Correlation testing for affine invariant properties on $F^n_p$

Shachar Lovett
Institute for Advanced Study

Joint with Hamed Hatami (McGill)
Property testing

• Math: infer **global structure** from **local samples**

• CS: Super-fast (randomized) algorithms for approximate decision problems

• Decide if large object *approximately has property*, while testing only a *tiny fraction* of it
Graph properties: 3-colorability

• Input: graph G
• Is G 3-colorable?

• Local test:
  – Sample $(1/\varepsilon)^{O(1)}$ vertices
  – Accept if induced subgraph is 3-colorable

• Analysis:
  – Test always accepts 3-colorable graphs
  – Test rejects (w.h.p) graphs $\varepsilon$-far from 3-colorable

[Goldreich-Goldwasser-Ron’96]
Algebraic properties: linearity

• Input: function $f : F_p^n \rightarrow F_p$

• Is $f$ linear?

• Local test:
  – Sample $x, y \in F_p^n$
  – Check if $f(x + y) = f(x) + f(y)$
  – Repeat $1/\varepsilon^{O(1)}$ times

• Analysis:
  – Test always accepts linear functions
  – Test rejects (w.h.p) functions $\varepsilon$-far from linear

[Blum-Luby-Rubinfeld’90]
Codes: locally testable codes

- Code: \( C \subset F_p^n \)
distinct elements have large distance

- Input: word \( w \in F_p^n \)

- \( C \) is locally testable if there exists a (randomized) test which queries a few coordinates and
  - Always accepts codewords
  - Rejects (w.h.p) if \( w \) is far from all codewords

- The “mathematical core” of the PCP theorem
- Open: can \( C \) have constant rate, distance and testability?
Proofs: Probabilistic Checkable Proofs

- PCP Theorem: robust proof system

- Encoding of theorems + randomized local test (queries few bits of proof)
  - Test always accepts legal proofs of theorems
  - Test rejects (w.h.p) proofs of false theorems

- Major tool to prove hardness of approximation
Property testing: general framework

- Universe: set of objects (e.g. graphs)
- **Property**: subset of objects (e.g. 3-col graphs)
- Test: randomized small sample (e.g. small subgraph)

- Property is testable if local consistency implies approximate global structure
Which properties are testable?

• **Graph properties**: well understood

• **Algebraic properties**: partially understood

• **Locally testable codes**: major open problems

• **PCP / hardness of approximation**: whole field
Correlation testing

• Property testing: **strong** global structure
  – Is the object close to having property

• **Correlation testing:** **weak** global structure
  – Is the object slightly related to having property

• Motivation: possible generalizations of the inverse Gowers conjecture theorem
Correlation testing
Linearity correlation testing

- Function $f : F_p^n \rightarrow $

- Correlation of $f, g$: $\langle f, g \rangle = E_{x \in F_p^n} [f(x)g(x)]$

- Correlation with linear functions (characters):

$$\| \hat{f} \|_\infty = \max_{\ell : F^n_p \rightarrow F_p \text{ linear}} |\langle f, \omega^\ell \rangle|$$

$(\omega_p = e^{2\pi i/p})$
Linearity correlation testing

- Linear correlation: **global property**
  Witnessed by **local average**

\[
\mathbb{E}_{x,y,z \in \mathbb{F}_p^n} \left[ f(x + y + z) \frac{f(x + y) f(x + z) f(x)}{f(x)} \right] = \sum_{\alpha} |\hat{f}(\alpha)|^4 \| f \|_{U^2}^4
\]

- Identifies functions correlated with linear funcs:
  - *f* correlated to linear:
    \[ \| \hat{f} \|_{\infty} \geq \varepsilon \implies \| f \|_{U^2} \geq \varepsilon \]
  - *f* is not correlated:
    \[ \| \hat{f} \|_{\infty} \leq \delta \implies \| f \|_{U^2} \leq \sqrt{\delta} \]
Linearity correlation testing

• Discrete setting: \( f : F_p^n \rightarrow F_p \)
  
