Linear Locally Decodable Codes **Fall 2016**

Lecture 4: Lower Bounds for r-query LDCs

Lecturer: Zeev Dvir *Scribe: Kalina Petrova*

Theorem 4.1 ([KT00]). If $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is an (r, δ)-LDC, then $n \geq C_{r,\delta} \cdot \frac{k^{1+\frac{1}{r-1}}}{\log_e^{\frac{1}{r-1}}(\frac{k}{0.1})}$, where $C_{r,\delta} = \frac{\delta^{\frac{1}{r-1}}}{6^{\frac{r}{r-1}} r^{\frac{1}{r-1}}}$.

Proof. Suppose first that $r \geq 2$. Pick a subset $S \subseteq [n]$ by taking each $\ell \in [n]$ independently with probability $p = \frac{r^{\frac{1}{r}}}{\delta^{\frac{1}{r}}} \cdot n^{-\frac{1}{r}} \log_e^{\frac{1}{r}} \frac{n}{0.1}$. Thus, with high probability $|S| \approx np$. Let the matching form of E be given by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and the r-matchings $M^1, \ldots, M^k$ with $\forall i, M^i = \left(T_1^i, \ldots, T_m^i\right)$, where $m = \frac{\delta n}{r}$. Let $P_p = Prob[\forall i \in [k], \exists j \in [m], T_j^i \subseteq S]$.

Claim 4.1.

$$1 - P_p \leq n(1-p^r)^{\frac{\delta n}{r}} \leq ne^{-\frac{\delta n}{r} p^r}$$

Proof. For any $i \in [k]$ and $j \in [m]$, we have that $Pr[T_j^i \subseteq S] = p^r$, since for each $t \in T_j^i$, $t \in S$ with probability p. Thus, $Pr[T_j^i \nsubseteq S] = 1 - p^r$. From here, since $T_1^i, \ldots, T_m^i$ are disjoint, we get $Pr[\forall j \in [m], T_j^i \nsubseteq S] = (1-p^r)^{|M^i|} = (1-p^r)^{\frac{\delta n}{r}}$. Finally, by the Union Bound the probability that $\exists i \in [k]$ such that $\forall j \in [m], T_j^i \nsubseteq S$ is at most $\sum_{i=1}^k Pr[\forall j \in [m], T_j^i \nsubseteq S] = n(1-p^r)^{\frac{\delta n}{r}}$, and since $Pr[\exists i, \forall j \in [m], T_j^i \nsubseteq S] = 1 - P_p$, we get that $1 - P_p \leq n(1-p^r)^{\frac{\delta n}{r}}$. The second inequality follows from $x + 1 \leq e^x$.

□

Now notice that $ne^{-\frac{\delta n}{r} p^r} = 0.1$ for our choice of p. This means that $1 - P_p \leq 0.1$, so with probability 0.9, S is such that $\forall i \in [k], \exists j \in [m]$, S contains T_j^i. This means that with probability 0.9, $\{\mathbf{b}_j | j \in S\}$ spans $\mathbf{e}_1, \ldots, \mathbf{e}_k$, which implies that with probability 0.9, $k \leq |S|$. Now we are going to use the Chernoff Bound to bound $|S|$ with respect to n.

Chernoff Bound: If $X_1, \ldots, X_n$ are independent and identically distributed 0/1 random variables with $\mathbb{E}[X_j] = \mu$, then $Pr[|\sum_{j=1}^n X_j - \mu n| \geq \epsilon n] \leq 2e^{-\epsilon^2 n}$.

If we take $\forall i \in [n], X_i$ to be the indicator variable that we have chosen $i \in S$, then $\mu = p$. Take $\epsilon = p$. We get that $Pr[|S| > 2np] \leq e^{-p^2 n}$. Notice that since $r \geq 2$, $p \geq \frac{1.6}{\sqrt{n}}$ as we have chosen it, which implies that $e^{-p^2 n} \leq 0.1$, so $Pr[|S| > 2np] \leq 0.1$. This means that with probability at least 0.9, $|S| \leq 2np$. Since with probability 0.9, $k \leq |S|$, and with probability 0.9, $|S| \leq 2np$, this means that there is a choice for S such that $k \leq |S| \leq 2np$. Therefore, $k \leq 2np = 2n^{1-\frac{1}{r}} \frac{r^{\frac{1}{r}}}{\delta^{\frac{1}{r}}} \log_e^{\frac{1}{r}} \frac{n}{0.1}$. To get the result stated above,

notice that if $n > k^3$, then it follows trivially. So we can assume $n \leq k^3$, which means that $k \leq 2n^{1-\frac{1}{r}} \frac{r^{\frac{1}{r}}}{\delta^{\frac{1}{r}}} \log_e^{\frac{1}{r}} \frac{k^3}{0.1} \leq 6n^{1-\frac{1}{r}} \frac{r^{\frac{1}{r}}}{\delta^{\frac{1}{r}}} \log_e^{\frac{1}{r}} \frac{k}{0.1}$. This gives us $n \geq \frac{6^{-\frac{r}{r-1}} \delta^{\frac{1}{r-1}} k^{1+\frac{1}{r-1}}}{r^{\frac{1}{r-1}} \log_e^{\frac{1}{r-1}} \frac{k}{0.1}}$, which is what we wanted to show.

