Today: Local codes. Locally decodable codes

Code: $C : \Sigma^k \to \Sigma^n$

$m \xrightarrow[\text{channel adds error}]{C(m)} C(m)+E$ (error vector)

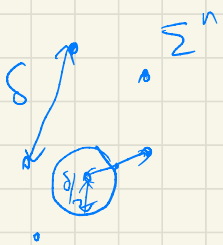
$m \in \Sigma^k$ is a message

$C(m) \in \Sigma^n$ is a codeword

Parameters

rate: $\frac{k}{n}$ $\Rightarrow$ amount of redundancy

(rel) distance: $\text{dist}(C) = \min_{c_i \neq c_j \in C} \{\delta(c_i, c_j)\}$ ← rel. Hamming distance



$\Downarrow$ amount of error/corruption the code can withstand.

Ideally: $\frac{k}{n} = \Theta(1)$ (even $\to 1 - o(1)$)

$\text{dist}(C) = \Theta(1)$

Unique decoding: Given a corrupted version of $C(m)$ find m, if fract of error $< \frac{\text{dist}(C)}{2}$.

Usually $\Sigma = \mathbb{F}_q$ (also can think of $\mathbb{Z}_q$, the ring of $\mathbb{Z}$ mod q)

Linear code: $C \subseteq \mathbb{F}_q^n$ is linear if it forms a subspace (vectorspace) i.e.
$\forall\ c_1, c_2 \in C \Rightarrow c_1 + c_2 \in C$.

Systematic code: Codeword contains the message.

Fact: Every linear code is systematic.

Local tasks:

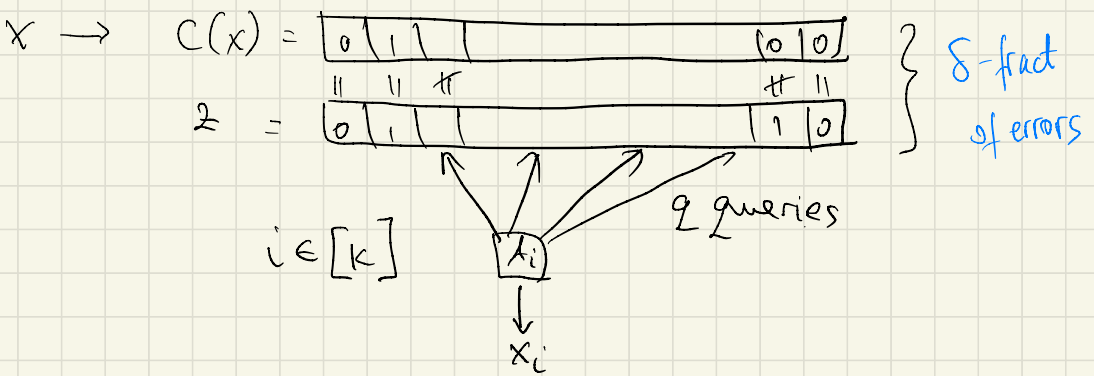
- local testing: Given $z \in \mathbb{F}_q^n$, test if $z \in C$ or is far from C (with few queries)
- local decoding: Given $z \in \mathbb{F}_q^n$ and $i \in [k]$ if z is close to $C(m)$ find m_i
- Local correction: — $i \in [n]$, find $C(m)_i$.

Locally decodable / correctable code:

Given $C: \Sigma^k \to \Sigma^n$

- C is (q, δ, ε)-locally decodable if $\forall i \in [k]$ $\exists$ rand alg A_i s.t. A_i makes q queries to z & $\forall$ msg $x \in \Sigma^k$, & $z \in \Sigma^n$ with $dist(z, C(x)) \leq \delta n$ then

$$Pr[A_i(z) = x_i] \geq 1 - \varepsilon$$

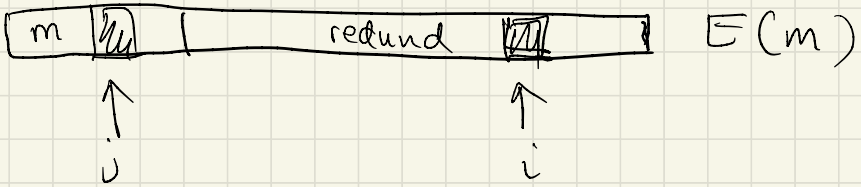


- C is (q, δ, ε)-locally correctable if $\forall i \in [n]$

$\forall$ ——————

$$Pr[A_i(z) = C(x)_i] \geq 1 - \varepsilon$$

Fact: Every lin locally correctable code is locally decodable. (since it is systematic)



Example: Hadamard cod.

