Locally Testable Good Codes

Alex Lubotzky

Joint work with:
Irit Dinur, Shai Evra, Ron Livne and Shahar Mozes

Error Correcting Codes (ECC)

Let $\mathbb{F}_q$ =field of order q, usually $q = 2$.

An $(n, k, d)_q$-code C is a subspace of $\mathbb{F}_q^n$ of dimension k and distance at least d, where

$$dist(C) = \min\{\text{(Hamming) dist } (\alpha, \beta) \Big| \alpha \neq \beta \in C\}$$
$$= \min\{\text{(Hamming) weight } (\alpha) \Big| 0 \neq \alpha \in C\}$$

We are interested in the case $n \to \infty$.

A code (or more precisely a family of codes) is good

if $\exists \varepsilon > 0$ s.t. $k \geq \varepsilon n$ **and** $d \geq \varepsilon n$.

The standard terminology:

$$\text{rate } = \rho(C) = \frac{k}{n} \qquad \text{(normalized) distance } = \delta(C) = \frac{d}{n}$$

and C is good if $\rho(C)$ and $\delta(C)$ are bounded below by some ε.

(Say: "Constant rate and constant distance")

Ex: 1) Random codes defined by random constraint

2) many explicit examples

LDPC codes

The code(s) is called LDPC (= low density parity check) if C is defined by equations (= linear functionals = constraints) with bounded support.

Equivalently

$$C^{\perp} = \{\beta \in \mathbb{F}_q^n \Big| < \alpha, \beta >= 0 \quad \forall \alpha \in C\}$$

is spanned by vectors of bounded support.

$$(< \alpha, \beta >= \sum_{i=1}^{n} a_i b_i).$$

Ex: Gallager: 60's Random code defined by random equations of bounded support

Explicit construction of LDPC good codes

A breakthrough by Sipser & Spielman in the 90's was the explicit construction of "Expander codes" which are LDPC good codes

Let $X = (V, E)$ be an r-regular graph. Its e.v.'s (the e.v.'s of the adjacency matrix) are in $[-r, r]$. The largest is always r and the second one $\lambda(X)$ (with $\lambda(X) < r$ iff X is connected).

Def: X is called λ-expander for some $\lambda < r$,

$$\text{if } \lambda(X) \leq \lambda$$

Again: we want $r \,\&\, \lambda$ fixed and

$$|X| \to \infty$$

Def : X is called Ramanujan graph if $\lambda(X) < 2\sqrt{r-1}$

There are explicit constructions of Ramanujan graphs:

For $r = p + 1$, p prime (Lubotzky-Phillips-Sarnak, see also Margulis)

$r = p^e + 1$ (Morgenstern)

The S-S construction

$$X = (V, E),\ r\text{-regular } \lambda\text{-expander graph on } m \text{ vertices}$$
$$r \text{ fixed}, \quad \lambda \text{ fixed}, \ m \to \infty$$

Let $C_0 \leq \mathbb{F}_2^r$ a subspace ("the small code").

Label the r edges around each vertex v by $\{1, \ldots, r\}$ (no consistency assumption).

Let

$$W = \{f : E \to \mathbb{F}_2\}$$

so:

$$n := \dim(W) = \frac{mr}{2}$$

and "the large code"

$$C = \{f \in W \,|\, f|_{\mathsf{link}(v)} \in C_0, \quad \forall v \in V\}$$

i.e. the local view of f at every vertex is in the small code

Theorem Sipser-Spielman

If

(a) $\dim C_0 > \frac{r}{2}$

(b) distance$(C_0) > \lambda(X)$

then C is an LDPC good code.

Sketch of proof:

(1) LDPC - clear !!

(2)

$$\dim(C) \geq n - m(r - \dim C_0) =$$

$$\frac{mr}{2} - mr + m \dim C_0 = m(\dim C_0 - \frac{r}{2}) =$$

$$= \underbrace{\frac{2}{r}(\dim C_0 - \frac{r}{2})}_{\text{constant.}} n$$

(3) Let $0 \neq f \in C$ and let $F = \text{support}(f)$ and G the subgraph of X spanned by F.

