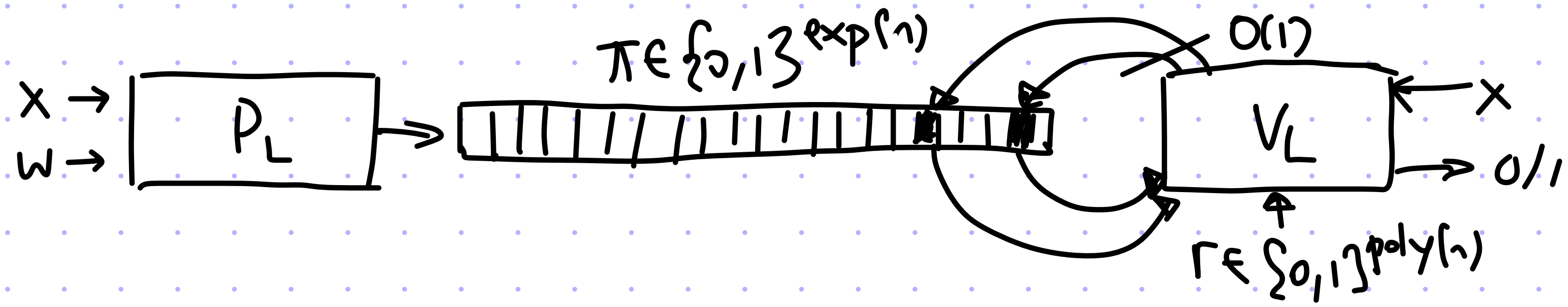
Lecture 10

Foundations of Probabilistic Proofs
Fall 2020
Alessandro Chiesa

Exponential-Size PCPs for NP

<u>theorem:</u> $NP \subseteq PCP[\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0,1\}, \ell = \exp(n), q = O(1), r = \mathrm{poly}(n)]$

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



We can achieve soundness error ≤ 0.5 with a <u>constant</u> number of queries!

<u>Proof strategy:</u>

① construct constant-query <u>linear</u> PCP for NP } last lecture

② construct a linearity test

③ linear PCP + linearity test → exponential-size PCP

} today's lecture

From LPCP to PCP

Lemma: $\mathrm{LPCP}[\varepsilon_c, \varepsilon_s, \Sigma = \mathbb{F}, \ell, q, r]$

$$\subseteq \mathrm{PCP}\left[\varepsilon_c, \varepsilon_s' = \max\left\{\tfrac{15}{16}, \varepsilon_s + \tfrac{1}{100}\right\}, \Sigma = \mathbb{F}, \ell' = \mathbb{F}^{\ell}, q' = O(q \log q), r' = r + O(\ell \cdot \log q)\right]$$

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 $NP \subseteq \mathrm{LPCP}[\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0,1\}, \ell = O(n^2), q = O(1), r = O(n)]$
- via the lemma we get $NP \subseteq \mathrm{PCP}[\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0,1\}, \ell = \exp(n), q = O(1), r = \mathrm{poly}(n)]$

  [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

<u>lemma</u>: $\mathrm{LPCP}[\varepsilon_c, \varepsilon_s, \Sigma=\mathbb{F}, \ell, q, r] \subseteq \mathrm{PCP}[\varepsilon_c, \varepsilon_s', \Sigma=\mathbb{F}, \ell'=\mathbb{F}^{\ell}, q', r']$

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

$P_{PCP}(x) :=$ • compute $\pi := P_{LPCP}(x) \in \mathbb{F}^{\ell}$
• output $\Pi := \{\langle \pi, a\rangle\}_{a\in\mathbb{F}^{\ell}} \in \mathbb{F}^{\mathbb{F}^{\ell}}$

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

- <u>Completeness:</u> if $x \in L$ then $V_{PCP}^{\Pi}(x) = V_{LPCP}^{f_\pi}(x)$ accepts w.p. $\geq 1-\varepsilon_c$
- <u>Soundness:</u> if $x \notin L$ then $\forall\, \tilde{\Pi} \in \mathbb{F}^{\mathbb{F}^{\ell}}$ $V_{PCP}^{\tilde{\Pi}}(x) = ?$