- linear code
- $l_a(x) = a \cdot x \bmod 2$

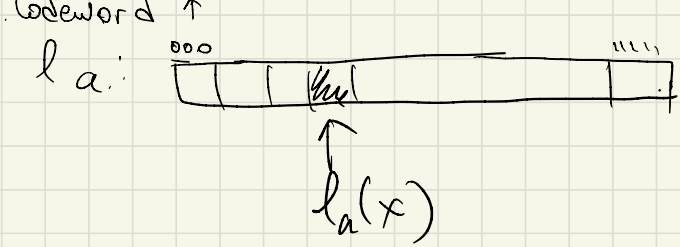
$a, x \in \mathbb{F}_2^k$

$$Had = \{ l_a : \mathbb{F}_2^k \to \mathbb{F}_2 \}_{a \in \mathbb{F}_2^k}$$

msg: $a \in \mathbb{F}_2^k$

codewords indexed by x:

	00	01	10	11	2^k	a
l_{00}:	0	0	0	0		00
l_{01}:	0	1	0	1		01
						10
						11



Thm: Had is $(3, \delta, 2\delta)$-locally dec. (LDC)

Pf: $\forall i \in [k]$ ⟶ decode m_i

Here $x \in \mathbb{F}_2^k$ ⟶ output $l_a(x)$

Alg: on x and $z \in \mathbb{F}_2^n$ $\quad n = 2^k$

$\updownarrow$ think of z as

$f: \mathbb{F}_2^k \to \mathbb{F}_2$.

s.t. $dist(f, l_a) \le \delta n$

$\uparrow \in$ Had

Pick b unif 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 u.a.r in $\mathbb{F}_2^k$

$$\Pr\left[f(b) \neq l_a(b) \right] \leq \delta$$

Also for fixed x, b u.a.r.

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

$$\Pr\left[f(x+b) \neq l_a(x+b) \right] \leq \delta$$

By union bound, at least one of bad events happen w.p 2δ.

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

when that is the case we have

$$f(b) = l_a(b)$$

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

✓

hence $f(b) + f(x+b) = l_a(b) + l_a(x+b) = l_a(x)$

Remark: Had is 2-query loc. dec.

- rate $= k/2^k \to 0$
- distance: $1/2$

Can get better 2-query LDC?

- NO: every 2 query LDC has $n = 2^{O(k)}$

(GKST02, Kerenidis deWolf)

Constant > 2 queries.

ub: subexponential (Yekhanin, others)

lb: $n > k^{1+O(1/q)}$ ← # of queries.

(Woodruff 07) ↑ slightly super linear.

OQs: $O(\log n)$ queries for $n = O(k)$?

- tighter construct / lbs for cl. query LDCs

Reed Muller codes:

Def: m (# vars), ℓ (deg) and $\mathbb{F}_q$ $\quad \ell \leq q-1$

$$RM(m,\ell) = \left\{ \langle p(\alpha) \rangle_{\alpha \in \mathbb{F}_q^m} \;\middle|\; p \in \mathbb{F}_q[x_1, \ldots x_m], \deg p \leq \ell \right\}$$

If $m = 1$, $RM(1,\ell)$ is a Reed Solomon code.

Example: RS: $m = 1$

$\mathbb{F}_p$ $\quad RS(\ell) = \{ \langle p(\alpha) \rangle_{\alpha \in \mathbb{F}_p} \mid \deg p \leq \ell \}$.

eg: $p = 2x^5 + 10$ in $\mathbb{F}_{11}$; $\ell \leq 6$

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

$RS(\ell)$: take all codewords gen by every $p \in \mathbb{F}_{11}[x]$ of deg ≤ 6.

index by deg 6 5 4 3 2 1 0

message $(0, 2, 0, 0, 0, 0, 10)$

↑ coefficients of p

$k = \ell + 1$ ← # of coefficients

$n = |\mathbb{F}_Q| = Q$

$\frac{k}{n} = \frac{\ell+1}{Q}$ (rate of when $\ell = \Theta(Q)$)

