Local Decoding

Defn: Let $E:\{0,1\}^n \to \{0,1\}^m$ be an ecc & $\rho \in (0,1)$.
A local decoder for E handling ρ errors is an algorithm: ← short: Ld$_\rho$

Given $j \in [n]$ & oracle to $y \in \{0,1\}^m$ st. $\Delta(y, E(x)) < \rho$: Output x_j with probability $\geq 2/3$ in polylog m-time.

(⇒ when m is large, very few bits of y are used to guess x_j !)

Theorem 1: $\forall p < 1/4$, WH-code has Ld_p.

Proof: Idea – Query positions z & $z+e_j$ in y.
Output $y(z)+y(z+e_j)$.

$\uparrow$ j-th elementary vector

Input: $j \in [n]$, Oracle $f: \{0,1\}^n \to \{0,1\}$ s.t.

$$\Pr_{z}[f(z) \neq x \odot z] \leq p < 1/4.$$

$\uparrow$ corrupted $E(x)$ $\uparrow$ unknown

Output: $b \in \{0,1\}$

· **Decoder:** 1) Randomly pick $z \in \{0,1\}^n$.
2) Let $e_j \in \{0,1\}^n$ be the string with 1 at j-th place & 0 in the rest.
3) Output $f(z)+f(z+e_j) \bmod 2$.

▷ Time is $\text{poly}(n) = \text{poly}(\log m)$. $[m := |y| = 2^n.]$

• Consider $\Pr_z[f(z) = x \odot z \wedge f(z+e_j) = x \odot (z+e_j)]$

$\geq 1 - 2p > 1/2$.

$\Rightarrow \Pr_z[f(z) + f(z+e_j) \equiv x \odot e_j \mod 2] > 1/2$.

$\Rightarrow \Pr_z[b = x_j] > 1/2$.

$\Rightarrow$ Repeating the experiment, boosts the prob. to $\geq 2/3$. $\square$

