Lecture 9: Locally Testable Codes and the left-right Cayley complex

Irit Dinur

January 8, 2023

In this lecture we will show a construction of LTCs with constant rate, distance, and locality.

1 LTCs

We have seen the Hadamard code, which is locally testable, but its rate is $\frac{\log n}{n}$. We have mentioned Reed-Muller codes, which are generalizations, and that they are locally testable. This is the famous low degree test and there are many works on various aspects. These codes do not combine both constant rate and constant locality.

Are there LTCs with constant rate, constant distance and constant locality?

Definition 1.1. A code $C \subseteq \mathbb{F}_2^n$ is a (q, β)-LTC if there is an $R \times n$ parity check matrix H such that

1. $C = KerH = \{w \in \mathbb{F}_2^n \mid \ Hw = 0\}$
2. Every row in H has at most q non-zeros
3. $wt(Hw) \geqslant \beta \cdot dist(w, KerH)$

Remark 1.2. H describes, in effect, a randomized tester, as follows

- Choose a random row $r \in \{1, \ldots, R\}$
- Accept iff

$$\langle H(r, \cdot), w \rangle = \sum_{i: H(r,i) \neq 0} w(i) = 0$$

It is immediate that

$$\mathbb{P}[\text{Tester rejects}] = \text{wt}(Hw)$$

Remark 1.3. There is a cohomological viewpoint to this. Let $\dim(C) = k$ and let G be an $n \times k$ generating matrix. The following is a chain complex

$$\mathbb{F}_2^k \xrightarrow{G} \mathbb{F}_2^n \xrightarrow{H} \mathbb{F}_2^R$$

because $ImG = KerH$. Recall that cosystolic-expansion of this chain is

$$\beta = \min_{w \in \mathbb{F}_2^n \setminus C} \frac{\text{wt}(Hw)}{\text{dist}(w, C)}.$$

The chain complex is a β-coboundary expander iff the code is a β-LTC.

2 Expander Codes

Lewt $G = (V, E)$ be a d-regular expander graph, with normalized second largest eigenvalue at most λ. Let $C_0 \subset \mathbb{F}_2^d$ be an error-correcting-code defined by a parity check matrix H_0. Suppose C_0 has distance δ_0 and rate $r_0 = \dim(C)/d$. We define the expander code $C(G, C_0)$ as the kernel of an $|V| \cdot d \times |E|$ parity check matrix defined by placing copies of H_0, for each vertex v on the columns corresponding to the neighbors of v. Alternatively,

$$C(G, C_0) = \left\{ f \in \mathbb{F}_2^E \;\middle|\; f|_{E_v} \in C_0 \right\}.$$

We can also have a different code $C_v \subseteq \mathbb{F}_2^d$ for each vertex $v \in V$.

Rate

If $r_0 > 1/2$ then $\dim(C) = \Omega(|E|)$.
The number of degrees of freedom is $|E| = nd/2$.
The number of constraints is $n \cdot m_0$ where $m_0 = codim(C_0) = (1 - r_0) \cdot d$.
Thus, $\dim(C) \geqslant \frac{nd}{2} \cdot (1 - 2(1 - r_0)) = |E|(2r_0 - 1)$.

Distance

Since this is a linear code, the distance is the weight of the minimal nonzero codeword. Let $0 \neq f \in C$. Let $V' = \{v \in V \mid f|_{E_v} \neq 0\}$. Clearly $V' \neq \phi$ and every vertex in V' has at least δd neighbors also in V'. Recall the Alon-Chung lemma, $|V'| \geqslant (\delta - \lambda)|V|$. Every such vertex touches at least δd non-zero edges, so the support of f has relative size at least $\delta(\delta - \lambda)$.

