual of the Hamming Code

Remember the *Sphere Packing Bound*:

$$A(n,d) \leq \frac{2^n}{\sum_{i=1}^{d-1} \binom{n}{i}}.$$

Codes that satisfy this are called **perfect codes**. We know that Hamming codes are perfect codes. There is also one other code, called the *Golay Code*. This code is [23, 12, 7].

We now define the Dual of the Hamming Code. In this code, we will take the Parity Check matrix of the Hamming Code and use it as a generator matrix. It is known as *Punctured Hadamard/Shortened Reed-Muller/Simplex Code*. We have that the parameters of this code are $[2^m - 1, m, ?]$. We will investigate what the distance is.
Let us examine the case for [7,3,?].

$$G^{(3)} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Note that

$$G^{(3)} = \begin{bmatrix} 0 \quad 0 & 0 & 1 & 1 \quad 1 & 1 \\ G^{(2)} & & 0 & & G^{(2)} & \\ & & 0 & & & \end{bmatrix}.$$

The codewords are $S^{(3)}$

0000000
0001111
0110011
1010101
0111100
1011010
1100110
1101001

But note that the codewords can be written as:

$$S^{(3)} = \begin{bmatrix} & 0 & \\ & 0 & \\ S^{(2)} & 0 & S^{(2)} \\ & 0 & \\ & 1 & \\ S^{(2)} & 1 & \bar{S^{(2)}} \\ & 1 & \\ & 1 & \end{bmatrix}.$$

Thus, the distance $d = 4$.

Claim 2 *In general:* $d_{min} = 2^{m-1} = \frac{n+1}{2}$

Thus, this code is $[n, \log(n+1), \frac{n+1}{2}]$. Since,

$$A(n,d) \leq \frac{2d}{2d-n} = n + 1.$$

This means that the Punctured Hadamard code matches the Plotkin bound.