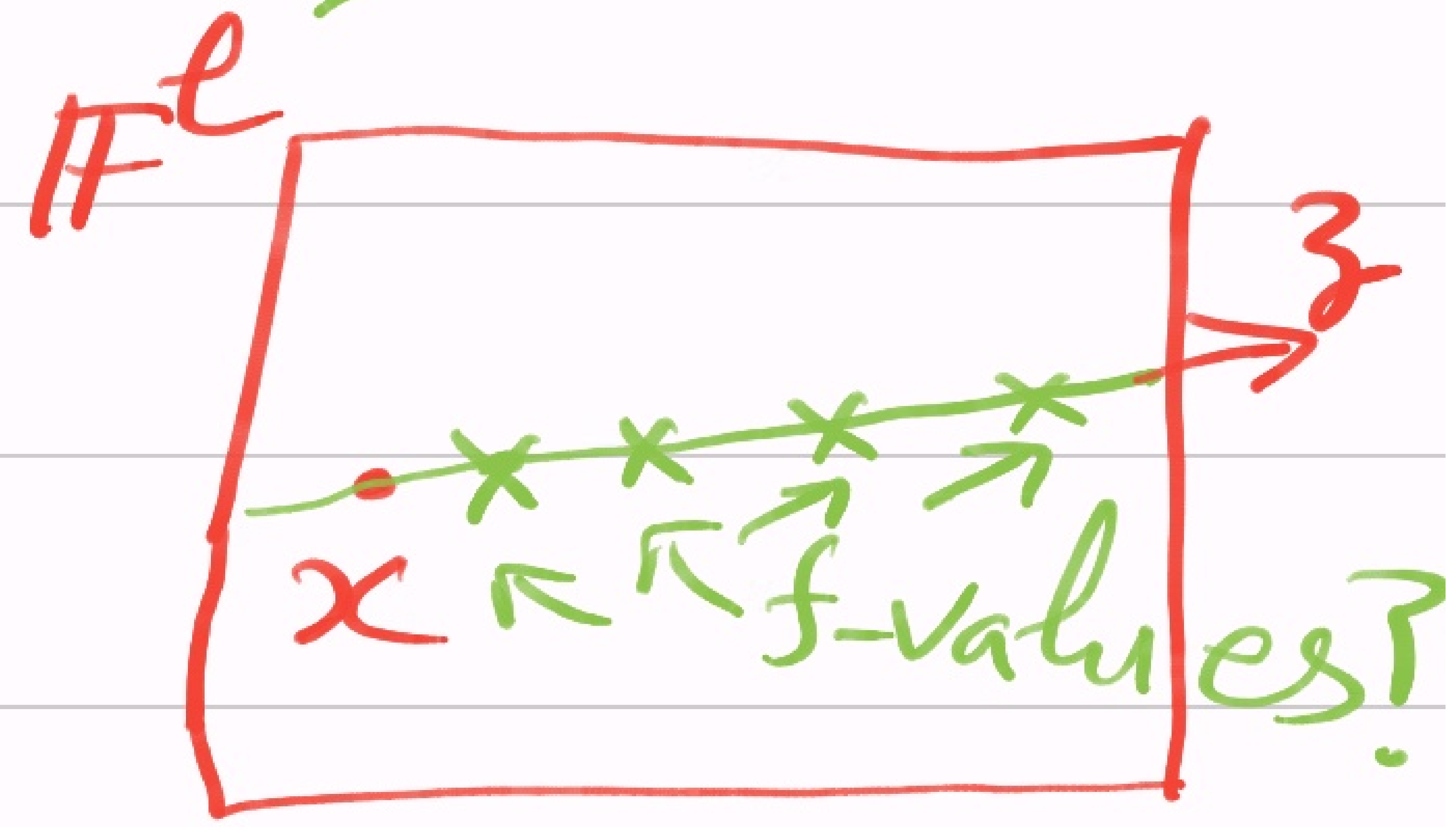
Local Decoder for RM

- Recall RM: $\mathbb{F}^{\binom{\ell+d}{d}} \longrightarrow \mathbb{F}^{|\mathbb{F}|^\ell}$ is of distance $1-\frac{d}{|\mathbb{F}|}$,
  where $d < |\mathbb{F}| < \infty$.
  non-binary
- View RM as mapping $\binom{\ell+d}{d}$ evaluations of a polynomial f to its $|\mathbb{F}|^\ell$ many evaluations.
  [$\binom{\ell+d}{d}$ evals uniquely specify deg-d ℓ-var f.]

Thm 2: $\forall \rho \le \frac{1}{6}\cdot\left(1-\frac{d+5}{|\mathbb{F}|-1}\right)$, RM-code has LDP.

Proof: • Deg-d polynomial f is unknown; a point $x \in \mathbb{F}^\ell$ is given in the input. We want $f(x) \in \mathbb{F}$.

Idea – Pick a random line L_x through x, evaluate f on each point in L_x. Use RS-decoder to learn $f|_{L_x}$.
$\Rightarrow$ Output $f|_{L_x}(x) = f(x)$.



- Input: $x \in \mathbb{F}^\ell$; oracle $\tilde{f}: \mathbb{F}^\ell \to \mathbb{F}$ that agrees with an unknown ℓ-var d-deg f on $\geq 1-\rho$ points.
- Output: $\alpha \in \mathbb{F}$.
- Decoder: 1) Pick random $z \in \mathbb{F}^\ell$ & define line $L_x := \{x + tz \mid t \in \mathbb{F}\}$.
  2) Query $\tilde{f}$ on L_x. I.e. collect the pairs

$\{(t, \tilde{f}(x+tz)) \mid t \in \mathbb{F}\} =: \tilde{f}(\underline{L_x})$.

3) Via RS-decoder, on $\tilde{f}(L_x)$, find a deg$\leq d$ 1-var polynomial $\tilde{Q}: \mathbb{F} \to \mathbb{F}$ s.t.
$\tilde{Q}(t) = \tilde{f}(x+tz)$ for the <u>largest</u> number of t's.

4) Output $\tilde{Q}(0)$.

▷ Time is $\mathrm{poly}(\ell, d, |\mathbb{F}|) = \mathrm{polylog}(|\mathbb{F}|^{\ell})$.