Note that in the case $r = 1$, the same proof goes through with $p = \frac{1.6}{\sqrt{n}}$, since $\frac{1.6}{\sqrt{n}} > \frac{1}{\delta n} \log_e \frac{n}{0.1}$, so we get that $k \leq 2np = 3.2\sqrt{n}$, therefore $n \geq \frac{k^2}{10.24}$, which is a much stronger result.

□

This lower bound can be improved to roughly $n \geq \Omega(k^{1+\frac{1}{\lceil \frac{r}{2} \rceil - 1}})$ [Woo07]. We will show the special case of $r = 3$ without repetition.

Theorem 4.2. If $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is a $(3, \delta)$-LDC without repetition, then $n \geq \frac{k^2}{3200 \log k}$.

Proof. Let the matching form of E be represented by vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and 3-matchings $M^1, \ldots, M^k$ with $\forall i, M^i = \left(T_1^i, \ldots, T_m^i\right)$, where $m = \frac{\delta n}{r}$. For simplicity, assume $\delta = \frac{1}{4}$. Assume for contradiction that $k \geq 40\sqrt{n \log n}$.

Claim 4.2. There exists a set $S \subseteq [n]$ with $|S| \leq 12\sqrt{n \log n}$, such that $\forall i \in [k]$, S intersects at least $\frac{3m\sqrt{\log n}}{\sqrt{n}} = \frac{(\frac{\delta n}{3}) 3 \log n}{\sqrt{n}} = \delta\sqrt{n \log n}$ of the 3-tuples in M^i.

Proof. Pick a random $S \subseteq [n]$ by taking each $\ell \in S$ independently and identically distributed with probability $\frac{6\sqrt{\log n}}{\sqrt{n}}$, thus $|S| \approx 6\sqrt{n \log n}$. Furthermore, using the Chernoff bound with $X_\ell = 1$ iff $\ell \in S$ and $\mu = \epsilon = \frac{6\sqrt{\log n}}{\sqrt{n}}$, we can get that $Pr[|S| \geq 12\sqrt{n \log n}] \leq e^{-\epsilon^2 n} = e^{-36 \log n}$. Notice that for $n = 2, e^{-36 \log n} = 2.32 \cdot 10^{-16}$, and as n grows this value decreases. Thus with high probability we have that $|S| \leq 12\sqrt{n \log n}$. Now fix $i \in [k]$, then

$$Pr[\text{S intersects a tuple } T_j^i \in M^i] \geq \frac{6\sqrt{\log n}}{\sqrt{n}}.$$

If we take $X_j = 1$ to mean that S intersects the tuple $T_j^i \in M^i$, then using the Chernoff Bound with $\mu = \frac{6\sqrt{\log n}}{\sqrt{n}}$ and $\epsilon = \frac{\mu}{2}$, we get that

$$Pr\Big[\Big|\sum_{j=1}^{m} x_j - \frac{6m\sqrt{\log n}}{\sqrt{n}}\Big| \geq \frac{3m\sqrt{\log n}}{\sqrt{n}}\Big] \leq 2e^{\frac{-9\log nm}{n}}$$

$$\begin{aligned} Pr[S \text{ intersects less than } \frac{3m\sqrt{\log n}}{\sqrt{n}} \text{ of the tuples in } M^i] &\leq e^{\frac{-9\log n\delta n}{3n}} \\ &= e^{-\frac{3\log n}{4}} \\ &= n^{-\frac{3}{4\log_e 2}} \\ &< \frac{1}{n^{1.08}} \end{aligned}$$

Now by the Union Bound, the probability that for any M^i, S intersects fewer than $\frac{3m\sqrt{\log n}}{\sqrt{n}}$ tuples in M^i is at most $\frac{n}{n^{1.08}}$. For $n = 2$, we have $\frac{n}{n^{1.08}} \leq 0.947$, and this value decreases with n. Thus, we get that $\frac{n}{n^{1.08}} + e^{-36\log n} < 1$ for $n \geq 2$, so there exists an S of size at most $12\sqrt{n\log n}$ that intersects at least $\frac{3m\sqrt{\log n}}{\sqrt{n}}$ tuples in M^i for every $i \in [k]$.

□

Take S to be as described above. Then we can prove the following claim.

