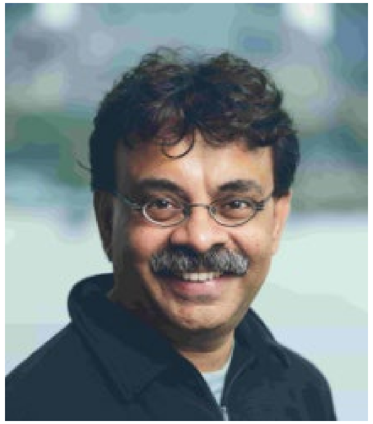
Local Correction of Linear Functions on the Boolean Cube

Prashanth Amireddy
(Harvard)

Amik Raj Behara
(Aarhus)

Manaswi Paraashar
(Aarhus)

Srikanth Srinivasan
(Copenhagen)



Madhu Sudan
(Harvard)

The Problem(s)

- Given oracle access to $f: S^n \to \mathbb{F}$, and $a \in S^n$, compute $P(a)$
  - Where $P(\cdot)$ is the (unique) linear function at dist. $\delta < \frac{1}{4}$ from f.

$$\delta(f,g) := \Pr_{b \in_U S^n}[f(b) \neq g(b)]$$

  - While minimizing #queries to f
  - "Local Correction of linear functions"
- List-Local Correction:
  - $\delta(f,P) \to 1 - \frac{1}{|S|}$; L not uniquely defined – but can hope for
    - List size to be small if $\delta(f,P) < 1 - \frac{1}{|S|} - \epsilon$
    - If so … provide oracle access to all such $P_1, P_2, \ldots, P_L$

The results

- Consider: $S = \{0,1\}$; arbitrary $\mathbb{F}$ (or even arbitrary abelian G)
- Thm 1: Local Corrector: Can compute $P(a)$ when $\delta(f,P) < \frac{1}{4} - \epsilon$ making $\tilde{O}_\epsilon(\log n)$ queries.
- Thm 2: List-decoding bound: There are at most $poly(\epsilon^{-1})$ linear functions P with $\delta(P,f) \leq \frac{1}{2} - \epsilon$
- Thm 3: Can be locally list-decoded with $\tilde{O}_\epsilon(\log n)$ queries.

Motivation/Context

- (Mainly mathematical)
- Context 1: Ore-"DeMillo-Lipton-Schwartz-Zippel" Lemma
  - Class of polynomials $f: S^n \to \mathbb{F}$ of degree d form code of relative distance $\delta = \delta(|S|, d) > 0$ independent of n
  - Locally correctable if $S = \mathbb{F}$; what if not?
  - Many common tools (affine-change of basis) unavailable … what are replacements?
- Context 2: Locally correctable codes over reals …
  - Unknown if there exists one that is correctable with O(1) queries and $\Omega(1)$ fraction error for every message length k
  - Our work – gives first code with $o(k)$ queries

Part 1: Decoding from $\frac{1}{4} - \epsilon$ fraction errors

3 Step approach:

- Construct a sequence of "oracles": $f \to f_1 \to f_2 \to P$
- Step 1: $f \to f_1$: where $\delta(f_1, P) < \tau$ for any constant τ
  - Oracle for f_1 makes $O_{\tau,\epsilon}(1)$ calls to oracle for f
- Step 2: $f_1 \to f_2$: where $\delta(f_2, P) = O\left(\frac{1}{\text{poly}\log n}\right)$
  - Oracle for f_2 makes $O(\text{poly}\log\log n)$ queries to f_1
- Step 3: $f_2 \to P$:
  - Oracle for P makes $\tilde{O}(\log n)$ queries to f_2

Step 1

- Key to Step 1 (and Step 2): Only need to recover $P(a)$ for random a (whp)

- Def: Cube C containing a given by function $h: [n] \to [k]$ and contains all points $\{a \oplus (y_{h(1)} \dots y_{h(n)}) \mid y_1 \dots y_k \in \{0,1\}\}$

