(67902) Topics in Theory and Complexity Nov 22, 2006

Lecture 4

Lecturer: Irit Dinur *Scribe: Omer Lev*

1 Expansion - General Notions

1.1 A Bad Code as a Useful Example

Construct a parity-check matrix H (from which an encoding matrix G can be derived) of size $n \times (n-k)$, so that $H_{ij} = 1$ with a probability of $\frac{10}{n}$, 0 otherwise. The code, C, is $C = \{y | yH = 0\}$.
The probability for a row of zeros is $(1 - \frac{10}{n})^{n-k}$, which is fairly high. Therefor, 2 words - y_1, y_2 which are differentiated by 1 bit, can both be legitimate words (because the 1 bit is multiplied by the zero row) - $y_1 H = y_2 H = 0$. This means the code distance is 1 - which is the worst distance possible.

1.2 Modeling a Graph

Definition 1 *Every matrix H of size $n \times m$ defines a 2-sided graph $G = (L \cup R, E)$.$|L| = n, |R| = m$. $H_{ij} = 1 \iff (i,j) \in E$. From now on, L will denote the left vertices, and R, the right ones*

Definition 2 *A bipartite graph G is called* (c,d)-regular *if all the vertices on the left have a rank of c, and all vertices on the right have a rank of d, (thus, $c \cdot n = d \cdot m$).*

Let S be a set of vertices. We denote the neighbors of S be $\Gamma(S) = \{ v \mid \exists u \in S, (u,v) \in E\}$.

Definition 3 *A (c,d)-regular graph will be called a* (c,d,δ,γ)-expander *for $0 < \delta < 1, 0 < \gamma < 1$), if for every $S \subseteq L$, $|S| \leq \delta n$ then $|\Gamma(S)| \geq |S| \cdot c \cdot \gamma$.*

Theorem 4 *For every $\varepsilon > 0$, there are $c, d, \delta > 0$ so there is an infinite family (so that n keeps growing) of (c,d,δ,γ)-expanding graphs.*

Sketch of Proof Choose a random matching between two sets of cn vertices. On the right, we combine every d vertices, and on the left, combine every c vertices. This gives a random (c,d)-*regular* graph, possibly with parallel edges.

For any fixed set $S \subseteq L$, $|S| \leq \delta n$ let us calculate the expected number of neighbors of S. A tail inequality is then used to show that with (very) high probability $|\Gamma(S)|$ doesn't deviate much from the expectation. This holds even after taking a union over all sets S of cardinality at most δn. ■

Definition 5 *If $S \subseteq L$, then $\Gamma_{unique}(S) = \{v \in R || \Gamma(v) \cap S| = 1$ - the vertex has only one "neighbor" in $S\}$.*

Lemma 6 *Let G is a (c,d,δ,γ)-expander, and let $S \subseteq L$, $|S| \leq \delta n$. Then $|\Gamma_{unique}(S)| \geq (2\gamma - 1) \cdot c \cdot |S|$.*

Proof Consider a set of vertices S. Define $A = \Gamma_{unique}(S)$ and $B = \Gamma(S) \setminus A$. Notice that $|A| + |B| \geq c\gamma|S|$ (since it's a (c,d,δ,γ)-*regular* graph). Also, recall that the total number of edges from S is $c \cdot |S|$ (because it is a (c,d)-*regular* graph), and A receives $|A|$ edges from S, while B receives $2 \cdot |B|$ edges.

$$c \cdot \gamma \cdot |S| \leq |A| + |B| \leq |A| + 2 \cdot |B| \leq c \cdot |S|$$

$$\Rightarrow |B| \leq c|S|(1 - \gamma)$$

$$\Rightarrow |\Gamma_{unique}(S)| = |A| \geq c \cdot \gamma |S| - |B| \geq c \cdot \gamma \cdot |S| - c \cdot |S| \cdot (1 - \gamma) = c \cdot |S| \cdot (2\gamma - 1)$$

■

2 Sipser-Spielman/Gallager Code

2.1 What is the Code?

Let G be a (c, d, δ, γ)-expander. The code $C(G)$ contains all strings $x \in \{0,1\}^n$ such that:

$$\forall j \in R, \qquad \sum_{i \in \Gamma(j)} x_i = 0 \mod 2 .$$

One can think of it as if x is "laid upon" L, and the each vertex in R "sees" 0 *mod* 2 "1"s in the vector.

Lemma 7 *If $\gamma > \frac{1}{2}$, then the relative distance of the code $C(G)$ is at least δ.*

Proof Since the code is linear, it is enough to prove that $wt(x) \geq \delta n$ for every $0 \neq x \in C(G)$.

Let's assume there is an $x \in C(G)$ so that $wt(x) < \delta n$. Define $S = \{\, i \in [n] \mid x_i = 1\}$. According to Lemma 6, $|\Gamma_{unique}(S)| \geq (2\gamma - 1) \cdot c \cdot |S| > 0$. For $j \in \Gamma_{unique}(S)$, the j constraint (according to the definition - $\sum_{i \in \Gamma(j)} x_i = 0 \; mod \; 2$) isn't fulfilled, and therefore $x \notin C(G)$. ■

2.2 Sipser-Spielman Decoding Algorithm

Suppose $x \in \{0,1\}^n$, and it is guaranteed that $dist(x, C(G)) \leq \frac{\delta n}{2}$ (which is, as half the distance, an error-correcting distance). We want to find the unique word $\tilde{x}$ which is the closest one to x.

