Today: Local codes. Locally decodable codes

Code: \[ C : \Sigma^k \rightarrow \Sigma^n \]

Parameters:
- rate: \( \frac{k}{n} \rightarrow \) amount of redundancy
- (rel) distance: \( \text{dist}(C) = \min \{ \delta(a_i, c_j) \} \)
  - \( c_i \neq c_j \) \( \rightarrow \) rel. Hamming distance
  - \( \downarrow \) amount of error/corruption the code can withstand.

Ideally:
- \( \frac{k}{n} = \Theta(1) \) (even \( \rightarrow 1 - o(1) \))
- \( \text{dist}(C) = \Theta(1) \)

Unique decoding: Given a corrupted version \( c(m) \) find \( m \), if fraction of error \( < \frac{\text{dist}(C)}{2} \).
Usually $\Sigma = F_2$ (also can think of $\mathbb{Z}_q$, the ring of $\mathbb{Z}$ mod. $q$).

Linear code: $C \subseteq F_q^n$ is linear if it forms a subspace (vector space) i.e. $\forall a, c_2 \in C \Rightarrow a + c_2 \in C$.

Systematic code: Codeword contains the message.

Fact: Every linear code is systematic.

Local tasks:
- Local testing: Given $z \in F_q^n$, test if $z \in C$ or is far from $C$ (with few queries).
- Local decoding: Given $z \in F_q^n$ and $i \in [k]$ if $z$ is close to $C(m)$ find $m_i$.
- Local correction: Given $i \in [k]$, find $C(m)_i$. 
Locally decodable/correctable code:

Given \( C : \Sigma^k \rightarrow \Sigma^n \), \( C \) is \((g, \delta, \varepsilon)\)-locally decodable if \( \forall i \in [k] \) \( \exists \) rand alg \( A_i \) st. \( A_i \) makes \( g \) queries to \( x \) \( \forall \) msg \( x \in \Sigma^k \), \( \forall \) \( z \in \Sigma^n \) with \( \text{dist}(z, C(x)) \leq \delta n \) then

\[
\Pr \left[ A_i(z) = C(x)_i \right] > 1 - \varepsilon.
\]

\( x \rightarrow C(x) = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix} \)

\( z = \begin{bmatrix} 0 & 1 & 1 & 1 \end{bmatrix} \)

\( i \in [k] \)

\( g \) queries

\( A_i \)

\( x_i \)

\( C \) is \((g, \delta, \varepsilon)\)-locally decodable if \( \forall i \in [n] \)

\[
\Pr \left[ A_i(z) = C(x)_i \right] > 1 - \varepsilon.
\]
Fact: Every linear locally correctable code is locally decodable. (Since it is systematic.)

Example: Hadamard code.

- Linear code
  - \( l_a(x) = a \cdot x \mod 2 \)

\[
\text{Had} = \sum \frac{1}{2} l_a^\mathbf{H}_2^k \rightarrow \mathbb{F}_2^2 \quad a \in \mathbb{F}_2^k
\]

msg: \( a \in \mathbb{F}_2^k \)  
index by \( x \): 00 01 10 11  
\( 2^k \)  
a

<table>
<thead>
<tr>
<th>l</th>
<th>( \langle 0, 0, 0, 0, 0 \rangle )</th>
<th>00</th>
</tr>
</thead>
<tbody>
<tr>
<td>l0</td>
<td>( \langle 0, 1, 1, 1, 0 \rangle )</td>
<td>01</td>
</tr>
<tr>
<td>l01</td>
<td>( \langle 0, 1, 1, 0, 1 \rangle )</td>
<td>01</td>
</tr>
<tr>
<td>l011</td>
<td>( \langle 0, 1, 1, 0, 1 \rangle )</td>
<td>11</td>
</tr>
</tbody>
</table>
Thm: Had is $(3, 5, 25) \text{ - locally dec. (LDC)}$

\[ \forall i \in [k], x \in \mathbb{F}_2^k \rightarrow \text{decode } m_i \text{ output } l_a(x) \]

Alg: on $x$ and $z \in \mathbb{F}_2^n$

\[ n = 2^k \]

\[ \text{think of } z \text{ as } \]

\[ f: \mathbb{F}_2^k \rightarrow \mathbb{F}_2, \]

\[ s.t. \text{ dist } (f, l_a) \leq 5n \]

\[ z \in \text{ Had} \]

Pick $b$ uniform at rand. in $\mathbb{F}_2^k$

Output $f(b) + f(b + x)$

Recall: $l_a(b) + l_a(f + b) = l_a(x)$

What we want to output
Analysis:

Since $b$ is m.a.r. in $\mathbb{F}_2^k$

$$\Pr[ f(b) \neq \lambda_a(b) ] \leq \delta.$$ 

Also, for fixed $x$, $b$ m.a.r.

$x + b$ is m.m.f. at rand. in $\mathbb{F}_2^k$, so

$$\Pr[ f(x + b) \neq \lambda_a(x + b) ] \leq \delta.$$ 

By union bound, at least one of bad events happen w.p. $2 \delta$.

So none hold w.p. $1 - 2 \delta$.

When that is the case, we have

$$f(b) = \lambda_a(b)$$

$$f(x+b) = \lambda_a(x+b)$$

hence $f(b) + f(x+b) = \lambda_a(b) + \lambda_a(x+b) = \lambda_a(x)$.
Remark: Had is 2-query Loc. dec.

- rate = \( \frac{k}{2^k} \to 0 \)

- distance: \( \frac{1}{2} \)

Can get better 2-query LDC?

