Lecture 10

Foundations of Probabilistic Proofs
Fall 2020
Alessandro Chiesa
Theorem: $NP \leq PCP[\varepsilon = 0, \delta = 0.5, \Sigma = \{0,1\}, l = \exp(n), q = O(1), r = \text{poly}(n)]$

That is, $\forall L \in NP \exists$ PCP system $(P_L, V_L)$ for $L$ that looks like this:

We can achieve soundness error $\leq 0.5$ with a constant number of queries!

Proof strategy:

1. Construct constant-query linear PCP for $NP$ (last lecture)
2. Construct a linearity test (today's lecture)
3. Linear PCP + linearity test $\rightarrow$ exponential-size PCP
From LPCP to PCP

Lemma: \( \text{LPCP} [\varepsilon_c, \varepsilon_s, \Sigma = \{0, 1\}, l, q, r] \quad \leq \quad \text{PCP} [\varepsilon_c, \varepsilon_s' = \max\{\frac{15}{16}, \varepsilon_s + \frac{1}{100}\}, \Sigma = \{0, 1\}, l' = \ell, q' = O(q \log q), r' = r + O(l \cdot \log q)] \)

The lemma lets us move from linear queries to point queries, while preserving query complexity and incurring an exponential blow-up in proof length.

This suffices for our goal:

- last time we proved \( \text{NP} \subseteq \text{LPCP} [\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0, 1\}, l = O(n^3), q = o(1), r = O(n)] \)
- via the lemma we get \( \text{NP} \subseteq \text{PCP} [\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0, 1\}, l = \exp(n), q = O(1), r = \text{poly}(n)] \)

  [the soundness error is reduced back to \( \varepsilon_s = 0.5 \) by repeating the verifier \( O(1) \) times]

We are left to prove the lemma.
First Attempt at the Lemma

**Lemma:** \[ \text{LPCP} [ \varepsilon_c, \varepsilon_s, \Sigma = \mathbb{F}, \ell, q, r] \subseteq \text{PCP} [\varepsilon_c, \varepsilon_s, \Sigma = \mathbb{F}, \ell' = \mathbb{F}^\ell, q', r'] \]

Let \((P_{\text{LPCP}}, V_{\text{LPCP}})\) be an LPCP for a language \(L\). Construct \((P_{\text{PCP}}, V_{\text{PCP}})\) as follows:

- **Prep** \((x) := \) compute \(\pi := P_{\text{LPCP}}(x) \in \mathbb{F}^\ell\)
  - output \(\Pi := \{\langle \pi, a \rangle\}_{a \in \mathbb{F}^\ell} \)

- **V_{\text{PCP}}** \((x) := \) simulate \(V_{\text{LPCP}}(x)\) by answering \(a \in \mathbb{F}^\ell\) with \(\tilde{\Pi}(a)\)

- **Completeness:** if \(x \in L\) then \(V_{\text{PCP}}\left(\tilde{\Pi}(x)\right) = V_{\text{LPCP}}(x)\) accepts w.p. \(\geq 1 - \varepsilon_c\)

- **Soundness:** if \(x \notin L\) then \(\forall \tilde{\Pi} \in \mathbb{F}^\ell\) \(\tilde{\Pi}(x) = \) ?

**Problem:** we do not know if \(\tilde{\Pi}\) is of the form \(\{\langle \pi, a \rangle\}_{a \in \mathbb{F}^\ell}\) for some \(\pi \in \mathbb{F}^\ell\)

**How to ensure that \(\tilde{\Pi}\) belongs to the set of linear functions**

\[ \text{Lin} := \left\{ f : \mathbb{F}^\ell \rightarrow \mathbb{F}^\ell \mid f \text{ is } \mathbb{F}\text{-linear} \right\} \]
Linearity Testing

A function $f: \mathbb{F}^n \to \mathbb{F}$ is \underline{linear} if $\exists c \in \mathbb{F}^n$ s.t. $f(x) = \sum_i c_i x_i$.
Equivalently, if $\forall x, y \in \mathbb{F}^n \quad f(x) + f(y) = f(x+y)$.