Claim 4.3. There exists a linear map $L : \mathbb{F}_q^k \to \mathbb{F}_q^{\frac{k}{2}}$ such that:

1. $\forall j \in S, L(\mathbf{v}_j) = \mathbf{0}$.
2. There are $i_1, \ldots, i_{\frac{k}{2}} \in [k]$, such that $L(\mathbf{e}_{i_1}) = \mathbf{e'}_1, \ldots, L(\mathbf{e}_{i_{\frac{k}{2}}}) = \mathbf{e'}_{\frac{k}{2}}$, where $\mathbf{e'}_1, \ldots \mathbf{e'}_{\frac{k}{2}}$ are the standard basis vectors in $\mathbb{F}_q^{\frac{k}{2}}$, and $\mathbf{e}_1, \ldots, \mathbf{e}_k$ are the standard basis vectors in $\mathbb{F}_q^k$.

Proof. Let $\mathbf{w}_1, \ldots, \mathbf{w}_s$ be a basis of $\{\mathbf{v}_j | j \in S\}$. Then since $k \geq 40\sqrt{n\log n}$ and $|S| \leq 12\sqrt{n\log n}$, we have that $|S| \leq \frac{k}{2}$, so there are $\frac{k}{2}$ standard basis vectors $\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{\frac{k}{2}}}$ such that $\{\mathbf{w}_1, \ldots, \mathbf{w}_s, \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{\frac{k}{2}}}\}$ are linearly independent. Define L such that $L(\mathbf{e}_{i_1}) = \mathbf{e'}_1, \ldots, L(\mathbf{e}_{i_{\frac{k}{2}}}) = \mathbf{e'}_{\frac{k}{2}}$ and $L(\mathbf{w}_1) = \cdots = L(\mathbf{w}_s) = \mathbf{0}$, the vector with 0s in all coordinates.

□

Now apply L on $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}_q^k$ to get $\mathbf{u}_1 = L(\mathbf{v}_1), \ldots, \mathbf{u}_n = L(\mathbf{v}_n) \in \mathbb{F}_q^{\frac{k}{2}}$.

Claim 4.4. The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ give the generating matrix of a 2-query LDC with matchings $M'^1, \ldots, M'^{\frac{k}{2}}$ such that $\forall i \in [\frac{k}{2}]$, $|M'^i| \geq \delta\sqrt{n\log n} = \frac{\sqrt{n\log n}}{4}$.

Proof. For any M'^i, we will take some of the pairs from $M^{i'}$, where i' is such that $L(\mathbf{e}_{i'}) = \mathbf{e'}_i$. S intersects at least $\delta\sqrt{n \log n}$ tuples in $M^{i'}$, which means that for at least $\delta\sqrt{n \log n}$ tuples in $M^{i'}$ $\{\mathbf{v}_{j_1}, \mathbf{v}_{j_2}, \mathbf{v}_{j_3}\}$, one of $L(\mathbf{v}_{j_1}), L(\mathbf{v}_{j_2})$, and $L(\mathbf{v}_{j_3})$ is the $\mathbf{0}$ vector, since $L(\mathbf{v}_h) = \mathbf{0}$ if $h \in S$. Suppose without loss of generality that $L(\mathbf{v}_{j_3}) = \mathbf{0}$. Since L is linear, it preserves the matching structure of E, so from $\mathbf{e}_{i'} \in \text{span}\{\mathbf{v}_{j_1}, \mathbf{v}_{j_2}, \mathbf{v}_{j_3}\}$, it follows that $\mathbf{e'}_i \in \text{span}\{L(\mathbf{v}_{j_1}), L(\mathbf{v}_{j_2})\}$. Thus, we have that $(L(\mathbf{v}_{j_1}), L(\mathbf{v}_{j_2})) \in M'^i$, and so $|M'^i| \geq \delta\sqrt{n \log n}$.

□

Recall our previous classification of the pairs in the matchings. For each pair $(\mathbf{v'}_j, \mathbf{v'}_{j'}) \in M'^i$, one of the following two things holds:

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.

As before, at most n of them are of Type 1, and so by the Edge-Isoperimetric Inequality for the Hypercube (Lemma 2.1), we get that

$$\sum_{i=1}^{\frac{k}{2}} |M'^i| \leq n \log n + n$$

On the other hand, $\frac{k\sqrt{n \log n}}{4} \leq \sum_{i=1}^{\frac{k}{2}} M'^i$, so

$$k \leq 4\sqrt{n \log n} + \frac{4\sqrt{n}}{\sqrt{\log n}}$$
$$k \leq 8\sqrt{n \log n},$$

which contradicts our previous assumption that $k \geq 40\sqrt{n \log n}$. Therefore, $k \leq 1600\sqrt{n \log n}$, so $\frac{k^2}{1600} \leq n \log n$. If $n > k^2$, then what we want to prove is true. So assume $n \leq k^2$. Then $\frac{k^2}{1600} \leq n \log n \leq n \log k^2 = 2n \log k$. This means that $n \geq \frac{k^2}{3200 \log k}$.

□

Exercise 4.1. Prove $n \geq \Omega(k^2)$ without $\log k$ factors.

References

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. *Proceeding STOC '00 Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 80–86, 2000.

[Woo07] David P. Woodruff. New lower bounds for general locally decodable codes. In *Electronic Colloquium on Computational Complexity*, 2007.