Lemma 2.1 (Alon-Chung)**.** *Let $G = (V, E)$ be a d-regular λ-one-sided expander. Let $T \subseteq V$ be such that the graph induced on T, denoted $G(T)$, has average degree at least δd. Then $|T| \geqslant (\delta - \lambda) \cdot |V|$, and the number of edges in $G(T)$ is at least $(\delta - \lambda)\delta \cdot |E|$.*

Decoding

Here is an algorithm for decoding a given word $f \in \mathbb{F}_2^E$,

1. Each $v \in V$ finds a locally best codeword $w_v \in C_v$
2. Go through all vertice s in arbitrary order and each v switches to another codeword if it will lessen the disagreement with its neighbors. Repeat until stuck.
3. If no disagreement output $\tilde{w}$ given by $\tilde{w}(e) = w_v(e)$ where $v \in e$ is arbitrary.

Does this terminate? yes, because the number of edges in disagreement decreases.
Does this terminate with perfect agreement?
If so, then we decoded in linear time!

Let

- $E_0 = \{uv \in E \mid w_u(v) \neq w_v(u) \text{ after step 1}\}$
- V_0 - vertices with incorrect local view after step 1
- V_1 - vertices with incorrect local view at the end

Claim 2.2. $\varepsilon|E| \geqslant \delta d/2 \cdot |V_0| \geqslant \delta/2 \cdot |E_0|$.

Proof. First, every incorrect vertex sees at least δd errors, and each error is counted twice. Next, if $uv \in E_0$ then either $u \in V_0$ or $v \in V_0$ so each edge can blame one vertex. Each vertex receives at most d blames, so $|V_0|d \geqslant E_0$. □

Claim 2.3. Every $v \in V_1$ has at least $\delta d/2$ nbrs in V_1

Proof. Otherwise v would have flipped during the algorithm. □

By the Lemma 2.1 again we get that $|V_1| \geqslant |V|(\delta/2 - \lambda)$, and this contradicts our assumption, because $|V_1| \leqslant |V_0| + |E_0| \leqslant \frac{\varepsilon}{\delta}|V| + \varepsilon|E| = O(\varepsilon)|V|$.

Are expander codes locally testable?

They certainly are LDPC.. but the answer is: typically no.

Fix a family of local codes $C_v \subseteq \mathbb{F}_2^d$ and for one vertex v_0 let $C'_{v_0} \supsetneq C_{v_0}$ have $\dim(C'_{v_0}) = \dim(C_{v_0}) + 1$. We can arrange so that both have distance at least δ and rate of all local codes is above $1/2$.

Let C be the expander code with local codes C_v for all v.

Let $C' \supset C$ be the expander code with local codes C_v for all v, except that we let the local code at v_0 be C'_{v_0}. If all constraints are linearly independent, then $dim(C') = \dim(C) + 1$, and let $w \in C' \setminus C$. By our assumption, both codes have large distance which means that $\mathrm{dist}(w, C) \geqslant \delta$. But, on the other hand, for H the parity check of C, we get $\mathrm{wt}(Hw) = 1/R$, as small as could possibly be!

3 Adding one more dimension

Recall from lecture 5, that the Hadamard code had short *constraint loops*. Indeed, there were many short linear dependencies between the constraints. In every loop, an even number of constraints can be unhappy. So, if one constraint is unhappy, it implies that each loop touching it has another unhappy constraint, and we can hope for some propagation.

This was also the case in the proof of cosystolic expansion: Given a 1-chain, the constraints were the triangles, and four triangles in a pyramid made a constraint-loop.

So to get an LTC we need an expander code that also has these loops. Something like:

$$\mathbb{F}_2^{X(2)} \xrightarrow{H} \mathbb{F}_2^{X(1)\cdot m_1} \longrightarrow \mathbb{F}_2^{X(0)\cdot m_0}$$

For this we need to combine a 2-dimensional expander with a local locally testable code that has a local 2-chain.

So far, we don't know how to find a good local code to match the HDX constructions, but we have another HDX for which this is possible: it is not a simplicial complex, but rather a squares complex.

4 Left Right Cayley Complex

4.1 squares complex

Let G be a group, and let $A, B \subset G$ be generator sets. We assume that they are symmetric, namely, $a \in A$ implies $a^{-1} \in A$. We define a squares complex $X[A, G, B]$:

- The vertices are $X(0) = G$.

– The edges are $X(1) = X_A(1) \sqcup X_B(1)$ where $X_A(1) = \{g, ag \mid \ g \in G, a \in A\}$ and $X_B(1) = \{g, gb \mid \ g \in G, b \in B\}$.