Problem: we do <u>not</u> know if $\tilde{\Pi}$ is of the form $\{\langle \tilde{\pi}, a\rangle\}_{a\in\mathbb{F}^{\ell}}$ for some $\tilde{\pi} \in \mathbb{F}^{\ell}$

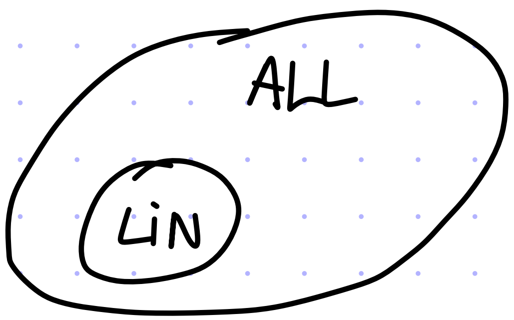
How to ensure that $\tilde{\Pi}$ belongs to the set of <u>linear functions</u>

$$\mathrm{LIN} := \{ f : \mathbb{F}^{\ell} \to \mathbb{F}^{\ell} \mid f \text{ is } \mathbb{F}\text{-linear} \} ?$$

Linearity Testing

A function $f: \mathbb{F}^n \to \mathbb{F}$ is <u>linear</u> if $\exists\, c \in \mathbb{F}^n$ s.t. $f(x) = \sum_{i=1}^{n} c_i x_i$.

Equivalently, if $\forall x, y \in \mathbb{F}^n$ $\; f(x) + f(y) = f(x+y)$.



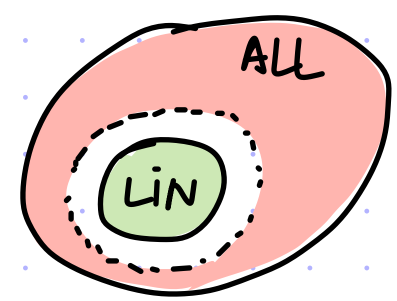
$ALL = \{ f : \mathbb{F}^n \to \mathbb{F} \}$ $\qquad |ALL| = |\mathbb{F}|^{|\mathbb{F}|^n}$

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

We want a $O(1)$-query test that, given $f \in ALL$, says YES if $f \in LIN$ and NO if $f \notin LIN$.

But this is impossible: if f differs in 1 location from $\bar{f} \in LIN$ then $f \notin LIN$
but we cannot detect this with constant soundness error.

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



an instance of a problem in Property Testing

We count in <u>Hamming distance</u>:

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

Q1: can we solve the relaxed problem? Q2: if so, how does it suffice for LPCP→PCP?

The Blum-Luby-Rubinfeld Test

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

$V_{BLR}^{f:\mathbb{F}^n \to \mathbb{F}}$:= 1. sample $x, y \in \mathbb{F}^n$
2. check that $f(x)+f(y)=f(x+y)$

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

Completeness: if $f \in LIN$ 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 $\Delta(f, LIN) \geq \frac{1}{8}$ then $\Pr[V_{BLR}^f = 1] \leq 1 - \frac{1}{16}$.

theorem: $\Pr[V_{BLR}^f = 0] \geq \min\{1/6, \frac{1}{2} \cdot \Delta(f, LIN)\}$

Proof intuition:

- if f is linear then each $y \in \mathbb{F}^n$ "votes" for the same value of x: $\forall y \in \mathbb{F}^n$, $f(x) = f(x+y) - f(y)$
- 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 $g_f(x) := \underbrace{\arg\max_{v \in \mathbb{F}} \left|\{y \in \mathbb{F}^n \mid v = f(x+y) - f(y)\}\right|}_{\text{this is the plurality value}}$

Soundness Analysis of BLR Test - Part 1

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

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

<u>claim:</u> $\Pr[V_{BLR}^f = 0] \geq \frac{1}{2} \cdot \Delta(g_f, f)$

<u>proof:</u> Letting $S = \{ x \in \mathbb{F}^n \text{ s.t. } \Pr_{y \leftarrow \mathbb{F}^n}[f(x) \neq f(x+y) - f(y)] \geq 1/2 \}$, we get