<u>Analysis</u>:

- RS-decoder tries to reconstruct $\underline{Q(t)} := f(x+tz)$.
- Consider $\Pr_z\left[\#t, \text{ with } Q(t) \neq \tilde{f}(x+tz), \text{ is } < \frac{|\mathbb{F}|-d}{2}\right]$

<u>Qn</u>: $\geq 2/3$?

$\frac{1}{2} \times$ RS-distance

- To show that we consider $\underline{\text{expectation}}$:

$$\mathrm{Exp}_z[\#\{t \in \mathbb{F} \mid f(x+tz) \neq \tilde{f}(x+tz)\}] \leq$$

$$1 + \sum_{t \in \mathbb{F}\setminus\{0\}} \Pr_z[f(\underbrace{x+tz}_{\text{rnd pt. in } \mathbb{F}^\ell}) \neq \tilde{f}(x+tz)] \leq 1 + \rho(|\mathbb{F}|-1).$$

$\Rightarrow$ (By $\underline{\text{Markov's}}$ inequality) $\Pr_z\left[\#\{t \in \mathbb{F} \mid Q(t) \neq \tilde{Q}(t)\} \geq \frac{|\mathbb{F}|-d}{2}\right]$

$$\leq \frac{1+\rho\cdot(|\mathbb{F}|-1)}{(|\mathbb{F}|-d)/2} \leq \frac{1+\frac{1}{6}(|\mathbb{F}|-d-6)}{(|\mathbb{F}|-d)/2} = 1/3.$$

$\Rightarrow$ With $\Pr_z \geq 2/3$, step-3 gets $\tilde{Q} = Q = f(x+tz)$.

$\Rightarrow$ $\tilde{Q}(0) = f(x)$ whp. $\square$

Local Decoder for Concatenated Codes

- Let $E_1 : \{0,1\}^n \to \Sigma^m$ resp. $E_2 : \Sigma \to \{0,1\}^k$ be ecc with local decoders of q_1 resp. q_2 queries, handling ρ_1 resp. ρ_2 errors.

[Like RM, assume $q_1 \geq |\Sigma|$.]

Thm 3: Ecc $E := E_2 \circ E_1 : \{0,1\}^n \to \{0,1\}^{mk}$ has $Ld_{\rho_1 \rho_2}$ with queries $O(q_1 \lg q_1 \cdot q_2 \cdot \lg|\Sigma|)$.

Proof: Idea - Given $y \in \{0,1\}^{mk}$, break it into blocks of size k. On a block: call E_2-Ld_{ρ_2} $(\lg|\Sigma|)$-times. Finally, call E_1-Ld_{ρ_1} on several decoded "blocks".

- Input: Index $i \in [n]$; oracle access to $y \in \{0,1\}^{mk}$ s.t. $\exists x \in \{0,1\}^n$, $\Delta(y, E_2 \circ E_1(x)) < \rho_1 \rho_2$.
- Output: $b \in \{0,1\}$.
- Decoder: 1) View y as m blocks each of k-bits.

  [It's a corrupted version of $\langle E_2(E_1(x)_j) \mid j \in [m] \rangle$.]

  2) Call E_2-Ld_{ρ_2} on the j-th block of y. Do this $\lg|\Sigma|$ times to "recover" $E_1(x)_j$.

  3) Repeat this $50 \cdot \lg q_1$ times so that the prob. of not decoding $E_1(x)_j$ is $< 1/10q_1$.

  [We need this, because E_1-Ld_{ρ_1} will need q_1-many j's.]
  [j's are picked by Ld_{ρ_1} algo.]

4) Use E_1's Ld_{ρ_1} to q_1 blocks (i.e. q_1-many j's). The answers are <u>consistent</u> with that of a string that is ρ_1-close to $E(x) = E_2 \circ E_1(x)$ with probability $> 1 - \frac{1}{10q_1} \times q_1 = 0.9$. [Since $< \rho_1$ of the blocks in y can be at distance $> \rho_2$ from the respective <u>true</u> block.]

$\Rightarrow$ E_1's Ld_{ρ_1} outputs x_j with prob. $\geq 0.9 - 1/3 > 1/2$.