RM: 2 vars in $\mathbb{F}_{11}[x_1, x_2]$, deg ≤ 9

$$p(x_1, x_2) = 3x_1^4 + 5x_1x_2^3 + x_2^9 + 6x_1^2x_2$$

indexing by $\mathbb{F}_{11}^2$: $(0,0)(0,1)$ ... $(10,10)$

Encoding of p : $\langle p(0,0), p(0,1), \ldots, p(10,10) \rangle$

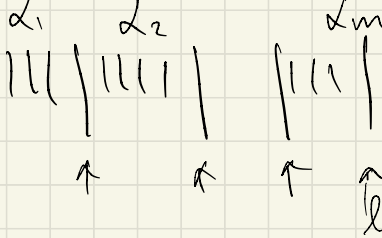
k: # of monomials in x_1, x_2 of deg $\leq \ell$.

$x_1 \ldots x_m$

In general

$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_m^{\alpha_m}$ st. $\sum \alpha_i = \ell$

$\ell < Q/2$



$$\binom{\ell+m}{m} = \binom{\ell+m}{\ell}$$

$$k = O\binom{m+\ell}{m}$$

$$n = |\mathbb{F}_Q^m| = Q^m$$

distance of RM: $\left(1 - \ell/Q\right) Q^{m-1}$

Pf: Schwartz-Zippel Lemma: Let $p \in \mathbb{F}_Q[x_1 \cdots x_m]$ of total deg ℓ. Then # of $x \in \mathbb{F}_Q[x_1 \cdots x_m]$ s.t. $P(x) = 0$ is $\leq \ell Q^{m-1}$

Pf: By induction on # of vars.

So $\Pr_{x \xleftarrow{u.a.r} \mathbb{F}_Q^m}\left[f(x) = 0 \right] \leq \ell/Q$.

Assume $\ell < Q/2$.

Distance of RM: 2 codewords are evals of 2 pdys over $\mathbb{F}_Q^m \to$ (Q prime)

$\langle p(000\ldots), p(00\ldots1), \ldots, p(Q-1, Q-1, \ldots Q-1) \rangle$

$\langle q(00\ldots0), q(0,0,\ldots1), \ldots, q(Q-1, Q-1, \ldots Q-1) \rangle$

$$\text{dist}(P, Q) = \#\ \text{non-zeros} / Q^m$$

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

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

Thm: $RM_Q(\ell, m)$ is $\left(\ell+1, \frac{1}{3(\ell+1)}, \frac{1}{3}\right)$ LDC.

- $\ell+1$: # of queries
- $\frac{1}{3(\ell+1)}$: dist that can with ↑ dep on ℓ.
- $\frac{1}{3}$: confid.

Def: A line in $\mathbb{F}_q^m$ is def by

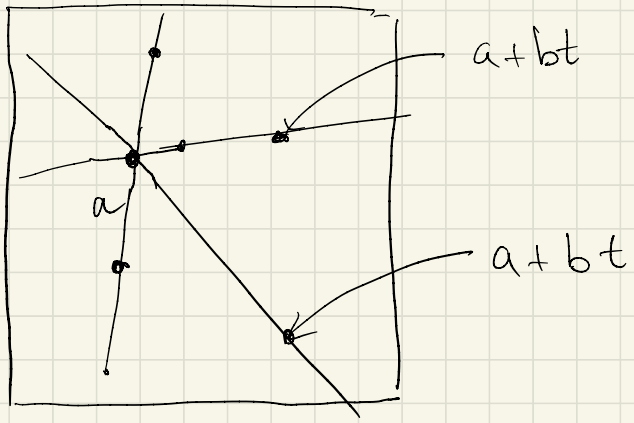
$$a \in \mathbb{F}_q^m, \quad b \in \mathbb{F}_q^m / \{0\}.$$

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

Lem: Given $a \in \mathbb{F}_q^m$, $t \in \mathbb{F}_q^*$

For $b \xleftarrow{u.a.r.} \mathbb{F}_q^m / \{0\}$.

$a + bt$ is uniform in $\mathbb{F}_q^m$



Lem: If $p \in \mathbb{F}_Q[x_1, \dots x_m]$ of deg ℓ

Then $p_L = p \circ L$ (the restriction of p to line L) is a poly in one var of deg $\leq \ell$

Eg: $p(x_1, x_2) = x_1^3 x_2$ in $\mathbb{F}_7$ → deg = 4

$L = (\underbrace{3}_{a}, 2) + t(\underbrace{5}_{b}, 1) = (3+5t, 2+t)$

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

- ↑ univ. int.
- deg 4

(To be continued)