Lecture 10:

Lower bounds for testing

$\Delta$-freeness:

Superpoly dependence on $\varepsilon$ is required!
A lower bound for testing $\Delta$-freeness

In a previous lecture:
- saw property test for $\Delta$-freeness
- const time in terms of $n$
- dependence on $\epsilon$ horrible - toward $1/2^c$
- is this required?

Today:
- answer this question partially (for 1-sided testers)
- when testing $H$-freeness property,
  - if $H$ bipartite, $\text{poly}(1/\epsilon)$ is enough
  - if $H$ not bipartite no $\text{poly}(1/\epsilon)$ suffices
  (We'll actually prove special case of $H=\Delta$ only)

Thm (adj matrix model)
exists $c<\infty$ s.t. any 1-sided tester for
whether graph $G$ is $\Delta$-free needs $\geq \frac{c}{1/\epsilon}$ queries.
Main Tools:

1. Goldreich-Trevisan Thm: (homework)
   - Adj matrix model
   - Property P
   - Tester T with $q(n,\epsilon)$ queries
   - $\Rightarrow$ Tester $T'$: "Natural Tester"
     - pick $q(n,\epsilon)$ nodes
     - query submatrix $\Rightarrow o(q^2)$ queries
     - nonadaptive

Consequences:

- I.b. for natural tester of $\Omega(q^2)$
  $\Rightarrow$ I.b. for any tester of $\Omega(q^2)$

- noka reduction preserves 1-sidedness

  So I.b. implication does too.
Main tools (cont.):

(2) **Additive Number Theory Lemma**

\[ \forall M \subset \mathbb{N}, \exists X \subset M = \{1, 2, \ldots, m^2 \} \]

of size \( \geq \frac{m}{e^{10 \log m}} \)

with no non-trivial solution to \( X_1 + X_2 = 2X_3 \)

\( \text{ie. } X_1 = X_2 = X_3 \text{ is the trivial solution.} \)

Will use to construct graphs such that:

- far from \( \Delta \)-free
- natural algorithm needs \( \mathcal{O}(\frac{m}{2}) \) queries

**Examples**

**Bad** \( X \):

\[ \{1, 3, 5, 9, 13, 3 \} \]

**Good** \( X \):

\[ \{1, 3, 4, 5, 7, 8, 10, 32, \ldots \} \]

\( \text{3} \) \( \text{how big??} \)

\( \text{only size } \log m \)

**Proof of lemma**

- let \( d \) be integer

\[ k \left\lfloor \frac{\log m}{\log d} \right\rfloor - 1 \]

\( \text{(later, set } d \text{ to } e^{10 \log m} \text{)} \)

\( \text{(so } k \approx \frac{\log m}{10 \log m} \approx \frac{\log m}{10} \text{)} \)
Proof of Lemma (cont.)

Define $X_B = \sum_{i=0}^{k} x_i d^i \mid x_i \leq \frac{d}{2}$ for $0 \leq i \leq k$ 

view each $x \in M$ as represented in base $d$

where $x = (x_0...x_k)$

"digits" of $x$

Claim $X_B \leq M$

Why? largest number in $X_B$

$\leq d^{k+1} \leq d \left( \left\lfloor \frac{\log_d m} \right\rfloor - 1 \right) + 1 \leq d \log_d m = m \log_d d = m$

What is $B$? Pick st. $|X_B|$ maximized

Why the constraints?

0 $x_i$'s $< \frac{d}{2}$ $\Rightarrow$ summing pairs of elements in $X_B$

doesn't generate a carry in any location!

we'll see why this is useful soon
2. will use to show that \( X_B \) is "sum-free"

Claim \( X_B \) is "sum free" i.e. \( \forall x, y, z \in X_B \) st:
\[
x + y = 2z
\]

Pf of claim assume to contrary

for \( x, y, z \in X_B \)
\[
x + y = 2z \iff \sum_{i=0}^{k} x_i a^i + \sum_{i=0}^{k} y_i a^i = 2 \sum_{i=0}^{k} z_i a^i
\]
\[
\iff x_0 + y_0 = 2z_0
\]
\[
x_1 + y_1 = 2z_1
\]
\[
\vdots
\]
\[
x_k + y_k = 2z_k
\]
\[
\left\{ \text{since no carries} \right\}
\]

Note \( \forall i \) \( x_i + y_i = 2z_i \implies \forall i \) \( x_i^2 + y_i^2 \geq 2z_i^2 \)
with equality only if \( x_i = y_i = z_i \)

Why? \( f(a) = a^2 \) is convex

use Jensen's \( \frac{\sum_{i=0}^{n} a_i f(a_i)}{n} \geq f \left( \frac{\sum_{i=0}^{n} a_i}{n} \right) \) with equality only if \( a_i's \) are all =
\[
\Rightarrow \frac{x_i^2 + y_i^2}{2} \geq \left( \frac{2z_i}{2} \right)^2 = z_i^2 \text{ with equal only if } x_i = y_i = z_i
\]
finishing proof of claim:

if \( x, y, z \) st. \( \not(x = y = z) \)

\[ \text{then } i \neq j \text{ st. } \not(x_i = y_i = z_i) \]

\[ \text{then note } \Rightarrow x_i^2 + y_i^2 > 2z_i^2 \]

\[ \text{but then: } \sum_{i \in B} x_i^2 + y_i^2 > 2 \sum z_i^2 = 2 \sum z_i^2 = 2 \sum z_i^2 \]

but how do we know that \( x_B \) is big?