  Test queries \textbf{4 locations}, accepts \( f \) if
  \[
  f(x + y + z) - f(x + y) - f(x + z) + f(x) = 0
  \]

• Acceptance probability:
  - \( \varepsilon \)-correlated with linear: \( \text{prob.} \geq \frac{1}{p} + \varepsilon^2 \)
  - negligible correlation: \( \text{prob.} \leq \frac{1}{p} + o(1) \)

• Property testing: \#queries depends on \( \varepsilon \)
  • Here: \#queries=4, acceptance prob. depends on \( \varepsilon \)
Testing correlation with polynomials

• Inverse Gowers Theorem (for finite fields):

  Global structure: correlation with low-degree polynomials (Higher-order Fourier coefs)

Witnessed by local average
Testing correlation with polynomials

- Correlation with degree $d$ polynomials:

$$\| f \|_{u({\text{Poly}}_d)} = \max_{Q: \mathbb{F}_p^n \to \mathbb{F}_p} \text{ polynomial degree } \leq d \quad | \langle f, \omega^Q_p \rangle |$$

- Gowers norm: average over $2^{d+1}$ points

$$\| f \|_{U^{d+1}} = \mathbb{E}_{x, y_1, \ldots, y_{d+1} \in \mathbb{F}_p^n} \left[ \prod_{I \subseteq [d+1]} C^{d-|I|} f(x + \sum_{i \in I} y_i) \right]$$

$C = \text{Conjugation}$
Testing correlation with polynomials

• Direct theorem [Gowers]
  \[ \| f \|_{u(Poly_d)} \geq \varepsilon \implies \| f \|_{U^{d+1}} \geq \varepsilon \]

• Inverse Theorem [Bergelson-Tao-Ziegler]
  \[ \| f \|_{U^{d+1}} \geq \varepsilon \implies \| f \|_{u(Poly_d)} \geq \delta(\varepsilon) \]
  (if \( p < d \) then \( Poly_d = \) non classical polynomials)
Main theorem

• Gowers norms: local averages which witness global correlation to low-degree polynomials

• Question: are there other such properties?
  – Correlation witnessed by local averages

• Theorem [today]: no
  (affine invariant properties, in large fields)
Correlation with property

- Property  \( P \subseteq \{ g : F_n^p \rightarrow F_p \} \)
  (can also consider  \( P \subseteq \{ g : F_n^p \rightarrow \square \} \) )

- Function  \( f : F_n^p \rightarrow F_p \)

- Correlation of \( f \) with property \( P \):
  \[
  \| f \|_{u(P)} = \max_{g \in P} |\langle \omega^f, \omega^g \rangle|
  \]
Local test

- Local test (with q queries):
  - Distribution over \( \{x_1, \ldots, x_q\} \subset F^n_p \)
  - Local test \( T: F^q_p \rightarrow \{0, 1\} \)
    \[
    T(f) = E[T(f(x_1), \ldots, f(x_q))]
    \]

- \( T \) tests **correlation with property \( P \)** if
  \[
  \forall \varepsilon \ \exists \delta \in (0, \varepsilon), \theta^- < \theta^+ \text{ such that }
  \]
  \[
  \|f\|_{u(P)} \geq \varepsilon \Rightarrow T(f) \geq \theta^+
  \]
  \[
  \|f\|_{u(P)} \leq \delta(\varepsilon) \Rightarrow T(f) \leq \theta^-
  \]
Affine invariant properties

- Property $P \subseteq \{ f : F_p^n \rightarrow F_p \}$
- $P$ is affine invariant if
  
  $$f(x) \in P \iff g(x) = f(Ax + b) \in P$$

- Examples:
  - Linear functions; degree-d polynomials
  - Functions with sparse / low-dim. Fourier representation

- Local tests for affine invariant properties are w.l.o.g local averages over linear forms
Local average over linear forms

- **Variables** \( X = (X_1, \ldots, X_k) \in (\mathbb{F}_p^n)^k \)
- **Linear form** \( L(X) = \lambda_1 X_1 + \cdots + \lambda_k X_k \quad (\lambda_i \in \mathbb{F}_p) \)
- **System of linear forms** \( L = \{L_1, \ldots, L_q\} \)
  