– The squares are $X(2) = A \times G \times B / \sim$ where we define $(a, g, b) \sim (a^{-1}, ag, b) \sim (a^{-1}, agb, b^{-1}) \sim (a, gb, b^{-1})$.

We denote a square by $[a, g, b] = \{(a, g, b), (a^{-1}, ag, b), (a^{-1}, agb, b^{-1}), (a, gb, b^{-1})\}$. We also denote an edge by $[a, g]$ for $\{g, ag\}$ and $[g, b]$ for $\{g, gb\}$. We assume that $ag \neq gb$ and $g \neq agb$ for all g, a, b. This assumption is not necessary but makes it simpler to think about squares as never collapsing. This property depends on the group and choice of sets A, B.

Links. The squares touching an edge $[a, g]$ are $[a, g, b]$ and can be identified with the set B. The squares touching an edge $[g, b]$ are $[a, g, b]$ and can be identified with the set A. The squares touching a vertex g are $[a, g, b]$ and can be identified with the set $A \times B$.

Instanciation. which group should we choose? We will need the Cayley graph to be a good expander (for distance and for local testability). So a good choice is any group that has a small set of generators such that $Cay(G, A)$ and $Cay(G, B)$ is a good expander. For example, we can take the LPS Ramanujan expanders [2].

4.2 squares code

Fix two local codes $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$. We define the squares code

$$C(X, C_A, C_B) = \left\{ f \in \mathbb{F}_2^{X(2)} \;\middle|\; \forall a, g, b \; f([a, g, \cdot]) \in C_B \;\text{ and }\; f([\cdot, g, b]) \in C_A \right\}$$

Local tensor code. First thing to observe is that the restriction of the code to squares touching a vertex g is a tensor code. Indeed, fixing g and letting $M(a, b) = f([a, g, b])$ we get an $A \times B$ matrix whose rows are in C_B: $M(a, \cdot) \in C_B$ and whose columns are in C_A: $M(\cdot, b) \in C_A$. In other words, $M \in C_A \otimes C_B$.

Rate

Suppose now that $|A| = |B| = d$ and $C_A = C_B = C_0 \subseteq \mathbb{F}_2^d$. Let $r_0 = \dim(C_0)/d$, and assume $r_0 > 3/4$.

– The number of degrees of freedom is $|X(2)| = |G||A||B|/4 = |G|d^2/4$.

– The number of constraints is $|X(1)|(1 - r_0)d = |G|d^2(1 - r_0) = |X(2)| \cdot 4(1 - r_0)$. We have used that the number of edges is $|X(1)| = |G|(|A| + |B|)/2 = |G|d$.

We get $\dim(C) \geqslant |X(2)|(1 - 4(1 - r_0)) = |X(2)|(4r_0 - 3)$, which is positive whenever $r_0 > 3/4$.

Distance

Assume now that the 1-skeleton of X, namely, the graph $(X(0), X(1))$ is a λ-expander.

Since this is a linear code, the distance is the weight of the minimal nonzero codeword. Let $0 \neq f \in C$. Let $V' = \{v \in V \mid \ f|_{S_v} \neq 0\}$, where S_v are the squares touching v. Recall that $f|_{S_v}$ is a tensor codeword. Observe that every non-zero tensor codeword must have at least δ fraction of nonzero rows and δ fraction of nonzero columns. Therefore, every

$v \in V'$ has at least δ fraction of its neighbors in V'. Again by the Alon-Chung lemma, Lemma 2.1, $|V'| \geqslant (\delta - \lambda)|V|$. Every such vertex touches at least $\delta^2 d^2$ non-zero squares (this is the minimum distance of the code $C_0 \otimes C_0$), so the support of f has relative size at least $\delta^2(\delta - \lambda)$.

4.3 Tensor Codes

Before we move to discuss local testability, we observe that tensor codes also have constraint loops.

