CS6845: Pseudorandomness 1 Mar 2018

23. Asymptotically good codes from expanders

23.1 Introduction

In the last lecture, we saw the Hamming code that corrects one bit of error. Typically, we would need a family of codes $\{\mathcal{C}_i\}$ where $\mathcal{C}_i = [n_i, k_i, d_i]_q$ code. In this lecture we will see a construction of an asymptotically good family of linear codes using bipartite expanders.

23.2 Asymptotics of codes

For an $[n, k, d]_q$ code, the rate of the code $R = k/n$ and the relative distance of the code $\delta = d/n$. The rate of the code measures how efficient it is in encoding the message, and the relative distance measures what is the amount of errors it can correct. An asymptotically good code is one where the rate and relative distance is constant for the entire family of codes.

Definition 23.1. *A family of codes $\mathcal{C}_i$ is an* asymptotically good code *if there exists $R, \delta > 0$ such that for every $\mathcal{C}_i \in \mathcal{C}$ that is an $[n_i, k_i, d_i]_{q_i}$ code, $d_i > \delta n_i$ and $k_i/n_i > R$.*

What can we say about the Hamming code that we saw in the last lecture. The relative distance of the code is bad as it corrects just one error, irrespective of n_i and k_i. So does there exist asymptotically good codes. We will now prove their existence without giving an explicit construction.

Theorem 23.2 (Gilbert's greedy code)**.** *For every $\delta < 1/2$ and large enough n, there exists a binary code $\mathcal{C}$ with relative distance δ and rate $R \geq 1 - H(\delta)$.*

Proof. We contruct the code in the following greedy manner. Let $S = \{0,1\}^n$ and $\mathcal{C} = \emptyset$ initially. Pick a string $\mathbf{x} \in S$ and add it to $\mathcal{C}$. Remove all strings $\mathbf{y} \in \{0,1\}^n$ from S such that $\Delta(\mathbf{x}, \mathbf{y}) \leq \delta n$. Continue this process until S is empty. We claim that $\mathcal{C}$ is a code with the desired property.

To see this note that whenever we add a new codeword to $\mathcal{C}$, we remove at most $\mathrm{Vol}(\delta n, n)$ many strings from $\{0,1\}^n$. Therefore we can continue this process for at least $2^n / \mathrm{Vol}(\delta n, n)$ many steps. Since $\mathrm{Vol}(\delta n, n) \leq 2^{H(\delta)n}$ when $\delta n \geq 1$, we have $|\mathcal{C}| \geq 2^{n(1-H(\delta))}$. □

In fact this theorem is true even when we restrict $\mathcal{C}$ to be a linear code.

Theorem 23.3 (Varshamov's linear code)**.** *Let $\delta < 1/2$ and $\varepsilon > 0$. For large enough n, there exists a binary linear code $\mathcal{C}$ with relative distance δ, and rate $R = 1 - H(\delta) - \varepsilon$.*

Proof. For a fixed n, and $k = Rn$, choose a matrix $G \in \mathbb{F}_2^{k \times n}$ uniformly at random. We will argue that, with high probability, G generates a code with 2^k codewords and distance at least δn. To show that G has distance at least δn, it is sufficient to show that for every $\mathbf{y} \in \{0,1\}^k$, $\mathbf{y}G \notin B(0, \delta n)$. This is also sufficient to show that $\mathcal{C}$ has 2^k codewords (why?). First observe that for any $\mathbf{y} \in \{0,1\}^k$, $\mathbf{y}G$ is a random vector in $\{0,1\}^n$ (why?). Therefore, we have the following:

$$\begin{aligned}\Pr[\mathbf{y}G \in B(0,\delta n)] &= \frac{\mathrm{Vol}(\delta n, n)}{2^n} \\ &\leq 2^{(H(\delta)+\varepsilon/2-1)n}.\end{aligned}$$

Therefore by the union bound we can say that the probability that there exists a $\mathbf{y}$ such that $\mathbf{y}G \in B(0,\delta n)$ is at most $2^k 2^{(H(\delta)+\varepsilon/2-1)n}$. Since $R = k/n = 1 - H(\delta) - \varepsilon$, we have that this probability is at most $2^{-\varepsilon n/2}$. □

It is not known how to construct explicit codes that achieve the Gilbert-Varshamov bounds. In the remainder of this lecture, we will see how to construct asymptotically good codes with efficient encoding and decoding algorithms from bipartite expander graphs.

23.3 Expander codes

We know that random bipartite graphs are good expanders for certain parameters. In this section we will see how to use the properties of the expander to construct good codes. We will not prove that explicit expanders with the properties that we require can be constructed. Our aim will be to analyse the codes coming out of these expanders, assuming good explicit constructions.

Let $H \in \mathbb{F}_2^{n \times (n-k)}$ be the parity check matrix of a code $\mathcal{C}$. We can think of H also as a bipartite graph (V_L, V_R, E) where $|V_L| = n$ and $|V_R| = n - k$. The number of parity check constraints in which each $i \in V_L$ is involved is the *left degree* of the graph. We will look at bipartite graphs that are D-left regular.

For this discussion we want expanders, where for every $S \subseteq V_L$ such that $|S| \leq d$, $|N(S)| \geq L(H,d)|S|$, where $L(H,d)$ is known as the left-expansion ratio. For the remainder of this discussion the set $N(S)$ will denote the neighbors of S excluding the set S.

Theorem 23.4 (Sipser & Spielman)**.** *If $L(H,d) > D/2$, then the minimum distance of $\mathcal{C}$ is $> d$.*

Proof. Suppose that for all $S \subseteq V_L$ such that $|S| \leq d$, $\exists\, u \in V_R$ such that $|N(u) \cap S| = 1$. Then, we can prove the theorem as follows: Let $\mathbf{x} \in \{0,1\}^n$ be such that $\mathrm{wt}(\mathbf{x}) \leq d$. Let $S = \{v \in V_L \mid \mathbf{x}_v \neq 0\}$. Therefore, $\exists\, u \in N(S)$ that has a single neighbor in S. This means that there is a column in H (corresponding to u) such that $\exists v \in S$ and $H[v,u] \neq 0$ and $H[w,u] = 0$ for all $w \in S \setminus \{v\}$. Therefore $\mathbf{x}H \neq 0$.

Now to show that for all $S \subseteq V_L$ such that $|S| \leq d$, $\exists\, u \in V_R$ such that $|N(u) \cap S| = 1$, we proceed as follows: We know that $|N(S)| > D|S|/2$, and $|E(S, N(S))| = D|S|$. Therefore $\exists u \in N(S)$ such that $|N(u) \cap S| = 1$ since if for all u, $|N(u) \cap S| \geq 2$, then the edges going from V_R to V_L is at least $2|N(S)| > D|S|$. But this contradicts the observation that $|E(S, N(S)| = D|S|$. □