  - E.g. \( L = \{X + Y + Z, X + Y, X + Z, X\} \)

- **Average over linear forms:**

\[
T_{L,\alpha}(f) = \mathbb{E}_{X \in (\mathbb{F}_p^n)^k} \left[ \omega_p^{\alpha_1 f(L_1(X)) + \cdots + \alpha_q f(L_q(X))} \right]
\]
\[
(\alpha \in \mathbb{F}_p^q)
\]
Local tests: affine invariant properties

- **Local tests** for affine invariant properties are *w.l.o.g* averages over homogenous linear forms

  \[ L = \{L_1, \ldots, L_q\} \text{ homogenous if } L_i(X) = X_1 + \sum_{i=2}^{k} \lambda_i X_i \]

- \( \exists \text{ systems of linear forms } L_i, \alpha_i \) such that the sets

  \[
  \{(T_{L_1,\alpha_1}(f), \ldots, T_{L_m,\alpha_m}(f)) \| f \|_{u(P)} \geq \varepsilon\} \\
  \{(T_{L_1,\alpha_1}(f), \ldots, T_{L_m,\alpha_m}(f)) \| f \|_{u(P)} \leq \delta(\varepsilon)\}
  \]

are disjoint
Local tests: affine invariant properties

• Claim: any local test $\Rightarrow$ local averages

• Proof: $P$ affine invariant, so $\forall A, b$

$$f \|_{u(P)} \Leftrightarrow f (Ax + b) \|_{u(P)}$$

• Choosing $A, b$ uniformly:
  – transform each query $(x_1, \ldots, x_q)$
  – to a homogeneous system $(Ax_1 + b, \ldots, Ax_q + b)$
Main theorem (1)

- Property \( P = (P_n \subset \{ g : F_p^n \to F_p \})_{n \in \mathbb{N}} \)
  - Consistent \( P_n \subset P_{n+1} \)
  - Affine invariant
  - Sparse \( |P_n| = p^{o(p^n)} \)

- Thm: If \( P \) is locally testable with \( q \) queries (\( p > q \)) then \( \exists d \leq q \) such that for any sequence of functions \( (f_n : F_p^n \to F_p)_{n \in \mathbb{N}} \) which are unbiased \( \lim_{n \to \infty} E \omega_{p}^{f_n} = 0 \)
  
  \[ \lim_{n \to \infty} \| f_n \|_{u(P)} = 0 \iff \lim_{n \to \infty} \| f_n \|_{U^d} = 0 \]
Main theorem (2)

- Consistent property

\[ P = (P_n \subset \{ g : F^n_p \to \mathbb{R}, \| g \|_{\infty} \leq 1 \})_{n \in \mathbb{N}} \]

- Thm: If P is testable by systems of q linear forms \((p > q)\) then \(\exists d \leq q\), for any bounded functions \((f_n : F^n_p \to \mathbb{R})_{n \in \mathbb{N}}\)

\[ \lim_{n \to \infty} \| f_n - Ef_n \|_{u(P)} = 0 \iff \lim_{n \to \infty} \| f_n - Ef_n \|_{U^d} = 0 \]

- Q: Is this true for any norm defined by linear forms?
Proof
Main theorem

- $P = (P_n \subset \{ g : F^n_p \rightarrow \mathbb{R}^k, \| g \|_{\infty} \leq 1 \})_{n \in \mathbb{N}}$

- $P$ testable by systems of $q$ linear forms ($q < p$)

- Thm: $u(P)$ norm equivalent to some $U^d$ norm:

  \[
  \text{if } \lim_{n \to \infty} E f_n = 0 \text{ then } \lim_{n \to \infty} \| f_n \|_{U^d} = 0 \iff \lim_{n \to \infty} \| f_n \|_{u(P)} = 0
  \]
Proof idea