Given a code $C \subset \mathbb{F}_2^n$, $\dim(C) = k$, such that $\mathrm{dist}(C) \geqslant \delta$, the tensor code $C \otimes C$ has dimension k^2 and distance at least δ^2. Depicting codewords as matrices, the natural parity check matrix has row constraints and column constraints, with total number $2n(n-k)$. This is larger than the codimension $n^2 - k^2$, so there must be linear dependencies. Indeed, for every α, β a pair of constraints for C, there is a loop consisting of the constraints $(i, \beta) : i \in \alpha$ and $(\alpha, j) : j \in \beta$.

Local testability of squares code

Let H be the parity check matrix of the code C given by collecting all the parity checks from all of the edges of the complex.

Lemma 4.1. *There exists some $\beta > 0$ that depends on $d = |A| = |B|$ but not on $|G|$, such that given $f \in \mathbb{F}_2^{X(2)}$, if $\mathrm{wt}(Hf) < \varepsilon$, then there is some $\tilde{f} \in C$ such that $\mathrm{dist}(\tilde{f}, f) \leqslant \varepsilon/\beta$.*

We consider the following local-correction algorithm that receives a word $f \in \mathbb{F}_2^{X(2)}$.

Algorithm:

1. Every $g \in G$ chooses $w_g \in C_0 \otimes C_0$ that is closest to $f(\cdot, g, \cdot)$.
2. Let $E' = \{\{g, g'\} \in X(1) \mid \ w_g \nsim w_{g'}\}$, where $w_g \nsim w'_g$ means that the local views disagree on some common square.

   For each g, if there is another choice of w_g that minimizes the number of sets in E' touching g, then switch to that local view.

   Repeat until no more available switches.
3. If $E' = \phi$ output $\tilde{f}$ the codeword obtained from the combined local views. Else output fail.

Observe that indeed if $E' = \phi$ then setting $\tilde{f}([a, g, b]) = w_g([a, g, b])$ gives a valid codeword. We will show that if $\mathrm{wt}(Hf)$ is small then $\tilde{f} \approx f$. Observe also that the algorithm must halt because the size of E' decreases at every step.

Let w_g^0 denote the local view of vertex g at the beginning of the algorithm.

Claim 4.2. Let

$$\begin{aligned} V_0 &= \left\{g \in X(0) \,\middle|\, \ w_g^0 \neq f([\cdot, g, \cdot])\right\} \\ V_1 &= \{g \in X(0) \mid \ w_g \text{ changed value during the algorithm}\} \\ E'_0 &= E' \text{ at the start of the algorithm.} \end{aligned}$$

Then

– $|X(2)| \cdot \mathrm{dist}(\tilde{f}, f) \leqslant (|V_0| + |V_1|) \cdot d^2$.

– $|V_0| \leqslant 2|Hf|$

– $|V_1| \leqslant |E_0'| \leqslant |V_0|d \leqslant 2d|Hf|$

Therefore, $\operatorname{dist}(f, \tilde{f}) \leqslant (2d+2)\operatorname{wt}(Hf) \cdot d^2 \cdot \frac{1-r_0}{4} = O(\operatorname{wt}(Hf))$.

Proof. For the first item, if $v \notin V_0$ then initially w_v agrees with f on every square s touching v. This can only change if at some point in the algorithm w_v changes value. So $\operatorname{dist}(\tilde{f}, f) \leqslant (|V_0| + |V_1|) \cdot d^2$.

For the second item, observe that every vertex in V_0 sees some violated constraint. Each constraint can be counted at most twice.

For the third item clearly the first and last inequality hold. The middle inequality is because every disagreement between a pair of local views means that at least one of them is in V_0. □

In conclusion, if $\operatorname{wt}(Hf) = 0(1)$ then $f \approx \tilde{f}$. It remains to show that

Lemma 4.3. *If $E' \neq 0$ at the end of the algorithm, then $|E'| = \Omega(|E|)$.*

The meaning of this is that there is some threshold τ such that if $\operatorname{wt}(Hf) < \tau$ the algorithm must output a closeby $\tilde{f} \in C$. (This is beacause $|E_0'| \geqslant |E'| \geqslant \Omega(|E|) = \Omega(|Hf|)$ by the claim above)

References

[1] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022.

[2] Alexander Lubotzky, Roger Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. 4

[3] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022.