& #queries $= O(q_2 \cdot \lg|\Sigma| \cdot \lg q_1 \cdot q_1)$. $\square$

Corollary: For WHoRM local decoder the #queries $= O(q \cdot \lg^2 q) = \tilde{O}(q)$ handling up to $\frac{1}{6} \cdot \left(1 - \frac{d+5}{q-1}\right) \cdot \frac{1}{4}$ errors, where $q := |\mathbb{F}|$.

[msg-length $\approx \binom{d+\ell}{\ell}$ & codelength $\approx q^{\ell}$
& errors $\approx 5\%$]

- Our final goal is to show: If f is a worst-case hard fn. & E is a l.d. code, then $E \circ tt(f) =: tt(g)$ gives an average-case hard fn. (hard on $\frac{1}{2} - \delta$ inputs?)

$\Rightarrow$ We need an E that is locally decodable up to $(\frac{1}{2}-\delta)$-errors!

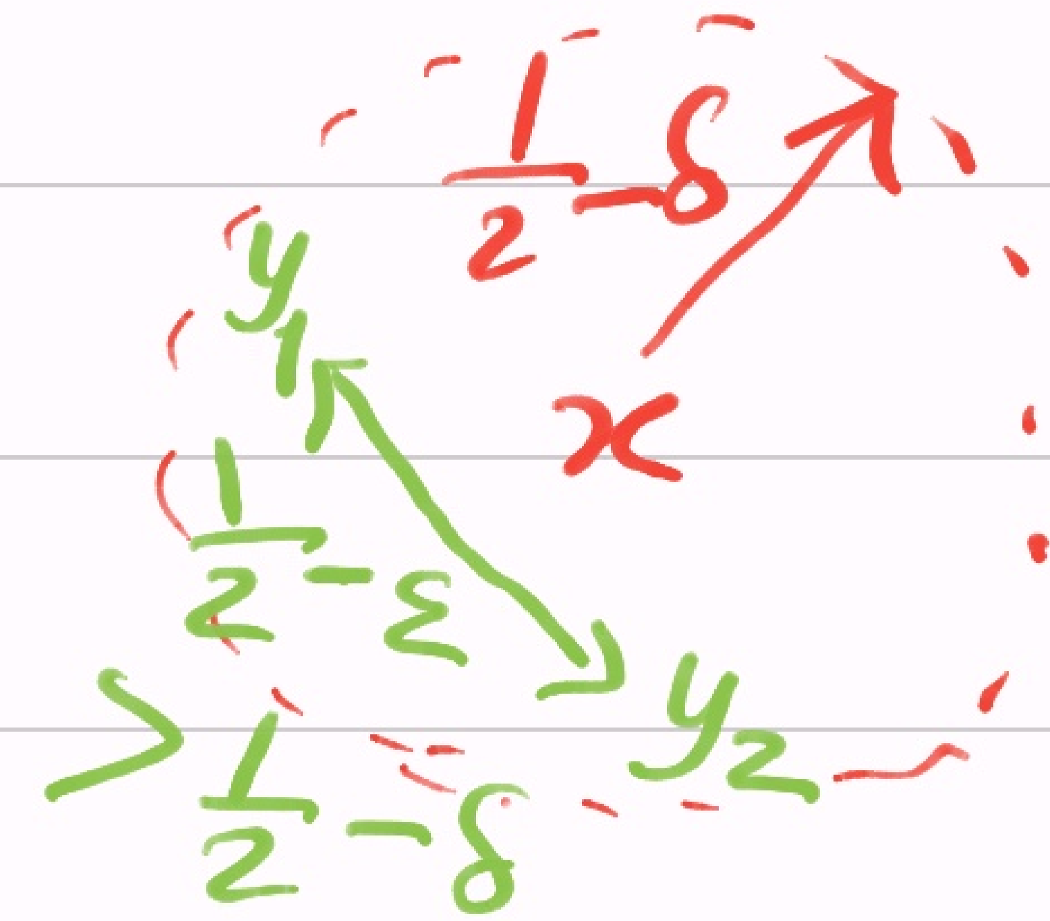
This decodability is <u>not</u> unique.

- So, we relax unique decodability to that of finding a <u>list</u>.