$$\text{ALL} = \{ f: \mathbb{F}^n \to \mathbb{F} \} \quad |\text{ALL}| = |\mathbb{F}|^{\mathbb{F}^n}$$

$$\text{LIN} = \{ f: \mathbb{F}^n \to \mathbb{F} \text{ is linear} \} \quad |\text{LIN}| = |\mathbb{F}|^{\mathbb{F}^n}$$

We want a $O(1)$-query test that, given $f \in \text{ALL}$, says YES if $f \in \text{LIN}$ and NO if $f \not\in \text{LIN}$.
But this is impossible: if $f$ differs in 1 location from $f \in \text{LIN}$ then $f \not\in \text{LIN}$
but we cannot detect this with constant soundness error.

So we relax the question: given oracle access to $f \in \text{ALL}$, say YES if $f \in \text{LIN}$ and NO if $f$ is far from $\text{LIN}$

We count in Hamming distance:

$$\Delta(f, g) = \Pr_{x \in \mathbb{F}^n} [f(x) \neq g(x)] \quad \text{and} \quad \Delta(f, S) = \min_{g \in S} \Delta(f, g).$$

Q1: can we solve the relaxed problem? Q2: if so, how does it suffice for $\text{LTF} \Rightarrow \text{P}$?
The Blum-Luby-Rubinfeld Test

A $O(1)$-query test for linearity testing:

$V_{BLR}^f : = 1. \text{ sample } x,y \in \mathbb{F}^n$

2. check that $f(x) + f(y) = f(x+y)$

randomness: 2n field elt
queries: 3 locations of $f$

Completeness: if $f \in \text{Lin}$ then $\forall x,y \in \mathbb{F}^n$, $f(x) + f(y) = f(x+y)$ so $\Pr[V_{BLR}^f = 1] = 1$

Soundness: non-trivial. E.g. if $\Delta(f, \text{Lin}) > \frac{1}{8}$ then $\Pr[V_{BLR}^f = 1] \leq 1 - \frac{1}{16}$.

Theorem: $\Pr[V_{BLR}^f = 0] \geq \min\left\{ \frac{1}{6}, \frac{1}{2} \cdot \Delta(f, \text{Lin}) \right\}$

Proof intuition:

- if $f$ is linear then each $y \in \mathbb{F}^n$ "votes" for the same value $\Delta x \in \mathbb{F}^n$, $f(x) = f(x+y) - f(y)$

- if $f$ is not linear then we can still consider, for each $x$, the most popular value:

$g_f : \mathbb{F}^n \to \mathbb{F}$ is defined as $g_f(x) := \arg \max_{v \in \mathbb{F}} \left| \left\{ y \in \mathbb{F}^n \mid v = f(x+y) - f(y) \right\} \right|$

this is the plurality value
Soundness Analysis of BLR Test - Part 1

Let \( g_f(x) := \arg \max_{y \in F} |\{y \in F^n \mid v = f(x+y)-f(y)\}| \) be the plurality correction of \( f \).

If \( g_f \) is far from \( f \) then \( V_{BLR}^f \) must reject with high probability:

**Claim:** \( \Pr[V_{BLR}^f = 0] \geq \frac{1}{2} \cdot \Delta(g_f, f) \)

**Proof:** Letting \( S = \{x \in F^n \mid \exists y \in F^n \mid f(x) \neq f(x+y)-f(y)\} \) , we get

\[
\Pr[V_{BLR}^f = 0] = \Pr_{x \in F^n}[x \notin S] \Pr_{x,y}[V_{BLR}^f = 0 \mid x \in S] + \Pr_{x \in F^n}[x \notin S] \Pr_{x,y}[V_{BLR}^f = 0 \mid x \notin S]
\]

\[
\geq |S| \cdot \min_{x \in S} \left\{ \Pr_{y \in F^n}[f(x) \neq f(x+y)-f(y)] + 0 \right\} \geq \frac{|S|}{|F|^n} \cdot \frac{1}{2}.
\]

Also, for every \( x \in S \) we have \( \Pr_{y \in F^n}[f(x) = f(x+y)-f(y)] \geq \frac{1}{2} \) so \( f(x) = g_f(x) \).