- Small set expansion of noisy hypercube (aka hypercontractivity)
  $\Rightarrow$ Cube is a good sampler of $\{0,1\}^n$ for random a
  $\Rightarrow$ Cube has roughly $\frac{1}{4} - \epsilon$ fraction errors.
  $\Rightarrow$ Can brute force decode $f|_C$; $k = O(1) \Rightarrow$ #queries = $O(1)$; Error depends on prob. Sampler not good … goes $\to 0$ as $k \to \infty$

Step 3:

- To compute, say, $P(1^n)$:
  - Find $v_1 \dots v_{\log n} \in \{0,1\}^n$ and $\alpha_1 \dots \alpha_{\log n} \in \mathbb{Z}$s.t.
    - Each v_i roughly balanced $(\frac{1}{2} \pm \frac{1}{\sqrt{n}})$-fraction 1s.
    - $\sum_i \alpha_i = 1$
    - $\sum_i \alpha_i v_i = 1^n$
  - Output $\sum_i f(v_i)$
- Key claim: Such v_i's and α_i's exist.
  - Proof: Constructive.
- Aside: Proof sort of "converse" to a result from prev paper by Bafna-Srinivasan-S who show $\tilde{\Omega}(\log n)$ necessary.

Step 2:

- (Most novel/intricate?)
- Key idea:
  - There exists an $O(1)$-query error reducer: $g_1 \to g_2$, i.e.,
    - g_2 makes $O(1)$-queries to g_1
    - $\delta(g_2, P) = O(\delta(g_1, P)^2)$
    - Proof: Maybe on board?
  - Repeat $k = O(\log\log\log n)$ times
    - Queries $\exp(k) = O(\mathrm{poly}\log\log n)$
    - Error = $\exp(-\exp(k)) = O\left(\frac{1}{\log n}\right)$

Part 2: List-Decodability from $\frac{1}{2} - \epsilon$ fraction errors

Overview

- Recall main theorem:
  - For every $f:\{0,1\}^n \to \mathbb{F}$ and every $\epsilon > 0$ there exist at most $\mathrm{poly}(\frac{1}{\epsilon})$ linear functions P s.t. $\delta(f,P) \leq \frac{1}{2} - \epsilon$
- Actually prove it for functions mapping to any abelian group G
  - Many steps and cases …
  - Step 0: Reduce to case of finite group G (size depends on n)
  - Let $G = G_2 \times G_3 \times G_0$ (G_p p-group; G_0 all elements have order at least 5)
  - Tackle each case separately; Combining easy

The G_0 case

- Substeps:
  - If $P_1, \ldots, P_L$ all $\frac{1}{2} - \epsilon$-close to f then there exist $t = \Omega(L)$ polynomials among them that are all $\frac{3}{4} + .0001$-close to one of them.
  - Say $P_1 \ldots P_t$ close to P_1
  - For every $i \in [t]$, $P_i - P_1$ is a sparse polynomial depending only on $O(1)$ variables.
  - Extreme cases:
    - All $P_i - P_1$'s depend on $\mathrm{poly}\left(\frac{1}{\epsilon}\right)$ variables … easy to count
    - P_i's depend on disjoint set of variables … unlikely to agree
  - General case … reduces to combination of extreme cases

The G_2 & G_3 cases

- G_2 case essentially known … but new proof in this paper.
- Unified with G_3 case; main ingredients
  - Extended Johnson Bound for ranges $\mathbb{Z}_2$ and $\mathbb{Z}_3$
  - "Special Intersection Properties" of agreement sets to lift results to G_2, G_3

Extended Johnson Bound

- Extended Johnson Bound: $\exists C$ s.t. if $\delta(f, P_i) \leq \frac{1}{2} - \epsilon_i$ then $\sum_i \epsilon_i^C \leq 1$
  - $\mathbb{Z}_2$ case is the standard one
  - $\mathbb{Z}_3$ case uses Fourier analysis + some sparsity …
    - Specifically: There are at most 31 "highly distinct" polynomials that are at distance at most $\frac{1}{2}$ from f. (proved using Fourier analysis)
    - "Highly distinct" $\coloneqq$ differ on six variables.
    - Now proceed in manner similar to G_0 …