<u>Thm</u> (Johnson bound '62): If $E:\{0,1\}^n \to \{0,1\}^m$ is an ecc with distance $\geq (1/2-\varepsilon)$ then $\forall x \in \{0,1\}^m$ & $0<\delta \geq \sqrt{\varepsilon}$, $\exists (\leq 1/2\delta^2)$-many codewords $y_1, \ldots, y_\ell$ s.t.

$$\Delta(x, y_i) \leq 1/2 - \delta \,, \quad \forall i \in [\ell] .$$

Proof: • Intuitively we cannot "pack" many y_i's inside the ball. We analyze using "inner-product".



• Define $z_1, \cdots, z_\ell \in \{-1,1\}^m$ s.t.

$$z_{i,k} := \begin{cases} 1, & \text{if } y_{i,k} = x_k \\ -1, & \text{else} \end{cases}$$

• Since $\Delta(x, y_i) \leq 1/2 - \delta \Rightarrow$

$$\sum_{k=1}^{m} z_{i,k} \geq \left(\tfrac{1}{2}+\delta\right)m - \left(\tfrac{1}{2}-\delta\right)m = 2\delta m \quad \text{---(1)}$$

• Since $\Delta(y_i, y_j) \geq \frac{1}{2} - \varepsilon \quad (i \neq j) \Rightarrow$

$$\langle z_i, z_j \rangle = \sum_{k=1}^{m} z_{i,k} \cdot z_{j,k} \leq \left(\tfrac{1}{2}+\varepsilon\right)m - \left(\tfrac{1}{2}-\varepsilon\right)m = 2\varepsilon m \quad \text{---(2)}$$

- Let $w := \sum_{i=1}^{\ell} z_i$. So, $\langle w,w\rangle = \sum_{i=1}^{\ell}\langle z_i,z_i\rangle + \sum_{i\neq j}\langle z_i,z_j\rangle$

$$\leq \sum_{i=1}^{\ell} m + \sum_{i\neq j} 2\varepsilon m \leq \ell m + 2\delta^2\cdot\ell^2 m \quad \text{--- (3)}$$

- Also, by eqn.(1): $\sum_{k=1}^{m} w_k = \sum_{k\in[m],\, i\in[\ell]} z_{i,k} \geq 2\delta m\cdot\ell$ --- (4)

- By Cauchy-Schwarz's: $\sum_{k=1}^{m} w_k^2 \geq \left(\sum w_k\right)^2/m$

$$\Rightarrow \langle w,w\rangle \geq (2\delta m\ell)^2/m = 4\delta^2\ell^2\cdot m.$$

- Combining with eqn.(3):

$$4\delta^2\ell^2\cdot m \leq \langle w,w\rangle \leq \ell m + 2\delta^2\ell^2\cdot m$$

$$\Rightarrow 2\delta^2\ell \leq 1 \Rightarrow \qquad \ell \leq 1/2\delta^2 \leq 1/2\varepsilon. \qquad \square$$

- Can we compute the list efficiently? locally?
  - The answers are YES!

List Decoding RS

Thm (Sudan '95): ∃ randomized poly-time algo. that given $\{(a_i, b_i) \in \mathbb{F}^2 \mid i \in [m]\}$ returns the list of all deg $\leq d$ polynomials $G(x)$ s.t.

$\#\{i \in [m] \mid G(a_i) = b_i\} > \sqrt{2dm}$.

[I.e. for distance $> 1 - \frac{d-1}{m}$, the list-decoder handles $< 1 - \sqrt{\frac{2d}{m}}$ errors!] ← non-binary Johnson's bound.

Proof: Idea – Use a bivariate auxiliary polynomial $Q(X,Y)$ to fit the data. Factor Q!

1) Compute a nonzero $Q \in \mathbb{F}[X,Y]$ s.t. $Q(a_i, b_i) = 0$, $\forall i \in [m]$, where $(1,d)\text{-wt.deg}(Q) \leq \sqrt{2dm} =: t$.

$:= \max\{i + dj \mid \text{monomial } x^i y^j \text{ in support of } Q\}$.