• Dfn: \( S = \{ \text{degrees } d: \forall n \text{ large } \exists \text{ degree-} d \text{ poly } Q_n \} \)
  1. \( Q_n \) correlated with property \( P \)
  2. \( Q_n \) has “high enough” rank

• \( D = \text{Max}(S) \)
  – \( D \) is bounded (bound depends on the linear systems)

• Lemma 1: \( \lim_{n \to \infty} \|f_n\|_{U^{D+1}} = 0 \Rightarrow \lim_{n \to \infty} \|f_n\|_{u(P)} = 0 \)

• Lemma 2: \( \lim_{n \to \infty} \|f_n\|_{u(P)} = 0 \Rightarrow \lim_{n \to \infty} \|f_n\|_{U^{D+1}} = 0 \)
**Polynomial rank**

- $Q$ – degree $d$ polynomial

- $\text{Rank}(Q)$ – minimal number of lower-degree polynomials $R_1, \ldots, R_c$ needed to compute $Q$
  
  $Q(x) = \Gamma(R_1(x), \ldots, R_c(x))$

- Thm [Green-Tao, Kaufman-L.]
  
  If $P$ has **high enough rank**, it has **negligible correlation** with lower degree polynomials
Polynomial factors

• Polynomial factor: \( B = \{ Q_1, \ldots, Q_C : F^n_p \rightarrow F_p \} \)
  – Sigma-algebra defined by \( Q_1, \ldots, Q_C \)
  – \( f : F^n_p \rightarrow \mathbb{F} : \) average over \( B \), \( E[f \mid B] \)

• Complexity(\( B \)): \( C = \) number of basis polys
• Degree(\( B \)): max degree of \( Q_1, \ldots, Q_C \)
• Rank(\( B \)): min. rank of linear comb. of \( Q_1, \ldots, Q_C \)
  – Large rank: \( Q_1(x), \ldots, Q_C(x) \) are nearly independent
Decomposition theorems

- Fix \( d < p \)

- \( f : F_p^n \rightarrow \) can be decomposed as \( f = f_1 + f_2 \)
  
  - \( f_1 = \mathbb{E}[f \mid B] \)
    
    B has degree d, high rank, bounded complexity
  
  - \( \| f_2 \|_{U^{d+1}} = 1 \)
Complexity of linear systems

- **Linear form:** \( L(X) = \lambda_1 X_1 + \cdots + \lambda_k X_k \)
- **Linear system:** \( L = \{L_1, \ldots, L_q\} \)
- **Average:**
  \[
  T_L(f) = E_{X \in (F_p^n)^k} \prod_{i=1}^{q} f(L_i(X))
  \]
- **Complexity:** \( \text{min. } d, \text{ if } f = f_1 + f_2, \| f_2 \|_{U^{d+1}} \leq 1 \) then \( T_L(f) \approx T_L(f_1) \)
- **C-S complexity** [Green-Tao]
- **True complexity** [Gowers-Wolf, Hatami-L.]
Proof idea

- **Dfn:** \( S = \{ \text{degrees } d: \forall \text{large } n \ \exists \text{degree-}d \text{ poly } Q_n \) 
  1. \( Q_n \) correlated with property \( P \)
  2. \( Q_n \) has "high enough" rank

- **D=Max(S)**
  - \( D \) is bounded (≤ complexity of linear systems)

- **Lemma 1:** \( \lim_{n \to \infty} \| f_n \|_{U^{D+1}} = 0 \Rightarrow \lim_{n \to \infty} \| f_n \|_{u(P)} = 0 \)

- **Lemma 2:** \( \lim_{n \to \infty} \| f_n \|_{u(P)} = 0 \Rightarrow \lim_{n \to \infty} \| f_n \|_{U^{D+1}} = 0 \)
Lemma 1: Small $U^{D+1} \rightarrow$ small $u(P)$

- **D**: max deg of high rank polys correlate with $P$
- Assume $\| f \|_{U^{D+1}} \leq 1$ but $\| f \|_{u(P)} \geq \varepsilon$