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

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

Also, for every $x \notin S$ we have $\Pr_{y \leftarrow \mathbb{F}^n}[f(x) = f(x+y) - f(y)] > \frac{1}{2}$ so $f(x) = g_f(x)$.
This tells us that $\frac{|S|}{|\mathbb{F}|^n} \geq \Delta(g_f, f)$. ∎

Soundness Analysis of BLR Test - Part 2

Next we analyze the collision probability:

<u>claim</u>: $\forall x \in \mathbb{F}^n$, $\Pr_{y,z}\left[f(x+y) - f(y) = f(x+z) - f(z) \right] \geq 1 - 2 \cdot \Pr\left[V_{BLR}^f = 0 \right]$

<u>proof</u>: Define $T := \left\{ (y,z) \in \mathbb{F}^n \times \mathbb{F}^n \;\middle|\; \begin{matrix} f(z) = f(y) + f(z-y) \\ f(x+z) = f(x+y) + f(z-y) \end{matrix} \right\}$.

- $\Pr_{y,z}\left[(y,z) \notin T \right] \leq 2 \cdot \Pr\left[V_{BLR}^f = 0 \right]$ because $(y, z-y)$ and $(x+y, z-y)$ are random in $\mathbb{F}^2$
- if $(y,z) \in T$ then $f(x+y) - f(y) = \left[f(x+y) + f(z-y) \right] - \left[f(z-y) + f(y) \right] = f(x+z) - f(z)$. ■

We deduce that:

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

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

$\sum_i p_i^2 \leq \max_i \{p_i\} \cdot \sum_i p_i$

$$= \Pr_{y,z}\left[f(x+y) - f(y) = f(x+z) - f(z) \right]$$

$$\geq 1 - 2 \cdot \Pr\left[V_{BLR}^f = 0 \right].$$

Soundness Analysis of BLR Test - Part 3

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

We established that $\Pr[V_{BLR}^f = 0] \geq \frac{1}{2}\cdot\Delta(g_f, f)$ & $\Pr_{y\in\mathbb{F}^n}[g_f(x)=f(x+y)-f(y)] \geq 1-2\cdot\Pr[V_{BLR}^f=0]$.

$> 2/3$

If $\Pr[V^f_{BLR}=0]\geq 1/6$ then we are done. So assume that $\Pr[V^f_{BLR}=0]<1/6$.

We prove that $g_f\in LIN$, so we are done as $\Pr[V^f_{BLR}=0]\geq \frac{1}{2}\Delta(g_f,f)=\frac{1}{2}\cdot\Delta(f,LIN)$.

<u>claim</u>: if $\Pr[V^f_{BLR}=0]<\frac{1}{6}$ then $\forall x,y$ $g_f(x)+g_f(y)=g_f(x+y)$

<u>proof</u>:

$$\Pr_z[g_f(x)=f(x+z)-f(z)]\geq 1-2\cdot\Pr[V^f_{BLR}=0]>\frac{2}{3}$$

$$\Pr_z[g_f(y)=f(y+z)-f(z)]\geq 1-2\cdot\Pr[V^f_{BLR}=0]>\frac{2}{3}$$

$z \leftrightarrow z-y$

$$\Pr_z[g_f(y)=f(z)-f(z-y)]$$

$$\Pr_z[g_f(x+y)=f(x+y+z)-f(z)]\geq 1-2\Pr[V^f_{BLR}=0]>\frac{2}{3}$$

$z \leftrightarrow z-y$

$$\Pr_z[g_f(x+y)=f(x+z)-f(z-y)]$$

$\exists z^*$ s.t.

$$g_f(x)=f(x+z^*)-f(z^*)$$
$$g_f(y)=f(z^*)-f(z^*-y)$$
$$g_f(x+y)=f(x+z^*)-f(z^*-y)$$

$\Rightarrow g_f$ linear at $(x,y)\in\mathbb{F}^n$!