* \( B \leq (kh)(\frac{d}{a})^3 < kd^2 \)

bound on digits of \( B \)

* \( |x_B| \geq (\frac{d}{a})^{kh} > (\frac{d}{a})^k \)

\[ \sum_{B \in \text{disjoint}} |x_B| \]

since \( \Rightarrow \) \( \Rightarrow \)

* \( \exists B \text{ s.t. } |x_B| \geq (\frac{d}{a})^k \)

use settings of \( d, k \), get \( |x_B| \geq \frac{m}{10^{10^{10^m}}} \)

For l.b.: Not enough! need another idea, but won't do it here.
Proof of Thm (prop testing bound)

given sum-free \( X \subseteq \{1, \ldots, m\} \)

call construct a graph:

\[
\begin{align*}
  V_1 &= \{1, \ldots, m\} \\
  V_2 &= \{1, \ldots, 2m\} \\
  V_3 &= \{1, \ldots, 3m\}
\end{align*}
\]

\[ \forall x, \bar{x} \in X \]

\[ \forall x \in X \]

will abuse notation:

node should be \((ij, \bar{i})\) \(i \in \{1, \ldots, m\} \)

will drop \( i \) if easy to see from context

\# nodes = \( 6m \) so \( m = \Theta(n) \)

\# edges = \( \Theta(m \cdot |X|) = \Theta(n^2 / e^{10Tm}) \) \( \leq \) not exactly dense
The number of cycles:

- Intended δ's: \( j, j+x, j+2x \)

- # intended δ's is \( m/\lvert X \rvert = \Theta(n^2/e^{10\sqrt{\log n}}) \)

Non-intended δ's:

- No edges internal to \( V_1, V_2 \) or \( V_3 \)

- Any δ has:
  - \( u \in V_1 \)
  - \( v \in V_2 \)
  - \( w \in V_3 \)

\[
\begin{align*}
\text{label } x_3 & \quad \text{label } x_1 & \quad \text{label } x_2 \\
\text{\( j \)} & \quad j + x_1 & \quad j + x_2 \\
\text{\( j + x_1 + x_2 \) } \Rightarrow & \quad x_1 + x_2 = 2x_3 \\
\text{\( j + x_1 + x_2 \)} & \quad \Rightarrow x_1 = x_2 = x_3 \quad \text{since } X \text{ is sum-free}
\end{align*}
\]

- No non-intended δ's

\( \square \)
• # disjoint cycles:
  all intended Δ's are disjoint (share no edges at all).
  Suppose not:

  \[ j + x = k + x' \]

  \[ x = x' \]

  \[ j + x = k + ax' \]

  Since \( x = x' \), \( k = j \) \( \Rightarrow \)

  **Distance to Δ-free**: must remove \( \geq 1 \) edge from each Δ.

  \[ \Downarrow \]

  "Absolute" distance from Δ-free = \( \Theta(\#\text{Δ's}) \)

  = \( \Theta \left( \frac{n^2}{e^{10\log n}} \right) \)

  = \( \Theta \left( m / |X| \right) \)

  **Problem needs** \( \tilde{O}(en^2) \) distance.
Idea for fix

S-blow-up \( G \rightarrow G(s) \)

- vertex in \( G \) \( \rightarrow \) size \( s \) independent set in \( G(s) \)
- edge in \( G \) \( \rightarrow \) complete bipartite graph in \( G(s) \)

Note: \( \Delta \) in \( G \) \( \Rightarrow \) \( s^3 \) \( \Delta \)'s in \( G(s) \) \( \Rightarrow \) likely to find one!

- # nodes in \( G(s) \) \( \sim m \cdot s \) (actually \( 6m \cdot s \))
- # edges \( \sim m |x| \cdot s^2 \)
- # triangles \( \sim m |x| \cdot s^3 \) (no longer disjoint)

Lemma: dist of \( G(s) \) from \( \Delta \)-free

\[ \geq \text{# edge disjoint } \Delta \text{'s} \]

\[ \geq m |x| s^2 \]

Proof: show each triangle in \( G \) \( \Rightarrow s^2 \) disjoint \( \Delta \)'s in \( G(s) \)
Given \( \varepsilon \), pick \( m \) to be largest int

satisfying

\[
\varepsilon \leq \frac{1}{e^{10\sqrt{\log m}}}
\]

This \( m \) satisfies

\[
m \geq \left( \frac{c}{\varepsilon} \right)^{c \log c/\varepsilon}
\]

Pick \( s = \left\lfloor \frac{n}{6m} \right\rfloor \approx \left\lfloor n \left( \frac{c}{\varepsilon} \right)^{c \log c/\varepsilon} \right\rfloor
\]

\[
\Rightarrow \ #\text{edges} \sim \text{distance} \sim \varepsilon \cdot n^2
\]

\[
\#\text{triangles} \sim \left( \frac{\varepsilon}{c} \right)^{c \log c/\varepsilon} \cdot n^3
\]

\[
m |x| = \frac{m}{c^{10\sqrt{\log m}}} \cdot s^3
\]

Finally, if we take a sample of size \( q \),

\[
E[\text{leads in sample}] < \left( \frac{q}{s} \right) \left( \frac{c}{\varepsilon} \right)^{c \log c/\varepsilon}
\]

\[
< 1 \quad \text{unless} \quad q > \left( \frac{c}{\varepsilon} \right)^{c \log c/\varepsilon}
\]

by Markov's \( \Rightarrow \) \( \Pr \left[ \text{see } \Delta \right] < 1 \)

But since i-sided error,

must find \( \Delta \) in order to fail