- No: every 2-query LDC has \( n = 2 \frac{8}{k} \) (GKST02, Kremeridis, deWolf)

Constant \( > 2 \) queries

UB: subexponential (Yekhanin, others)

16: \( n > k + O(1/2) \) \( \# \) of queries.

(Woodruff 07) Slightly super linear.

0Q: \( O(\log n) \) queries for \( n = \Omega(k) \)?

- tighter constant UBs for d-query LDCs
Reed Muller codes:

\[ \# \text{vars} \leq \deg \]

\[ \text{Def: } m, l \text{ and } \mathbb{F}_q, \quad l \leq q-1 \]

\[ \text{RM}(m, l) = \{ \langle p(x) \rangle \mid p \in \mathbb{F}_q[x_1, \ldots, x_m], \deg p \leq l \} \]

If \( m = 1 \), \( \text{RM}(1, l) \) is a Reed Solomon code.

Example: \( \text{RS: } m = 1 \)

\[ \# p \quad \text{RS}(l) = \{ \langle p(x) \rangle \mid p \in \mathbb{F}_q, \deg p \leq l \} \]

eg: \( p = 2x^5 + 10 \) in \( \mathbb{F}_11 \); \( l \leq 6 \)

codeword \( p: \langle p(0), p(1), \ldots, p(10) \rangle \)

\( \text{RS}(l) : \text{ take all codewords gen by every } \)

\( p \in \mathbb{F}_q[x_1] \) of \( \deg \leq 6 \).
message \( (0, 2, 0, 0, 0, 0, 10) \)

Coefficients of \( p \)

\[ K = l + 1 \leq \# \text{ of coefficients} \]

\[ n = \left| \mathbb{F}_q^l \right| = q \]

\[ \frac{k}{n} = \frac{l + 1}{q} \quad \text{(rate of when)} \]

\[ l = \Theta \left( \frac{q}{l} \right) \]

\( RM: \) \( 2 \) vars in \( \mathbb{F}_q^{l} [x_1, x_2] \) \( \text{deg} \leq q \)

\[ p(x_1, x_2) = 3x_1^4 + 5x_1^3x_2 + x_2^3 + 6x_1^2x_2 \]

indexing by \( \mathbb{F}_q^2 \) \( : (0,0)(0,1) \ldots (10,10) \)

Encoding of \( p \): \( \langle p(00), p(01) \ldots p(10,10) \rangle \)

\[ K: \# \text{ of monomials in } x_1, x_2 \text{ of } \text{deg} \leq l. \]

In general

\[ x_1^{a_1} x_2^{a_2} \ldots x_m^{a_m} \quad \text{st. } \sum a_i = l \]

\[ l < \frac{q}{2} \]
\[ l, l_2 \quad \frac{\partial f}{\partial x_m} \quad (l + m) - (l + m) \]

\[ K = O \left( \frac{m + e}{m} \right) \]

\[ n = \left| \mathbb{F}_q^m \right| = q^m \]

distance of RM: \((1 - \frac{e}{q})^m \)

**Proof:** Schwartz-Zippel Lemma: Let \( p \in \mathbb{F}_q[x_1, \ldots, x_m] \) of total deg \( l \). Then \# of \( x \in \mathbb{F}_q[x_1, \ldots, x_m] \) s.t. \( p(x) = 0 \) is \( \leq l q^{m-1} \)

**Proof:** By induction on \# of vars.

So, \( \Pr \left[ f(x) = 0 \right] \leq \frac{e}{q} \)

\[ x \leftarrow \text{var} \quad \mathbb{F}_q^m \]

Assume \( e \leq \frac{q}{2} \).
Distance of RM: 2 codewords are evuls of 2 polys over $\mathbb{F}_q \rightarrow (\mathbb{Q}^*)$

$\langle p(000), p(001) \rangle$

$\langle 2(000), 2(010) \rangle$

$\text{dist} (p, q) = \# \text{ non-0s} / Q^m$

$\geq (Q^m - \ell Q^{m-1}) / Q^m$

$= \left(1 - \ell / Q\right)$

---

Thm. $RM (\ell, m) \text{ is } (\ell + 1, \frac{1}{3(\ell + 1)}, \frac{1}{3}) \text{ LD C}$. 

---

# of queries

dist that can with $T$ depend on $\ell$. 

confid.
Def: A line in $\mathbb{F}_q^m$ is def by $a \in \mathbb{F}_q^m$, $b \in \mathbb{F}_q^m / \{0\}$.

\[ L_{a,b} = \{ a + bt \mid t \in \mathbb{F}_q \} \]

Lem: Given $a \in \mathbb{F}_q^m$, $t \in \mathbb{F}_q^*$. For $b \in \mathbb{F}_q^m / \{0\}$, $a + bt$ is uniform in $\mathbb{F}_q^m$. 
Lem: If \( p \in \mathbb{F}_q[x_1, \ldots, x_m] \) of degree \( n \),

Then \( P_L = p \circ L \) (the restriction of \( p \) to line \( L \)) is a poly in one var of degree \( n \).

Eg: \( p(x, x_2) = x_1^3 x_2 \) in \( \mathbb{F}_4 \rightarrow \deg 2 \)

\( L = (3, 2) + t(5, 1) = (3 + 5t, 2 + t) \)

\( p \circ L : \quad p(3+5t, 2+t) = (3+5t)^3 (2+t) \)

\( \uparrow \) univ. int.

\( \cdot \deg 4 \)

(To be continued)