Second Attempt at the Lemma

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

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

$P_{PCP}(x) :=$ • compute $\pi := P_{LPCP}(x) \in \mathbb{F}^{\ell}$
[same as before] • output $\Pi := \{\langle \pi, a\rangle\}_{a\in\mathbb{F}^{\ell}} \in \mathbb{F}^{\mathbb{F}^{\ell}}$

$V_{PCP}^{\tilde{\Pi}}(x) :=$ check that $V_{BLR}^{\tilde{\Pi}} = 1$ and then simulate $V_{LPCP}(x)$ by answering $a \in \mathbb{F}^{\ell}$ with $\tilde{\Pi}(a)$

- **Completeness:** if $x \in L$ then $V_{PCP}^{\Pi}(x) = V_{BLR}^{\Pi} \wedge V_{LPCP}^{f_\pi}(x)$ accepts w.p. $\geq 1-\varepsilon_c$
- **Soundness:** if $x \notin L$ then for any $\tilde{\Pi} \in \mathbb{F}^{\mathbb{F}^{\ell}}$ we have two cases:
  - $\tilde{\Pi}$ is $\frac{1}{8}$-far from LIN $\rightarrow V_{BLR}^{\tilde{\Pi}}$ rejects with probability at least $\frac{1}{16}$
  - $\tilde{\Pi}$ is $\frac{1}{8}$-close to LIN $\rightarrow$ let $\hat{\Pi} = f_{\hat{\pi}} \in$ LIN be closest to $\tilde{\Pi}$, and note that $\hat{\Pi}$ is unique because the distance between any two linear functions is $\geq 1 - \frac{1}{|\mathbb{F}|}$

$$\Pr\left[V_{LPCP}^{\boxed{\tilde{\Pi}}}(x)=1\right] \le \Pr\left[V_{LPCP}^{\hat{\Pi}}(x)=1 \,\middle|\, \begin{array}{l}\text{all queries by } V_{LPCP} \text{ to } \tilde{\Pi} \\ \text{are answered with } \hat{\Pi}\end{array}\right] + \Pr\left[\begin{array}{l}\exists \text{ query } a \text{ by } V_{LPCP} \text{ to } \tilde{\Pi} \\ \text{s.t. } \tilde{\Pi}(a) \neq \hat{\Pi}(a)\end{array}\right]$$

$$\le \varepsilon_s + q \cdot \Delta(\tilde{\Pi}, \hat{\Pi})$$

← assumes that each query is random but this may not be
[indeed, NONE of the queries in our LPCPs are!]

The Lemma via Linearity Testing and Self Correction

<u>Lemma</u>: $\mathrm{LPCP}[\varepsilon_c, \varepsilon_s, \Sigma = \mathbb{F}, \ell, q, r]$

$q' = 3 + q \cdot 2t \quad r' = r + 2\ell + t \cdot \ell$

$$\leq \mathrm{PCP}\left[\varepsilon_c, \varepsilon_s' = \max\left\{\tfrac{15}{16}, \varepsilon_s + \tfrac{1}{100}\right\}, \Sigma = \mathbb{F}, \ell' = \mathbb{F}^{\ell}, q' = O(q \log q), r' = r + O(\ell \cdot \log q)\right]$$

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

$P_{PCP}(x) :=$ • compute $\pi := P_{LPCP}(x) \in \mathbb{F}^{\ell}$
[same as before] • output $\Pi := \{\langle \pi, a \rangle\}_{a \in \mathbb{F}^{\ell}} \in \mathbb{F}^{\mathbb{F}^{\ell}}$

$V_{PCP}^{\tilde{\Pi}}(x) :=$ check that $V_{BLR}^{\tilde{\Pi}} = 1$ and then simulate $V_{LPCP}(x)$ by answering $a \in \mathbb{F}^{\ell}$ as follows:

self-correction
1. for $i = 1, \dots, t$: • sample $r_i \leftarrow \mathbb{F}^{\ell}$
   • set $v_i := \tilde{\Pi}(a + r_i) - \tilde{\Pi}(r_i)$
2. answer with $\mathrm{plurality}(v_1, \dots, v_t)$

