Linear Locally Decodable Codes **Fall 2016**

Lecture 2: Lower Bound for 2-LDCs

Lecturer: Zeev Dvir *Scribe: Kalina Petrova*

We will start by proving a result that may seem unrelated to Locally Decodable Codes, but will be useful in proving a lower bound for 2-LDCs.

Definition 2.1. A *hypercube* is the graph H_k on the vertex set $V = \{0,1\}^k$, with edges $E = \{(\mathbf{x},\mathbf{y})|dist(\mathbf{x},\mathbf{y}) = 1\}$.

Lemma 2.1 (Edge-Isoperimetric Inequality for the Hypercube)**.** Let $S \subseteq \{0,1\}^k$, denote $F(S) = \{(\mathbf{x},\mathbf{y}) \in S^2|dist(\mathbf{x},\mathbf{y}) = 1\}$. Then $|F(S)| \leq |S| \log |S|$.

Proof. We are going to prove the lemma by induction on $|S|$. The base cases are $|S| = 2$, which is trivial, since $|F(S)| \leq 1 < 2$, and $|S| = 1$, which is also trivial, since $|F(S)| = 0 \leq 1 \times \log 1 = 0$. Now suppose $|S| > 2$. Choose i such that S has an edge in direction i (that is, there are $\mathbf{x},\mathbf{y} \in S$ such that $(\mathbf{x},\mathbf{y}) \in F(S)$ and $\mathbf{x}-\mathbf{y} = \mathbf{e}_i$). Let $S_0 = \{x \in S|x_i = 0\}$ and $S_1 = \{x \in S|x_i = 1\}$. Now the edges in S are the union of the edges in S_0, the edges in S_1 and the edges between S_0 and S_1. The edges between S_0 and S_1 have to all be of the form $(\mathbf{v},\mathbf{w})$, where $\mathbf{v}-\mathbf{w} = \mathbf{e}_i$. Each element $\mathbf{u}$ of S_0 can have at most one neighbour in S_1, namely $\mathbf{u}-\mathbf{e}_i$, and analogically each element of S_1 can have at most one neighbour in S_0. Therefore,

$$|F(S)| \leq |F(S_0)| + |F(S_1)| + \min\{|S_0|, |S_1|\}$$

Suppose without loss of generality that $|S_0| = d \leq \frac{|S|}{2}$. Set $n = |S|$. Then by induction:

$$\begin{aligned} |F(S)| &\leq d\log d + (n-d)\log(n-d) + d \\ &= d\log(2d) + (n-d)\log(n-d) \\ &\leq d\log n + (n-d)\log n \\ &= n\log n \end{aligned}$$

□

Exercise 2.1 (Improved Edge-Isoperimetric Inequality for the Hypercube)**.** As above, let $S \subseteq \{0,1\}^k$, and denote $F(S) = \{(\mathbf{x},\mathbf{y}) \in S^2|dist(\mathbf{x},\mathbf{y}) = 1\}$. Show that $|F(S)| \leq \frac{1}{2}|S| \log |S|$.

The main result we prove in this lecture is a lower bound for Locally Decodable Codes on n in terms of k. We will first show a special case of the lower bound for $\mathbb{F}_2$, and then we will generalize it to $\mathbb{F}_q$.

Theorem 2.1 ([GKST02])**.** Let $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a $(2,\delta,\varepsilon)$-LDC without repetitions, then $n \geq 2^{\Omega(\delta k)}$.

Proof. Let $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \{0,1\}^k = \mathbb{F}_2^k$ be the rows of the generating matrix of E, and suppose for now that E is without repetitions. Using Theorem 1.1, let E be given in matching form by the k 2-matchings on $[n]$ $M^1,\ldots,M^k$ and the n vectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$. Each pair $(\mathbf{v}_j,\mathbf{v}'_j) \in M^i$ is of one of the following two types.

1. $\mathbf{e}_i = \mathbf{v}_j$ or $\mathbf{e}_i = \mathbf{v}_{j'}$.
2. $\mathbf{v}_j$ and $\mathbf{v}_{j'}$ differ only at the i-th coordinate.

Out of all pairs from all M^i's, at most n are of Type 1. This is because for each j, $\mathbf{v}_j$ can equal $\mathbf{e}_i$ for at most one value of i, which means that each $\mathbf{v}_j$ can participate in at most one pair of Type 1 (since for each i, $\mathbf{v}_j$ can participate in at most one pair in M^i), and there are n $\mathbf{v}_j$'s.

We are now going to use Lemma 2.1 to prove the theorem. Set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, and let $F(S)$ be the set of vector pairs $(\mathbf{v}_j, \mathbf{v}_{j'})$ such that for some $i \in [k], (\mathbf{v}_j, \mathbf{v}_{j'}) \in M^i$. Then Lemma 2.1 gives us $|F(S)| \leq n \log n$. Note that we can apply the Lemma because E is without repetitions. Going back to our two types of pairs in the matchings M^i, since at most n of them are of Type 1, then at least $\sum_{i=1}^{k} |M^i| - n \geq \frac{\delta nk}{2} - n$ must be of Type 2. However, the pairs of Type 2 are members of $F(S)$. Thus, we get

$$n \log n \geq |F(S)| \geq \frac{\delta nk}{2} - n$$

$$\log n \geq \frac{\delta k}{2} - 1$$

$$n \geq 2^{\frac{\delta k}{2} - 1}$$

This completes the proof of the theorem for $q = 2$ and code without repetitions.