Lemma (Alon-Chung) If $G \subset X$ a subgraph with average degree $l > \lambda(X)$, then $|G| \geq \frac{1}{r}(l - \lambda(X))|X|$.

Here the degree of every vertex of G is $\geq$ i.e., distance $(C_0) > \lambda(X)$ hence $|G|$ is linear in $n = |X|$.

Now use Ramanujan r-regular graphs and some fixed code $C_0 \leq \mathbb{F}_2^r$
(Even if chosen by random, it is for a fixed(!) r.
But actually there are explicit). Q.E.D.

Locally testable codes (LTC)

A code $C \leq \mathbb{F}_2^n$ is a (q, ε)-locally testable code if there exists a random algorithm A which for every $\alpha \in \mathbb{F}_2^n$, A reads only q random bits of α and answers YES if $\alpha \in C$ and answers NO with probability

$$Pr(A \text{ answers NO}) \geq \varepsilon \cdot \frac{1}{n}\text{distance}(\alpha, C)$$

i.e. A can decide, by reading only a few bits of α if α is in C or far away from it.

Example:

Let $W = \{f : \mathbb{F}_2^m \to \mathbb{F}_2 | f \text{ a function}\}$

so: $n = \dim W = 2^m$

Let $C =$ the linear functionals in W

so $\dim(C) = m = \log_2(n)$.

Let A be the test: choose random $x, y \in \mathbb{F}_2^n$

and answer YES if $f(x+y) = f(x) + f(y)$

Theorem (Blum-Luby-Rubinfeld [BLR], 1993)

Prob (A answers NO) $\geq \frac{2}{27}(\frac{1}{n}\text{distance}(f, C))$

So X is $(3, \frac{2}{27})-$LTC code.

But far away from being good.

Other LTC codes:

- Reed-Solomon
- Reed-Muller
- The tensor code Let $C_1 \leq \mathbb{F}_2^{n_1}$ be a "smooth code" and $C_2 \leq \mathbb{F}_2^{n_2}$.

The tensor code $C_1 \otimes C_2 \leq \mathbb{F}^{n_1} \otimes \mathbb{F}_2^{n_2} = M_{n_1,n_2}(\mathbb{F}_2)$ is the code of the $n_1 \times n_2$ matrices over $\mathbb{F}_2$ such that each column is in C_1 and each row in C_2 (so $n = n_1 n_2$ and $k = (\dim C_1) \cdot (\dim C_2)$).

The algorithm A chooses a random row or column and check.
If $C_1 \& C_2$ good, so is $C_1 \otimes C_2$ with $q = \sqrt{n}$ (say $n_1 = n_2$).

A long standing problem

Are there good LTC with constant number of queries, i.e., q is constant independent of n?

The C^3 problem:

constant rate, constant distance and constant number of queries.

Remark

Random (LDPC) are not! Check !
So find a needle in a haystack rather than hay in a haystack, if the needle exists ...

Theorem: (I. Dinur, S. Evra, R. Livne, A. Lubotzky, S. Mozes)

Explicit construction of LTC good codes (with q constant).

"Constant rate and constant distance"
Remark : Panteleev and Kalachev have independently provided a construction of LTC good codes.

Our construction Let G be a finite group with two symmetric sets of generators A and B.

Assume:

$$1 \notin A \cup B$$

and

$$\text{(TNC)} \quad \forall a \in A,\ b \in B,\ g \in G, \quad g^{-1}ag \neq b$$

i.e., no element of A is conjugate to an element of B and for simplicity assume also $|A| = |B| = r$ and both $Cay(G; A)$ and $Cay(G; B)$ - the Cayley graphs are Ramanujan,
i.e.

$$\lambda(Cay(G; A)),\ \lambda(Cay(G; B)) < 2\sqrt{r-1}$$

Define the Left/Right Cayley Complex $Cay(A;G;B)$ by letting A act on G from the left and B from the right, giving rise to the following squares:

$$\begin{array}{ccc} gb & \xrightarrow{a} & agb \\ \uparrow{\scriptstyle b} & & \uparrow{\scriptstyle b} \\ g & \xrightarrow{a} & ag \end{array}$$

Remark (TNC) ensures that all 4 vertices are different.