[$\Rightarrow$ # monomials in Q $= \sum_{0 \leq j \leq t/d} (1 + t - dj)$

$= (1+t)(1+\lfloor t/d \rfloor) - \frac{d}{2} \cdot \lfloor t/d \rfloor (\lfloor t/d \rfloor + 1) = (1 + \lfloor t/d \rfloor) \cdot (1 + t - \frac{d}{2} \lfloor t/d \rfloor)$

$\geq \; t/d \cdot (1 + t/2) = t/d + t^2/2d > m.$]

$\Rightarrow$ # unknowns > # eqns. in this homogenous linear-system.

[$\Rightarrow$ Step 1 finds a $Q(X,Y)$.]

2) Factor Q (using an efficient bivariate factoring algo. over finite field $\mathbb{F}$).

3) For factors of the form $\underline{Y - P(X)}$, $\deg P \le d$ & $\#\{i \in [m] \mid P(a_i) = b_i\} > t$, OUTPUT $P(X)$.

[Any $\underline{\deg \le d\ G(X)\ \text{that}}$ "$\underline{\text{fits}" > t}$-points $\Longrightarrow$

$$\begin{cases} Q(X, G(X)) = 0 \text{ on } > t\text{-distinct } a_i\text{'s}. \\ \deg Q(X, G(X)) \le \text{wt.deg}(Q) \le t. \end{cases}$$

$\Rightarrow\ Q(X, G(X)) = 0 \Rightarrow \underline{(Y - G(X)) \mid Q(X,Y)}$.]

$\square$

Corollary: RS, of distance $1-\frac{d-1}{m}$, has a list decoder handling $< 1-\sqrt{\frac{2d}{m}}$ errors, outputs list-size $\le \sqrt{2m/d}$.

Pf: $\deg_Y Q \le t/d = \sqrt{2m/d}$. $\square$

Local List Decoding

Defn: Let $E:\{0,1\}^n \to \{0,1\}^m$ be an ecc & let $\varepsilon := \frac{1}{2}-\rho$ for $\rho \in (0,\frac{1}{2})$.

An algorithm $\mathcal{D}$ is a local list-decoder for E handling ρ errors, if $\forall x \in \{0,1\}^n$, $\forall y \in \{0,1\}^m$ with $\Delta(y,E(x)) \le \rho$, $\exists$ advice $i_0 \in [\mathrm{poly}(n/\varepsilon)]$

s.t. $\forall j \in [n]$: On input $\langle i_0, j, \text{oracle } y \rangle$
$\mathcal{D}$ runs in $\text{poly}(\lg m, n/\varepsilon)$-time & outputs
x_j with prob. $\geq 2/3$.

[Think of i_0 as the location of x in the unknown "list".]
[You can ask for the whole x, instead of x_j;
when $n = \text{polylog}(m)$.]

Local List Decoding WH

Theorem 1 (Goldreich-Levin '89): lld_{WH} Let $WH:\{0,1\}^n \to \{0,1\}^{2^n}$ & $f:\{0,1\}^n \to \{0,1\}$ be the given oracle s.t. $\exists x \in \{0,1\}^n$ $\Pr_z[f(z) = WH(x)_z] \geq \frac{1}{2} + \frac{\varepsilon}{2}$.

$\exists$ randomized poly(n/ε)-time algorithm to find list $L_f := \{x \mid \Delta(f, WH(x)) \leq \frac{1}{2} - \frac{\varepsilon}{2} =: \rho\}$.

Proof:

Idea – Since f is corrupted close to $1/2$, we need more than two queries (& L_f is large). So, we'll make many (correlated) queries & "guess" some answers (advice).

- $k := \lceil \lg(m+1) \rceil$, where $m := \lceil 200n/\varepsilon^2 \rceil$.
- Randomly pick "locations" $s_1, \ldots, s_k \in \{0,1\}^n$ & guesses $\sigma_1, \ldots, \sigma_k \in \{0,1\}$. [Hope: $\exists!\, x \in L_f$, $\forall i \in [k]$, $\sigma_i = x \odot s_i$.]
- Define $s_T := \bigoplus_{i \in T} s_i$ & $\sigma_T := \bigoplus_{i \in T} \sigma_i$ ← (bit-wise XOR) $\forall T \subseteq [k]$.
- Compute $x_i := \mathrm{maj}_T \{ \sigma_T \oplus f(s_T \oplus e_i) \}$, $\forall i \in [n]$.