• **Completeness**: if $x \in L$ then

$$V_{PCP}^{\Pi}(x) = V_{BLR}^{\Pi} \wedge V_{LPCP}^{sc(\Pi)}(x) = V_{BLR}^{f_\pi} \wedge V_{LPCP}^{sc(f_\pi)}(x) = 1 \wedge V_{LPCP}^{f_\pi}(x) \text{ accepts w.p. } \geq 1 - \varepsilon_c$$

The Lemma via Linearity Testing and Self Correction

lemma: $\mathrm{LPCP}[\varepsilon_c, \varepsilon_s, \Sigma = \mathbb{F}, \ell, q, r]$ $\quad q' = 3 + q \cdot 2t \quad r' = r + 2\ell + t \cdot \ell$

$$\leq \mathrm{PCP}\left[\varepsilon_c, \varepsilon_s' = \max\left\{\tfrac{15}{16}, \varepsilon_s + \tfrac{1}{100}\right\}, \Sigma = \mathbb{F}, \ell' = \mathbb{F}^{\ell}, q' = O(q \log q), r' = r + O(\ell \cdot \log q)\right]$$

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

$P_{PCP}(x) :=$ • compute $\pi := P_{LPCP}(x) \in \mathbb{F}^{\ell}$
[same as before] • output $\Pi := \{\langle \pi, a \rangle\}_{a \in \mathbb{F}^{\ell}} \in \mathbb{F}^{\mathbb{F}^{\ell}}$

$V_{PCP}^{\tilde{\Pi}}(x) :=$ check that $V_{BLR}^{\tilde{\Pi}} = 1$ and then simulate $V_{LPCP}(x)$ by answering $a \in \mathbb{F}^{\ell}$ as follows:

self-correction
1. for $i = 1, \dots, t$: • sample $r_i \leftarrow \mathbb{F}^{\ell}$
   • set $v_i := \tilde{\Pi}(a + r_i) - \tilde{\Pi}(r_i)$
2. answer with plurality$(v_1, \dots, v_t)$

(*) $\forall a \in \mathbb{F}^{\ell} \; \Pr_r[\hat{\Pi}(a) \neq \tilde{\Pi}(a+r) - \tilde{\Pi}(r)] \leq 2 \cdot \frac{1}{8}$

- **Soundness:** if $x \notin L$ then for any $\tilde{\Pi} \in \mathbb{F}^{\mathbb{F}^{\ell}}$ we have two cases:
  - $\tilde{\Pi}$ is $\frac{1}{8}$-far from LIN $\rightarrow V_{BLR}^{\tilde{\Pi}}$ rejects with probability at least $\frac{1}{16}$
  - $\tilde{\Pi}$ is $\frac{1}{8}$-close to LIN $\rightarrow$ let $\hat{\Pi} = f_{\pi} \in$ LIN be closest to $\tilde{\Pi}$

$$\Pr\left[V_{LPCP}^{\boxed{\tilde{\Pi}}}(x) = 1\right] \leq \Pr\left[V_{LPCP}^{\hat{\Pi}}(x) = 1 \;\middle|\; \begin{array}{l}\text{all queries by } V_{LPCP} \text{ to } \tilde{\Pi} \\ \text{are answered with } \hat{\Pi}\end{array}\right] + \Pr\left[\begin{array}{l}\exists \text{ query } a \text{ by } V_{LPCP} \text{ to } \tilde{\Pi} \\ \text{s.t. } sc(\tilde{\Pi})(a) \neq \hat{\Pi}(a)\end{array}\right]$$

$$\leq \varepsilon_s + q \cdot \Pr\left[sc(\tilde{\Pi})(a) \neq \hat{\Pi}(a)\right] \underset{(*)}{\leq} \varepsilon_s + q \cdot O(\exp(-t)) \Rightarrow \text{so can take } t = O(\log q)$$