Definition 8 *For a given word $x \in \{0,1\}^n$, a vertex $j \in R$ is* **satisfied** *if*

$$\sum_{i \in \Gamma(j)} x_i = 0 \mod 2.$$

Otherwise, it is **unsatisfied**.

The Algorithm: If there is an $i \in L$ that has more unsatisfied adjacent vertices than satisfied ones - it is flipped ($x_i = 1 - x_i$).

It is not hard to implement the algorithm in *linear* time.

Theorem 9 *If $\gamma \geq \frac{3}{4}$, then the algorithm corrects $\frac{\delta}{2} \cdot n$ errors.*

Proof Denote by $\tilde{x}$ the codeword closest to x, and let $S = \{\, i \in L \mid x_i \neq \tilde{x}_i\}$. By assumption, $|S| \leq \delta n/2$.

- The number of unsatisfied constraints is at most $|S| \cdot c \leq \frac{\delta n c}{2}$. This is since a constraint can only be unsatisfied if it is adjacent to a vertex in S.
- Each iteration of the algorithm decreases the number of unsatisfied constraints. This is because a bit is flipped when more of its neighbors are unsatisfied than satisfied. The flip causes all satisfied neighbors to become to unsatisfied and vice versa.
- As long as $0 < |S| < \delta n$, the algorithm doesn't stop. (In other words, there is some bit that has more unsatisfied neighbors than satisfied ones). This follows from Lemma 6 since

  $$|\Gamma_{unique}(S)| \geq (2\gamma - 1)c\,|S| \geq \frac{c}{2}\,|S|$$

  (since $\gamma \geq \frac{3}{4}$). Therefore, there is at least one vertex $v \in S$ that has more than $\frac{c}{2}$ adjacent vertices in Γ_{unique}. Now, since all vertices in Γ_{unique} are unsatisfied, it follows that v has more unsatisfied neighbors than satisfied ones, and should be flipped.

- Throughout the algorithm, $|S| < \delta n$.

  S increases in jumps of 1. If $|S| = \delta n$ then $\Gamma_{unique}(S) > \frac{c}{2}|S| = \frac{c}{2}\delta n$ which lower-bounds the number of unsatisfied constraints at that point. But since we started with $|S| = \frac{\delta n}{2}$, the number of unsatisfied constraints was at most $|S| \cdot c \leq \delta cn/2$ and each step only decreases the number of unsatisfied constraints.

It follows that the algorithm stops when $|S| = 0$, i.e. we have reached the word $\tilde{x}$ as claimed. ■

For this code, the encoding is quadratic (basically, find a generator matrix for the code, and encoding amounts to vector-matrix multiplication. There is another code (also based on expanders), discovered by Spielman, which enables linear time encoding *and* decoding.

3 Expanders

We will focus on undirected *d-regular* graphs (every vertex has d edges).

3.1 Definitions

Definition 10 *$G = (V, E)$ is a d-regular graph. $S, T \subset V$, $S \cap T = \phi$. We define $E(S,T) = \{(u,v) | u \in S, v \in T, (u,v) \in E\}$. Notice that the expected value of $|E(S,T)| = |S| \cdot d \cdot \frac{|T|}{n}$, because $|S|d$ is the number of edges "leaving" S, and $\frac{|T|}{2}$ is the chance that those edges will "reach" somewhere in T.*

Definition 11 *For a set $S \subseteq V$, the edge boundary of S is $\partial S = E(S, V \setminus S)$.*

Definition 12 *Graph Gs expansion is $h(G) = \min_{S||S| \leq \frac{|V|}{2}} \frac{|\partial S|}{|S|}$.*

Definition 13 *A family of d-regular graphs $\{G_i\}_{i>0}$ is an ε-expander family if:*

- *$|G_i| = n_i$ and $n_i \longrightarrow \infty$ (but not too fast - we want every n to have an expander of a size that's "close" to it).*
- *$h(G_i) > \varepsilon$.*

Definition 14 *A family $\{G_i\}$ will be called explicit if there is an algorithm, that given 1^i creates, in polynomial time, the graph G_i.*
A family is very explicit if there is a polynomial algorithm that given (i, u, k) (for $u \in 1 \ldots u_i$ and $k \in 1 \ldots d$) gives the vertex $v \in V$ that is the kth neighbor of u in G_i.

3.2 Examples

- Random *d-regular* graphs.
- Margulis/Gaber - Galil:
  Take $\mathbb{Z}_m$, and define $V = \mathbb{Z}_m \times \mathbb{Z}_m$.
  $(x, y) \in V$ is connected with an edge to $(x + y, y), (x - y, y), (x, x + y), (x, x - y)$.
  This is a 4-regular graph, and there is a ε so it is an ε-expander. We must remove $(0, 0)$ in order to have this graph connected. Observe that there are parallel edges from vertices $(0, a)$ or $(a, 0)$.
- LPS (Lubotzky - Phillips - Sarnak):
  Using a prime number p, we define $V = \mathbb{Z}_p$.
  $x \in V$ is connected to $x - 1, x + 1, x^{-1}$. Again 0 needs to be treated separately, as no 0^{-1} exists, so we add a self-loop.
- Zig-Zag graphs / RVW (Reingold - Vadhan - Wigderson).