  σ_T ← from the guess; e_i ← i-th elementary vector
- OUTPUT $x_1 \cdots x_n$.
- Repeat the above algo. for $1000n/\varepsilon^4$ times.

→ See the analysis in my old lecture notes on the homepg.
- Main Pt.: When guesses are correct, then $x_1 \cdots x_n \in L_f$ whp. □

Local list decoding RM

- Recall that RM "maps" $\binom{\ell+d}{d}$ evaluations of a d-deg ℓ-variate polynomial $P(\bar{x})$ to all $|\mathbb{F}|^\ell$ evaluations.
Our goal is to output $P(x)$, given $x \in \mathbb{F}^\ell$, an oracle to corrupted $RM \circ P$, and an advice $(x_0, y_0 = P(x_0))$.

Thm 2 (Sudan, Trevisan, Vadhan '99): RM has a lld_{RM} handling $1 - 10\sqrt{d/q}$ errors.

non-binary alphabet

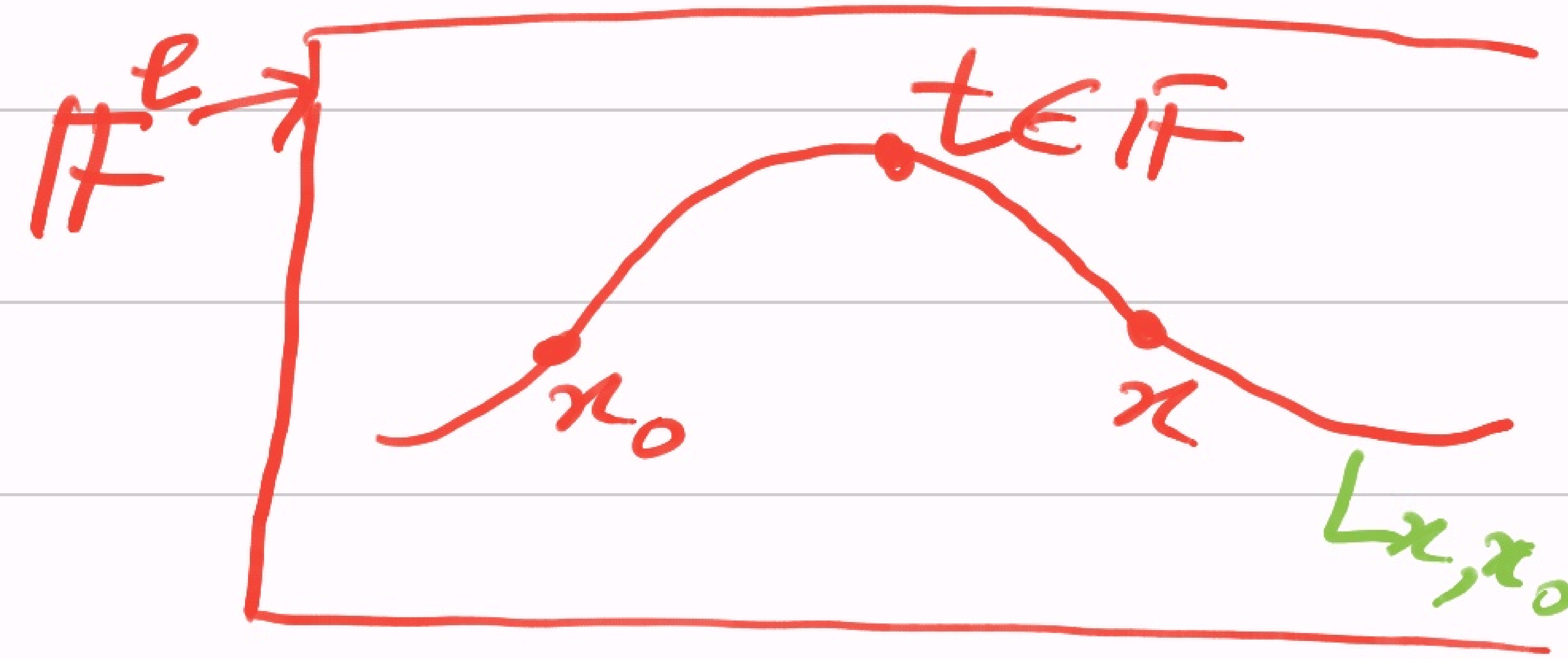
(Compare: with list-decoder of RS handling $1 - \sqrt{d/q}$ errors)