This tells us that \( \frac{|S|}{|F|^n} \geq \Delta(g_f, f) \).
Next we analyze the collision probability:

**Claim:** \( \forall x \in \mathbb{F}^n, \quad \Pr_{y \sim \mathbb{F}} \left[ f(x+y) - f(y) = f(x+z) - f(z) \right] \geq 1 - 2 \cdot \Pr_{f} \left[ \mathbb{V}_{\text{BLR}} = 0 \right] \)

**Proof:** Define \( T := \{ (y, z) \in \mathbb{F}^n \times \mathbb{F}^n \mid f(y) = f(y) + f(z-y) \} \)

\[
\Pr_{y \sim \mathbb{F}} \left[ (y, z) \in T \right] \leq 2 \cdot \Pr_{f} \left[ \mathbb{V}_{\text{BLR}} = 0 \right] \text{ because } (y, z-y) \text{ and } (x+y, z-y) \text{ are random in } \mathbb{F}^2
\]

if \((y, z) \in T\) then \( f(x+y) - f(y) = [f(x+y) + f(z-y)] - [f(z-y) + f(y)] = f(x+z) - f(z) \).

We deduce that:

\[
\Pr_{y \sim \mathbb{F}^n} \left[ g_f(x) = f(x+y) - f(y) \right] = \max_{v \in \mathbb{F}^n} \Pr_{y \sim \mathbb{F}^n} \left[ v = f(x+y) - f(y) \right]
\]

\[
\sum p_i^2 \leq \max \{ p_i \}, \quad \sum p_i = 1
\]

\[
\sum p_i^2 \leq \max \{ p_i \}, \quad \sum p_i = 1
\]

\[
\Pr_{y \sim \mathbb{F}} \left[ f(x+y) - f(y) = f(x+z) - f(z) \right] \geq \sum \Pr_{v \in \mathbb{F}^n} \left[ v = f(x+y) - f(y) \right]^2
\]

\[
\geq 1 - 2 \cdot \Pr_{f} \left[ \mathbb{V}_{\text{BLR}} = 0 \right]
\]

\[
= \Pr_{y \sim \mathbb{F}} \left[ f(x+y) - f(y) = f(x+z) - f(z) \right]
\]
Soundness Analysis of BLR Test - Part 3

Let \( g_f(x) := \arg \max_{y \in \mathbb{F}} |v - f(x+y) - f(y)| \) be the plurality correction of \( f \).

We established that \( \Pr[V_{BLR}^f = 0] \geq \frac{1}{2} \Delta(g_f, f) \) and \( \Pr[V_{BLR}^f = 0] > 1 - 2 \Pr[V_{BLR}^f = 0] \geq \frac{1}{2} \Delta(g_f, f) = \frac{1}{2} \Delta(f, \text{Lin}) \).

If \( \Pr[V_{BLR}^f = 0] \geq \frac{1}{6} \) then we are done. So assume that \( \Pr[V_{BLR}^f = 0] < \frac{1}{6} \).

We prove that \( g_f \in \text{Lin} \), so we are done as \( \Pr[V_{BLR}^f = 0] \geq \frac{1}{2} \Delta(g_f, f) = \frac{1}{2} \Delta(f, \text{Lin}) \).

Claim: if \( \Pr[V_{BLR}^f = 0] < \frac{1}{6} \) then \( \forall x, y \quad g_f(x) + g_f(y) = g_f(x+y) \)

Proof:\n\[
\frac{1}{2} \left( \Pr \left[ g_f(x) = f(x+z) - f(z) \right] \geq 1 - 2 \Pr[V_{BLR}^f = 0] > \frac{2}{3} \right) \\
\frac{1}{2} \left( \Pr \left[ g_f(y) = f(y+z) - f(z) \right] \geq 1 - 2 \Pr[V_{BLR}^f = 0] > \frac{2}{3} \right) \\
\frac{1}{2} \left( \Pr \left[ g_f(y) = f(z) - f(z-y) \right] \right) \\
\frac{1}{2} \left( \Pr \left[ g_f(x+y) = f(x+y+z) - f(z) \right] \geq 1 - 2 \Pr[V_{BLR}^f = 0] > \frac{2}{3} \right) \\
\frac{1}{2} \left( \Pr \left[ g_f(x+y) = f(x+z) + f(z) - f(x+y) \right] \right)
\]