There are $\frac{1}{4}|G| \cdot |A|\,|B|$ such squares.

(We take $|A|\&|B|$ fixed and $|G| \to \infty$).

Let $C_A \leq \mathbb{F}_2^A$ a smooth code, $C_B \leq \mathbb{F}_2^B$ and $C_0 = C_A \otimes C_B$

Note The squares containing a vertex g are in $1-1$ correspondence with $A \times B$.

We can define:

$$W = \{f : \{\text{squares}\} \to \mathbb{F}_2\}$$
$$C = \{f \in W | f|_{\text{link}(g)} \in C_0,\ \forall g \in G\}$$

$$= \left\{ f \in W \middle| \begin{array}{l} \forall \text{ edge } e \text{ of type } A, f|_{\text{link}(e)} \in C_B \\ \forall \text{ edge of type } B, f|_{\text{link}(e)} \in C_a \end{array} \right\}$$

Theorem

Under suitable assumption on C_A, C_B & r (large by fixed!)

This is a good LTC code.

The algorithm: Choose a random edge and check only r bits!

On HDX and LTC

The inspiration for codes on Left/Right Cayley complexes came from

HIGH DIMENSIONAL EXPANDERS (HDX)

History: Let G be a p-adic simple Lie group (e.g. $G = SL_n(\mathbb{Q}_p)$) and Γ a cocompact lattice in G. In 1972, Garland proved Serre conjecture: For every $1 \leq i < rank(G)$ (e.g. $n-1$), $H^i(\Gamma, V) = 0$ for every (finite dimensional) unitary representation of Γ.
He proved it by showing "spectral gap" for the high-dimensional Laplacians.

Namely, let $\mathcal{B}$ be the Bruhat-Tits building associated with G. This is a constructible simplicial complex of dimension $rank(G)$.
(So for $SL_n(\mathbb{Q}_p)$, this is a tree; the $(p+1)$-regular tree).

$\Gamma \setminus \mathcal{B}$ is a finite simplicial complex and Serre's conjecture can be stated as a conjecture on $\Gamma \setminus \mathcal{B}$.

Garland deduced the spectral gap for the global Laplacians of $\Gamma \setminus \mathcal{B}$ from strong spectral gaps of the links of $\Gamma \setminus \mathcal{B}$ (which are the same as of B and are well understood Spherical/Flag complexes over finite fields.

The moral lesson for us:

- Garland proved a "local to global" result
- The result means nothing to graphs!

(Moreover, the LPS - Ramanujan graphs are excellent expanders ($\equiv$ spectral gap) with large girth. But there are large girth graphs (hence locally isomorphic to the LPS's) which are not expanders.

So, "local to global" is a high-dimensional phenomenon!

$\Gamma \setminus \mathcal{B}$ above are (many times) Ramanujan complexes.

The original idea was to take Reed-Solomon codes on the links of $(r-2)$ cells (which can be identified with $\mathbb{P}^1(\mathbb{F}_p)$) and to propagate their local testability to get codes on $\Gamma \setminus \mathcal{B}$.

p-adic uniformization related $\Gamma \setminus \mathcal{B}$ when $\Gamma \leq SL_3(\mathbb{Q}_p)$ to $\Gamma^{\#} \setminus SU(2,1)/K$ when $\Gamma^{\#}$- a suitable arithmetic lattice.
We planned to use this (in the opposite direction to a famous work of Mumford on algebraic varieties) to deduce combinatorial information.

We failed and suddenly found a much simpler method for our goal of LTC codes.
But a number of interesting problems are still waiting!!