Proof: Idea — Randomly pick $r \in \mathbb{F}$. "Draw" a random cubic-curve through $(0,x)$ & $(r,x_0) \in \mathbb{F}\times\mathbb{F}^\ell$; call it L_{x,x_0}. [L_{x,x_0} has points $\{q(t) := (q_1(t), \ldots, q_\ell(t)) \in \mathbb{F}^\ell \mid$ for all $t \in \mathbb{F}\}$; q_i's cubics.]

$\triangleright$ $q(0) = x$ & $q(r) = x_0$.

Query f on L_{x,x_0}. Run RS list-decoder to find a unique $g(t) := P \circ q(t)$ with $g(r) = P \circ q(r) = P(x_0) = y_0$.

OUTPUT $g(0)$.



- Input: • Oracle f s.t. $\Pr_{x \in \mathbb{F}^\ell}[f(x) = P(x)] > 10\sqrt{d/q}$
  & $|\mathbb{F}| > d^4$.
  • Advice $(x_0, y_0) \in \mathbb{F}^\ell \times \mathbb{F}$.
  • $x \in \mathbb{F}^\ell$.

Output: $y \in \mathbb{F}$.

Decoder:

1) Pick random $r \in \mathbb{F}$ & monic cubics $q_i(t)$, $i \in [\ell]$,
   s.t. $q(t) =: (q_1(t), \ldots, q_\ell(t))$ satisfies $q(0) = x$ & $q(r) = x_0$.
2) Query f on L_{x,x_0} to obtain $S := \{(t, f \circ q(t)) \mid t \in \mathbb{F}\}$.
3) Run RS-list-decoder on S to find the list $g_1, \ldots, g_k$ of all deg-$3d$ polynomials that agree on $\geq 8\sqrt{dq}$ pairs in S.

4) If $\exists!\, i : g_i(r) = y_0$ then OUTPUT $g_i(0)$.
else FAIL.

$\rightarrow$ Read the analysis as an exercise. $\square$

Local list decoding WH∘RM

Thm 3 (STV'99): $E_1 : \{0,1\}^n \to \Sigma^m$ resp. $E_2 : \Sigma \to \{0,1\}^k$ are ecc with lld using advice from index-sets I_1 resp. I_2 & handling $1-\varepsilon_1$ resp. $\frac{1}{2}-\varepsilon_2$ errors.
Then, $E = E_2 \circ E_1$ has lld using advice in $I_1 \times I_2$ & handling $(1-\varepsilon_1 \cdot |I_2|) \times (\frac{1}{2} - \varepsilon_2)$ errors.

Pf: Similar to local decoding. $\square$

- From this we now deduce that for a worst-case hard f, $WH \circ RM \circ tt(f) =: tt(g)$ is the tt of an average-case hard g!

Idea – Otherwise a small circuit C approximates g.
Think of $tt(C)$ as a corruption of $tt(g)$.
$\Rightarrow$ Applying $lld_{WH \circ RM}$ on oracle C,
we get $f(x)$, for our x.
$\Rightarrow$ small circuit for f. $\Rightarrow \lightning$

Hardness Amplification

Theorem (Impagliazzo & Wigderson '97; STV '99): Let $f \in \mathbb{E}$ be s.t. $H_{wrs}(f) \geq S(n)$ for some $S: \mathbb{N} \to \mathbb{N}$. Then, $\exists g \in \mathbb{E}$, $\exists c > 0$ s.t. $H_{avg}(g) \geq S(n/c)^{1/c}$, for large n.

Pf:

- Analyze the parameters of $\mathrm{lld}_{\mathrm{WHoRM}}$ on corrupted version of $\mathrm{WHoRM}(tt(f))$.
- Left as a reading exercise. $\square$