□

Exercise 2.2. Prove the 2-LDC lower bound for codes with repetitions over $\mathbb{F}_2$.

Field reduction for 2-LDCs

Lemma 2.2 ([DS07])**.** Let $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ be a $(2, \delta, \varepsilon)$-LDC. Then there exists a $(2, \delta', \varepsilon')$-LDC $E' : \mathbb{F}_2^{\frac{k}{2}} \to \mathbb{F}_2^n$ such that $\delta' = O(\delta)$ and $\varepsilon' = O(\delta)$.

Proof. Suppose E is given by $(\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n) \in \left(\mathbb{F}_q^k\right)^n$ and 2-matchings $M^1, \ldots, M^k$. By Exercise 1.5, we can always construct another $(2, \delta_2, \varepsilon_2)$-LDC $E_2 : \mathbb{F}_q^k \to \mathbb{F}_q^n$ given by $(\mathbf{v'}_1, \mathbf{v'}_2, \ldots, \mathbf{v'}_n) \in \left(\mathbb{F}_q^k\right)^n$ and 2-matchings $M'^1, \ldots, M'^k$, such that $\forall i, \forall (\mathbf{v'}_j, \mathbf{v'}_{j'}) \in M'^i$, either $\mathbf{v'}_j - \mathbf{v'}_{j'} = \mathbf{e}_i$ or $\mathbf{v'}_j - \mathbf{v'}_{j'} = -\mathbf{e}_i$, and $\delta_2 = O(\delta)$ and $\varepsilon_2 = O(\varepsilon)$. By the same argument as in the proof of Theorem 2.1, for at least $\frac{\delta kn}{2} - n$ matching pairs $\left(\mathbf{v'}_j, \mathbf{v'}_{j'}\right) \in M^i$ for some i, $\mathbf{v'}_j$ and $\mathbf{v'}_{j'}$ differ only in the i-th coordinate. Call these matching pairs *good pairs*. Now take a random map $T : \mathbb{F}_q \to \{0, 1\}$. For any good pair $(\mathbf{v'}_j, \mathbf{v'}_{j'})$ in M'^i, the probability of $\left(T(\mathbf{v'}_j), T(\mathbf{v'}_{j'})\right)$ being a good pair is $\frac{1}{2}$, since all digits of the two vectors apart from the i-th will map to the same value (since they are the same), and the i-th digits will map to different values with probability $\frac{1}{2}$. In the second case, $\{T(\mathbf{v'}_j), T(\mathbf{v'}_{j'})\}$ will span $\mathbf{e}_i$. Thus, by linearity of expectation, the expected value of good pairs among $T(\mathbf{v'}_1), T(\mathbf{v'}_2), \ldots, T(\mathbf{v'}_n)$ is at least $\frac{1}{2}\left(\frac{\delta kn}{2} - n\right)$. Thus we have a new code $E_1 : \mathbb{F}_q^k \to \mathbb{F}_q^n$, given by vectors $T(\mathbf{v'}_1), T(\mathbf{v'}_2), \ldots, T(\mathbf{v'}_n)$ and 2-matchings $M''^1, \ldots, M''^k$. Then we have that $\sum_{i=1}^{k} |M''^i| \geq \alpha \sum_{i=1}^{k} |M'^i|$, where $\alpha = O(\delta_2)$. Now by Exercise 1.6, we can use E_1 to obtain a $(2, \delta', \varepsilon')$-LDC $E' : \mathbb{F}_q^{\frac{k}{2}} \to \mathbb{F}_q^n$ with $\delta' = O(\alpha) = O(\delta_2) = O(\delta)$ and $\varepsilon = O(\alpha) = O(\delta_2) = O(\delta)$.

□

Corollary 2.1. Let $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ be a $(2, \delta, \varepsilon)$-LDC, then $n \geq 2^{\Omega(\delta k)}$.

Proof. By Theorem 2.2, there is a $(2, \delta', \varepsilon')$-LDC $E' : \mathbb{F}_2^{\frac{k}{2}} \to \mathbb{F}_2^n$ for some $\delta' = O(\delta)$ and $\varepsilon' = O(\delta)$. Next, by Theorem 2.1, $n \geq 2^{\Omega(\delta' \frac{k}{2})}$, therefore $n \geq 2^{\Omega(\delta k)}$.

□

Notice that in the proof of Corollary 2.1 we did not use the fact that the field we were working with is finite, so Corollary 2.1 holds for $\mathbb{R}$ and $\mathbb{C}$ as well.

Open Problem 2.1 (Removing repetition)**.** Let $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ be an r-LDC. Show that there exists a r'-LDC $E' : \mathbb{F}_q^k \to \mathbb{F}_q^{n'}$ without repetitions for some $r' = O(r)$ and with $n' = poly(n)$ that can be obtained from E.

Exercise 2.3 ([KS07])**.** Show that a random code $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ of size $n = poly(k)$ is not an r-LDC with high probability. For what value of n (in terms of k) does this argument break? Show that a random code $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ with $k = \log n$ is an r-LDC with high probability.

References

[DS07] Zeev Dvir and Amir Shpilka. Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. *SIAM Journal on Computing*, 36:1404–1434, 2007.

[GKST02] O. Goldreich, H. Karloff, L.J. Schulman, and L. Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings 17th IEEE Annual Conference on Computational Complexity*, 2002.

[KS07] Tali Kaufman and Madhu Sudan. Sparse random linear codes are locally decodable and testable. In *Foundations of Computer Science, 2007. FOCS '07. 48th Annual IEEE Symposium on*, 2007.