Special Intersection Properties [DGKS]

- Suppose $f: \{0,1\}^n \to G \times H$ … so $f = (f_G, f_H)$
- Suppose $P_1 \dots P_L$ have significant agreement with f_G; let

  $S_i := \{a \in \{0,1\}^n | \, f(a) = P_i(a)\}$ and $\delta(P_i, f_G) = \frac{1}{2} - \epsilon_i$, so $|S_i| = 2^n \left(\frac{1}{2} + \epsilon_i\right)$
- How can this list become larger when looking at f?
  - Each P_i might extend to several P_{ij}'s with agreement on sets S_{ij} with size $2^n \left(\frac{1}{2} + \epsilon_{ij}\right)$
  - Suppose S_i's satisfy extended Johnson i.e., $\sum_i \epsilon_i^D \leq 1$
  - Under what condition can you show $\sum_{ij} \epsilon_{ij}^D \leq 1$?
  - Turns out S_{ij}'s have special intersection properties and this can be exploited.

S.I.P. (contd.)

- $S_1 \dots S_m$ have (ρ, τ, C) –special intersection properties if:
  - $\mu(S_i) \geq \rho$
  - $\mu(S_i \cap S_j) \leq \rho$
  - Extended Johnson Bound applies: $\mu(S_i) = \rho + \epsilon_i \Rightarrow \sum_i \epsilon_i^C \leq 1$
  - (Property τ): for $I \subseteq [m]$ let $S_I = \cap_{i \in I} S_i$. If $\mu(S_I) > \tau$ then $\forall\, J \subseteq I$ with $|J| = 2$, $S_J = S_I$ $(\mu(S_I) > \tau \Rightarrow \{S_i |\, i \in I\}$ form sunflower)
- DGKS Thm (specialized to example from previous page):
  - $\forall C\ \exists D$ If for every i, $\{S_{ij}\}_j$ form a $\left(\frac{1}{2}, \frac{1}{4}, C\right)$-SIP then $\sum_j \epsilon_{ij}^D \leq \epsilon_i^D$
  - Note that to prove EJB for $\{P_{ij}\}_j$ we can look only at f_H!

Brief aside on use of SIP

- DGKS – roughly apply it once to lift from $\mathbb{Z}_2$ to $\mathbb{Z}_{2^t}$ and once more (say) to $\mathbb{Z}_{2^t}^s$
  - Works fine, gets worse exponent D
- Our new application: directly works with $G \bmod H$ and H (even when $G \neq (G \bmod H) \times H$): so a single lifting step gets to $\mathbb{Z}_{2^t}^s$

Part 3: Algorithmic List-Decoding

- Develops idea from S.,Trevisan,Vadhan
- To compute $P_1(a) \dots P_L(a)$: can find a random cube C containing a;
- But how to construct an oracle that consistently outputs values of say P_1?
- Idea: Use $P_1|_{\hat{C}}$ as advice.
  - Now decode at a using random cube C that contains a and $\hat{C}$
  - Leads to some non-trivial complications, but … all ends well.

Summary

- Considered: $S = \{0,1\}$; arbitrary $\mathbb{F}$ (or even arbitrary abelian G)
- Thm 1: Local Corrector: Can compute $P(a)$ when $\delta(f, P) < \frac{1}{4} - \epsilon$ making $\tilde{O}_\epsilon(\log n)$ queries.
- Thm 2: List-decoding bound: There are at most $poly(\epsilon^{-1})$ linear functions P with $\delta(P, f) \leq \frac{1}{2} - \epsilon$
- Thm 3: Can be locally list-decoded with $\tilde{O}_\epsilon(\log n)$ queries.

- Natural directions:
  - Higher degree? Larger S?
  - Better LCCs over reals?

Thank you!