\( \exists z^* \text{ s.t.} \)
\[
g_f(x) = f(x+z^*) - f(z^*) \\
g_f(y) = f(z^*) - f(z^* - y) \\
g_f(x+y) = f(x+z^*) - f(z^* - y) \\
\Rightarrow g_f \text{ linear at } (x, y) \in \mathbb{F}^n \)
Second Attempt at the Lemma

**Lemma:** \( \text{LPCP}[\varepsilon_c, \varepsilon_s, \Sigma = \text{FFFF}, L, q, r] \leq \text{PCP}[\varepsilon_c, \varepsilon'_s, \Sigma = \text{FFFF}, L' = \text{FFFF}, q', r'] \)

Let \((P_{\text{BLR}}, V_{\text{BLR}})\) be a LPCP for a language \(L\). Construct \((P_{\text{PCP}}, V_{\text{PCP}})\) as follows:

- \(P_{\text{PCP}}(x) := \text{compute } \pi := P_{\text{LPCP}}(x) \in \text{FFFF}^q \)
- \([\text{same as before}] \cdot \text{output } \Pi := \{<\pi, a>\}_0 \in \text{FFFF}^q \)
- \(V_{\text{PCP}}(x) := \text{check that } V_{\text{BLR}}^{\Pi} \text{ is valid and then simulate } V_{\text{LPCP}}(x) \text{ by answering } a \in \text{FFFF}^q \) with \(\Pi(a)\)

**Completeness:** if \(x \in L\) then \(V_{\text{PCP}}^{\Pi}(x) = V_{\text{BLR}}^{\Pi} \wedge V_{\text{LPCP}}^{\Pi}(x) \) accepts w.p. \(\geq 1 - \varepsilon_c\)

**Soundness:** if \(x \notin L\) then for any \(\Pi \in \text{FFFF}^q\) we have two cases:

- \(\Pi\) is \(\frac{1}{8}\)-far from \(\text{LIN} \rightarrow V_{\text{BLR}}^{\Pi}\) rejects with probability at least \(\frac{1}{16}\)
- \(\Pi\) is \(\frac{1}{8}\)-close to \(\text{LIN} \rightarrow\) let \(\hat{\Pi} = f_{\pi} \in \text{LIN}\) be closest to \(\Pi\), and note that \(\hat{\Pi}\) is unique because the distance between any two linear functions is \(\geq 1 - \frac{1}{16F^q}\)

\[\Pr[V_{\text{LPCP}}^{\Pi}(x) = 1] \leq \Pr[V_{\text{LPCP}}^{\hat{\Pi}}(x) = 1 \mid \text{all queries by } V_{\text{LPCP}} \to \hat{\Pi}] + \Pr[\exists \text{ query } a \text{ by } V_{\text{LPCP}} \to \hat{\Pi} \text{ s.t. } \hat{\Pi}(a) \neq \hat{\Pi}(a)]\]

\[\leq \varepsilon_s + q \cdot \Delta(\Pi, \hat{\Pi})\]

Assumes that each query is random but this may not be

[Indeed, none of the queries in our LPos are!]

\[\leq \varepsilon_s + q \cdot \Delta(\Pi, \hat{\Pi})\]
The Lemma via Linearity Testing and Self Correction

\[
\text{Lemma: } \text{LPCP} \left[ \varepsilon_c, \varepsilon_s, \Sigma = \text{FF}, \ell, q, r \right] \leq \text{PCP} \left[ \varepsilon_c, \varepsilon_s = \max \left\{ \frac{\varepsilon_c}{100}, \varepsilon_s + \frac{1}{100} \right\}, \Sigma = \text{FF}, \ell' = \ell \cdot \ell, q' = O(q \log q), r' = c + O(\ell \cdot \log q) \right]
\]

