Local Decoding and Testing Polynomials over Grids

Madhu Sudan
Harvard University

Joint work with Srikanth Srinivasan (IIT Bombay)

DeMillo-Lipton-Schwarz-Zippel

- Notation:
  - $\mathbb{F}$ = field,
  - $\mathbb{F}(d,n)$ = set of deg. $\leq d$, n var. poly over $\mathbb{F}$
  - $\delta(f,g)$ = normalized Hamming distance
  - $\delta_d(f) = \min_{g \in \mathbb{F}(d,n)} \{\delta(f,g)\}$
- DLSZ Lemma: Let $S \subseteq \mathbb{F}$, with $2 \leq |S| < \infty$. If $f,g: S^n \to \mathbb{F}$ satisfy $f \neq g \in \mathbb{F}(d,n)$ then $\delta(f,g) \geq 2^{-d}$.
  - Strengthens to $\delta(f,g) \geq 1 - d/|S|$ if $d < |S|$
  - Holds even if $S = \{0,1\}$
- For this talk: d = constant so $\delta(f,g)$ bounded away from 0.
- DLSZ Lemma converts polynomials into error-correcting codes.
  - With bonus: decodability, locality …

Local Decoding and Testing

- Informally: Given oracle access to $f: S^n \to \mathbb{F}$ …
  - Testing: determine if $\delta_d(f) \leq \epsilon$
  - Decoding: If $\delta_d(f) \leq \delta_0$ and g is nearest poly then determine value of $g(a)$ given $a \in S^n$
- Parameters:
  - Query complexity = # of oracle queries
  - Testing accuracy $\equiv \frac{\Pr[\text{rejection}]}{\delta_d(f)}$
  - Decoding distance (δ_0 above)
- Ideal: Query complexity $O_d(1)$, accuracy, dec. distance $\Omega_d(1)$
- Terminology $\mathbb{F}(d,n)$ testable (decodable) over S if ideal achievable.
- Thms: For finite $\mathbb{F}$, $\mathbb{F}(d,n)$ decodable over $\mathbb{F}$ (folklore) and testable over $\mathbb{F}$ ([Rubinfeld+S.'96] …. [Kaufman+Ron'04])

Polynomials over Grids

- Testing + decoding thms hold only for $f: \mathbb{F}^n \to \mathbb{F}$
- Not surprising … repeated occurrence with polynomials.
  - Even classical decoding algorithms (non-local) work only in this case!
  - [Kopparty-Kim'16]: Non-local Decoder (up to half the distance) for $f: S^n \to \mathbb{F}$ for all finite S

- This talk: Testing + decoding when $S = \{0,1\}$ "grid"
  - (Note: equivalent to $S_1 \times \cdots \times S_n$ with $|S_i| = 2$)

Obstacles to decoding/testing

- Standard insight behind testing/decoding $d < |\mathbb{F}|$
  - f has deg. $\leq d$ iff f restricted to lines has $\leq d$
  - Can test/decode function on lines.
  - Reduces complexity from $\mathbb{F}^n$ to $\mathbb{F}$.
- When $d > |\mathbb{F}|$ use subspaces
  - (= collection of linear restrictions)
  - "Affine invariance" → "2 transitivity" ...
- Problem over grids:
  - General lines don't stay within grid!
  - Only linear restrictions that stay in grid:
    - $x_i = 0$ or $x_i = 1$
    - $x_i = x_j$ or $x_i = 1 - x_j$ (denoted "$x_i = x_j \oplus b$")

Main Results

- Thm 1. $\mathbb{F}(1,n)$ <u>not</u> locally decodable over grid if $\mathrm{char}(\mathbb{F}) \to \infty$
  - No 2-transitivity! Serious obstacle!

- Thm 2. $\mathbb{F}(d,n)$ is locally decodable over grid if $\mathrm{char}(\mathbb{F}) < \infty$.
  - Query complexity = $2^{O(\max\{d,\mathrm{char}(\mathbb{F})\})}$
  - Decoding without 2-transitivity!