- **Step 1**: reduce to “structured function”
  - Linear system $L$ of complexity $S$ ($S > D$)
  - Decompose: $f = f_1 + f_2$
    - $f_1 = E[f | B]$
    - $f_2 \|_{U^{S+1}} \leq 1$

- Reduce to studying $f_1$ - func. of deg $\leq S$ polys:
  - $f_1 \|_{U^{D+1}} \leq 1$
  - $T_L(f) \approx T_L(f_1) \Rightarrow f_1 \|_{u(P)} \geq \varepsilon'$
Lemma 1: Small $U^{D+1} \rightarrow$ small $u(P)$

- D: max deg of **high rank polys** correlate with $P$
- Structured function: $f_1 = E[f \mid B]$, $\deg(B) \leq S$
  - $\| f_1 \|_{U^{D+1}} \leq 1$
  - $\| f_1 \|_{u(P)} \geq \epsilon'$
- Will show: $\| f_1 \|_{u(P)} \approx 0$
- Use the structure: $f_1(x) = \sum \alpha_i \omega_{p}^{Q_i(x)}$, $\deg Q_i \leq S$
  - $\deg(Q_i) \leq D \Rightarrow \alpha_i \approx 0$ because $\| f_1 \|_{U^{D+1}} \leq 1$
  - $\deg(Q_i) > D \Rightarrow \| \omega_{p}^{Q_i} \|_{u(P)} \approx 0$ by def of $D$
Lemma 2: small $u(P)$ $\Rightarrow$ small $U^{D+1}$

- Key ingredient: invariance principle
  - High rank polynomials "look the same" to averages

- $\{Q_1, \ldots, Q_c\}, \{Q'_1, \ldots, Q'_c\}$ high rank, $\deg(Q_i) = \deg(Q'_i)$
  
  $f(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$
  
  $f'(x) = \Gamma(Q'_1(x), \ldots, Q'_c(x))$

Then local averages cannot distinguish $f,f'$:

\[ T_L(f) \approx T_L(f') \]
Part 2: small $u(P) \rightarrow$ small $U^{D+1}$

- D: max deg of high rank polys correlate with $P$
- Assume $\|f\|_{u(P)} \leq 1$ but $\|f\|_{U^{D+1}} \geq \epsilon$
  - Reduce to structured function, $f_1 = E[f | B]$

- $f_1$ correlated with high-rank $Q$ of degree $\leq D$
  - Assume for now: $\deg(Q) = D$

- Dfn of D: Exists high rank poly $Q'$, $\deg(Q') = D$, $Q'$ correlated with some function $g \in P$

- Contradiction: Define $f'_1 = f_1$ with $Q$ replaced by $Q'$
  - Invariance principle: $T_L(f_1) \approx T_L(f'_1)$
  - $f'_1$ is correlated with $g \in P$
Part 2: small $u(P) \Rightarrow$ small $U^{D+1}$

- Problem: what if $f_1$’ correlated with high rank poly of degree $< D$?
  - Solution: can find $Q'$ correlated with property $P$ for of all degrees $\leq D$
  - Reason: systems of averages are robust

- Thm: for any family of linear systems, the set
  \[ \{(T_{L_1}(f), \ldots, T_{L_k}(f)) : f : \mathbb{F}_p^n \rightarrow \mathbb{F}, \|f\|_{\infty} \leq 1\} \subset \mathbb{R}^k \]
  has a non-empty interior for some finite $n$
  (unless not for trivial reasons)
  - analog of [Erdos-Lovasz-Spencer] for additive settings
Summary

• Property testing: witness strong structure by local samples
• Correlation test: witness weak structure

• Main result: any affine invariant property which is correlation testable, is essentially equivalent to low-degree polynomials
Open problems

• Which norms can be defined by local averages
  – Are always equivalent to some $U^d$ norm?

• Testing in low characteristics

• Is it possible to test if a function $f : F_2^n \rightarrow F_2$ is correlated with cubic polynomials?
  – $U^4$ norm doesn’t work
  – Unknown even if #queries depends on correlation

THANK YOU!