Let \((P_{\text{LPCP}}, V_{\text{LPCP}})\) be an LPCP for a language \(L\). Construct \((P_{\text{PCP}}, V_{\text{PCP}})\) as follows:

\[P_{\text{PCP}}(x) := \text{compute } \pi := P_{\text{LPCP}}(x) \in \text{FF}^\ell\]

\[V_{\text{PCP}}(x) := \text{check that } V_{\text{BLR}} = 1 \text{ and then simulate } V_{\text{LPCP}}(x) \text{ by answering as } \text{FF}^\ell \text{ as follows:}\]

1. for \(i = 1, \ldots, \ell\):
   - sample \(r_i \in \text{FF}^\ell\)
   - set \(v_i := \overline{\Pi}(a + r_i) - \overline{\Pi}(r_i)\)

2. answer with plurality \((v_1, \ldots, v_\ell)\)

**Self-correction**

- **Completeness:** if \(x \in L\) then

\[V_{\text{PCP}}(x) = V_{\text{BLR}} \land V_{\text{LPCP}}(x) = V_{\text{BLR}} \land V_{\text{LPCP}}(x) = 1 \land V_{\text{LPCP}}(x) \text{ accepts w.p. } \geq 1 - \varepsilon_c\]
The Lemma via Linearity Testing and Self Correction

Lemma: \( \text{LPCP}\left[ \epsilon_e, \epsilon_s, \Sigma = \mathcal{F}, \ell, g, r \right] \)

\[ \leq \text{PCP}\left[ \epsilon_e, \epsilon_s' = \max\{\frac{\epsilon}{16}, \epsilon_s + \frac{1}{100}\}, \Sigma = \mathcal{F}, \ell' = \ell \cdot \ell, q' = O(g \log g), r' = r + O(\ell \cdot \log q) \right] \]

Let \((P_{\text{pre}}, V_{\text{pre}})\) be an LPCP for a language \(L\). Construct \((P_{\text{pre}}, V_{\text{pre}})\) as follows:

\[ P_{\text{pre}}(x) = \text{compute } \pi := P_{\text{LPCP}}(x) \in \mathcal{F}^\ell \]

[same as before] \[
\text{output } \hat{\Pi} := \{<\pi, a>\}_{a \in \mathcal{F}^\ell} \in \mathcal{F}^{\ell^2}
\]

\[ \exists \forall a \in \mathcal{F}^\ell \quad \Pr[\hat{\Pi}(a) \neq \hat{\Pi}(a+r) - \hat{\Pi}(r)] \leq 2 \cdot \frac{1}{8} \]

Self-correction

- Soundness: if \(x \notin L\) then for any \(\hat{\Pi} \in \mathcal{F}^{\ell^2}\) we have two cases:
  - \(\hat{\Pi}\) is \(\frac{1}{8}\)-far from \(\text{LIN}\) \(\Rightarrow\) \(V_{\text{BLR}}\) rejects with probability at least \(\frac{1}{16}\)
  - \(\hat{\Pi}\) is \(\frac{1}{8}\)-close to \(\text{LIN}\) \(\Rightarrow\) let \(\hat{\Pi} = \pi \in \text{LIN}\) be closest to \(\hat{\Pi}\)

\[ \Pr[V_{\text{pre}}(x) = 1] \leq \Pr[V_{\text{LPCP}}(x) = 1] \text{ all queries by } V_{\text{LPCP}} \text{ to } \hat{\Pi} \]

\[ + \Pr[\exists \text{ query } a \text{ by } V_{\text{pre}} \text{ to } \hat{\Pi} \text{ s.t. } \text{sc}(\hat{\Pi})(a) \neq \hat{\Pi}(a)] \]

\[ \leq \epsilon_e + q \cdot \Pr[\text{sc}(\hat{\Pi})(a) \neq \hat{\Pi}(a)] \leq \epsilon_e + q \cdot O(\exp(-b)) \Rightarrow \text{so can take } \ell = O(\log q) \]