- Thm 3. $\mathbb{F}(d,n)$ is locally testable over grid ($\forall \mathbb{F}$)
  - Testing without decoding!

Proofs: Not LDC over $\mathbb{Q}$

- Idea ($d = 1$): Consider f = linear function that is erased on "imbalanced" points (of weight $\notin \left[\frac{n}{2} - c\sqrt{n}, \frac{n}{2} + c\sqrt{n}\right]$)
- Claim 1: No local constraints over $\mathbb{Q}$ on balanced points and 0^n
  - Claim 1.1: if $f(0^n) = \sum_{i=1}^{\ell} a_i f(x_i) \, \forall f$ then $0^n = \sum_{i=1}^{\ell} a_i x_i$
  - Claim 1.2: $1 \leq |a_i| \leq \ell!$
  - Claim 1.3: If $x_1, \ldots, x_\ell$ are ϵ-balanced

    then $\sum_i^j a_i x_i$ is $\left(\epsilon \cdot \sum_i^j |a_i|\right)$-balanced
- Claim 2: No local constraints $\Rightarrow$ Not LDC
  - Known over finite fields [BHR'04]
  - Not immediate for $\mathbb{Q}$

Proofs: LDC over small characteristic

- Given: f δ-close to $g \in \mathbb{F}(d,n)$, $a \in \{0,1\}^n$ ($\mathrm{char}(\mathbb{F}) = p$)
- Pick $\sigma: [n] \to [k]$ uniformly and let $x_i = a_i \oplus y_{\sigma(i)}$
  - So considering $f'(y_1, \ldots, y_k) = f(a \oplus y_\sigma)$:
    - $f'(0) = f(a)$; $g'(0) = g(a)$
    - Claim 1: For balanced y, have $a \oplus y_\sigma \perp a$
    - Claim 2: If $\frac{k}{2} = p^t$ and $\frac{k}{2} > d$ then $g'(0)$ determined by $\left\{g'(y) \mid |y| = \frac{k}{2}\right\}$
      - (Usual ingredient "Lucas's Theorem" $x^p + y^p = (x+y)^p$)

Proofs: LTC over all fields: Test

- Test
  - Pick $a \in \{0,1\}^n$ uniformly and $\sigma: [n] \to [k]$ "randomly"
  - Verify $f'(y_1, \ldots, y_k) = f(a \oplus y_\sigma)$ is of degree d
- Randomly = ?
  - Not uniformly (don't know how to analyze).
  - Instead "random union-find"
    - Initially each x_i in ith bucket.
    - Iterate: pick two uniformly random buckets and merge till k buckets left.
    - Assign $y_1, \ldots, y_k$ to k buckets randomly.
  - Allows for proof by induction

Proofs: LTC over all field: Analysis

- Key ingredients: (mimics BKSSZ analysis for RM)
  - Main claim: $\Pr[\text{test rejects } f] \geq 2^{-O(d)} \cdot \min\{1, \delta_d(f)\}$
  - Proof by induction on n with 3 cases:
    - Case 1: f moderately close to deg. d
      - Show that eventually f disagrees from nearby poly on exactly one of 2^k samples.
        - Usual proof – pairwise independence
        - Our proof – hypercontractivity of sphere [Polyanskiy] + prob. fact about random union-find
    - Case 2: $\Pr_{i,j,b}\left[f|_{x_i = x_j \oplus b} \text{ very close to deg. } d\right]$ small Induction
    - Case 3: f far from deg. d, but $\Pr_{i,j,b}[\ldots]$ high
      - Use algebra to get contradiction. (Stitch different polynomials from diff. (i,j,b)'s to get nearby poly to f)

Conclusions + Open questions

- Many aspects of polynomials well-understood only over domain = $\mathbb{F}^n$
- Grid setting (S^n) far less understood.
- When $S = \{0,1\}$ (equivalently $|S| = 2$) testing possible without local decoding!
  - (novel in context of low-degree tests!)

- General S open!! (even $|S| = 3$ or even $S = \{0,1,2\}$?)
- Is there a gap between characterizability and testability here?

Thank You!