Some 3CNF Properties are Hard to Test

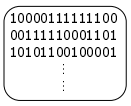
Eli Ben-Sasson
Harvard & MIT

Prahladh Harsha
MIT

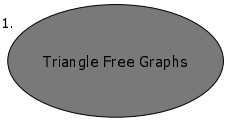
Sofya Raskhodnikova
MIT

Property - Definition

- Property – Set of Strings



e.g.:

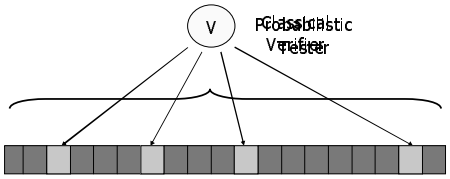
1. 

2. Satisfying assignments of a fixed CNF

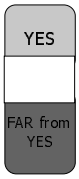
$$\bigwedge (\neg e_{i,j} \vee \neg e_{j,k} \vee \neg e_{k,i})$$

Property Testing - Definition

[Rubinfeld Sudan 96] [Goldreich Goldwasser Ron 98]



Definition - Contd



- YES: Verifier accepts with probability 1
- FAR from YES: Accepts with low probability
  - FAR – terms of hamming distance

Extensions
- 2 sided Error
- Adaptive Questions

Property Testing - Uses

- Naturally arise in contexts of PCPs
- Massive Data Sets
  - Eg: WWW, DNA samples, High Resolution Images
- Large Access Time
- Need to efficiently check if data satisfies certain properties

Testable Properties

Testable with constant number of queries
- bipartiteness [GGR 98]
- k-colorability [GGR 98]
- membership in a regular language [AKNS 99]
- Testing if function is linear [BLR 90]

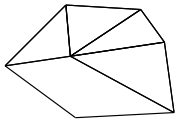
Non-Testable Properties

- [GGR 98] prove there exist properties not testable even with a linear number of queries. (Probabilistic Construction)
- Explicit Linear Lower bounds
  - 3 Colorable Bounded Degree Graphs [BOT 02]
  - Polynomials of degree n/2 represented as function evaluation [Sudan]

Testable Properties – Local Views

- Property is testable $\Rightarrow$ string far from property has lot of local views showing violation.

traingle free graphs



- Lower Bounds of [BOT 02] and [Sudan] exploit the fact that there are no small local views showing violation.

Properties as CNF formulae

Strings (length n)	Has property ?
001101.....1	✓
010101.....0	×
⋮	⋮

Each property can be represented as a CNF formula.

Triangle Free Graphs $\longrightarrow \bigwedge (\neg e_{i,j} \vee \neg e_{j,k} \vee \neg e_{k,i})$

CNF Property Testing

- CNF Property Testing:
  For a fixed CNF (i.e., property), given an assignment, is it
  - A satisfying assignment? Or
  - Far from satisfying?
- Note: Different from the testing if CNF is satisfiable or far from satisfiable.

Bounds for CNF Property Testing

- Some CNF properties - hard to test [GGR 98]
- 2CNF Property Testing: Testable with $O(\sqrt{n})$ queries [FLNRRS 02]
- What about kCNFs (k > 2)?
  - "Possibly testable": there exist "witness" of size k that falsifies kCNF.

3CNF Property Testing - Hard

Main Theorem:
There exist 3CNF formulae that require linear number of queries, even with adaptive 2-sided error tests.

kCNF ⟶ kLIN

3CNF:

$(x_1 \vee \neg x_2 \vee x_4) \wedge (\neg x_2 \vee x_3 \vee x_1) \cdots (x_{25} \vee \neg x_{10})$

$\Downarrow$

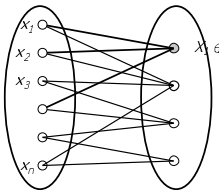
3LIN:

$(x_3 \oplus x_5 \oplus x_1) \wedge (x_2 \oplus x_3 \oplus x_1) \cdots (x_{23} \oplus x_{11})$

Advantages: Can use Linear Algebra

Linear Properties

- Defined by linear constraints. Testing membership in linear space.



V – set of vectors that satisfy all constraints.

Right degree $\leq k \Rightarrow$ V can be represented by kLIN

Lower Bound Proof (Linear property)

1. For linear property, adaptivity and 2-sided error does not help.
2. Prove sufft. properties for V to be hard for 1-sided non-adaptive tests.
3. Prove random linear spaces satisfy above properties.
4. k large in Step 3. Reduce $k \rightarrow 3$.

Adaptivity and 2-sided Error

Theorem:

For testing linear properties, adaptivity and 2-sided error do not help.

Key Idea:

- Accept only if no constraints are violated.
- To check if a linear constraint is satisfied, the order of checking the variables is immaterial.

Lower Bound Proof

Want to prove :

∀ prob. Tests T, ∃ string x far from having property,

Pr[T accepts x] is high.

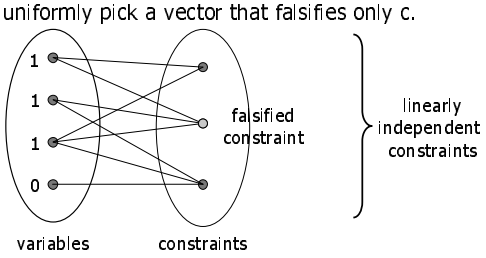
Sufft. to prove : (by Yao's MinMax Principle)

∃ bad distribution B of strings far from property,
∀ deterministic tests T,

$\Pr_{x \leftarrow B}$[T accepts x] is high

Bad Distribution - Defintion

Distribution B: uniformly pick a basis constraint c, uniformly pick a vector that falsifies only c.



Sufficient Properties – Hard to Test

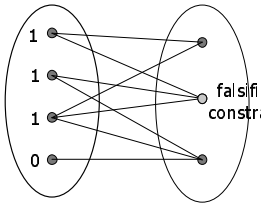
If basis constraints satisfy,

- Property 1: ε-separatedness
- Property 2: (q, μ)-locality

then, linear space is hard to test.

Property 1: ε-Separatedness

ε-separated: Any string x that falsifies exactly one basis constraint has large weight.



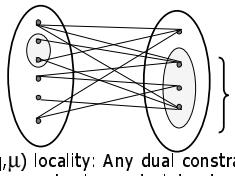
ε-separatedness

$\Downarrow$

All strings in B (bad distribution) are far from linear space V.

Property 2: (q,μ) – Locality

How can a test detect a string is from the bad distribution B?



Dual Constraint, which is a sum of large number of basis constraints, depends on few variables.

$x_2 + x_3 = 0 \pmod 2$

(q,μ) locality: Any dual constraint, that is a sum of at least μ n basis constraints, depends on more than q variables.

Probabilistic Construction

- Properties 1 & 2 are expansion-like properties. Hence, random LDPC codes satisfy Properties 1 and 2.
  - However, k (max. right degree) – large.
- Reduction $k \rightarrow 3$:

$$(x_1 \oplus x_2 \oplus \cdots x_d)$$

$$\Downarrow$$

$(x_1 \oplus x_2 \oplus \cdots x_{d/2} \oplus z)$ and $(x_{d/2+1} \oplus x_{d/2+2} \oplus \cdots x_d \oplus z)$

This reduction preserves properties 1 and 2.

Summarizing....

- For testing membership in a linear space, adaptivity and 2-sided error do not help.
- Random LDPC codes are hard to test even with a linear number of queries.
- Finally,

  There exist properties describable by 3CNFs that are hard to test with linear number of queries, even for adaptive 2-sided error tests.