Lecture B.6

Exp-size PCP

(The Hadamard PCP)

Tom Gur

Summer Graduate School on
Foundations and Frontiers of Probabilistic Proofs
August 2, 2021

Second part of the course

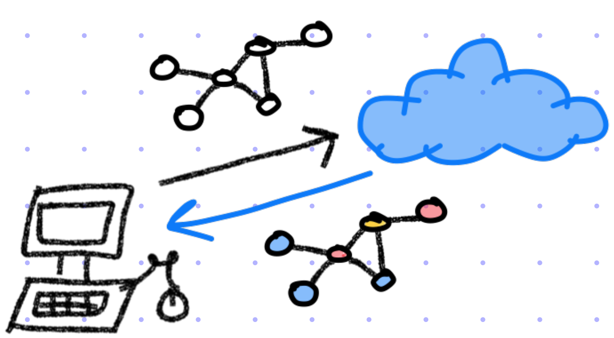
Local-to-global phenomena

- Linearity testing
- Low-degree testing
- FFT-based testing of univariate polynomials

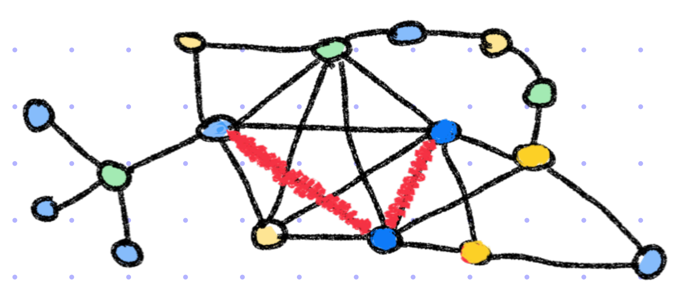
PCP constuctions

- exp-size, $O(1)$-local PCPs
- poly-size, polylog-local PCPs
- PCP composition
- Sublinear-time verification

Applications



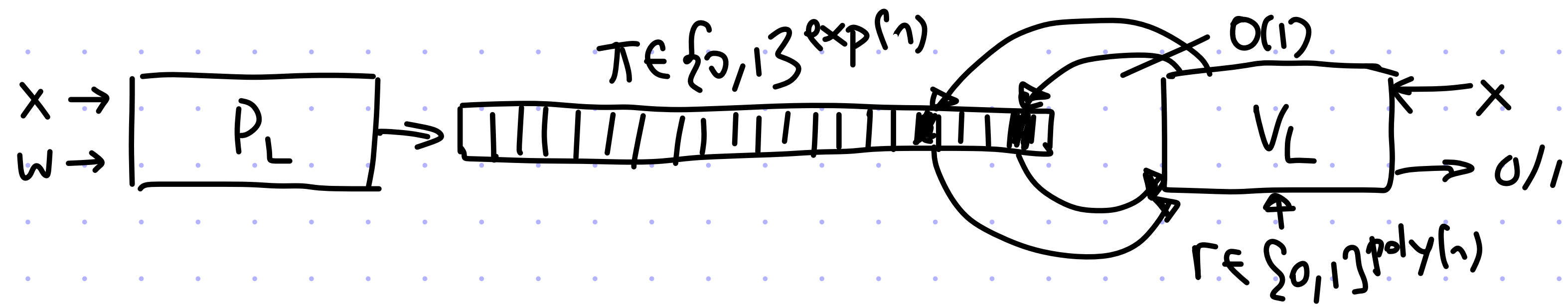
- Delegation of computation
- Hardness of approximation



Exponential-Size PCPs for NP

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

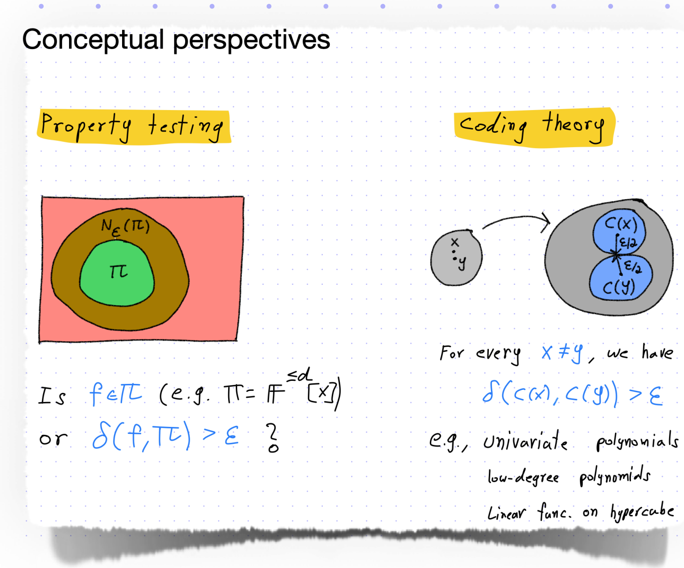
That is, $\forall L \in NP$ $\exists$ PCP system (P_L, V_L) for L that looks like this:



We can achieve soundness error ≤ 0.5 with a constant number of queries!

Proof strategy:

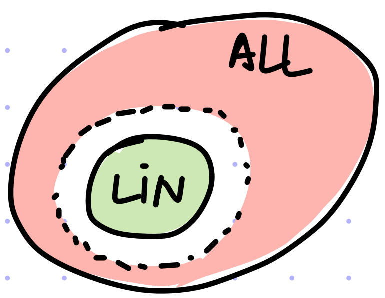
1. construct constant-query linear PCP for NP



2. construct a linearity test ✓
3. linear PCP + linearity test → exponential-size PCP

The Hadamard code

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



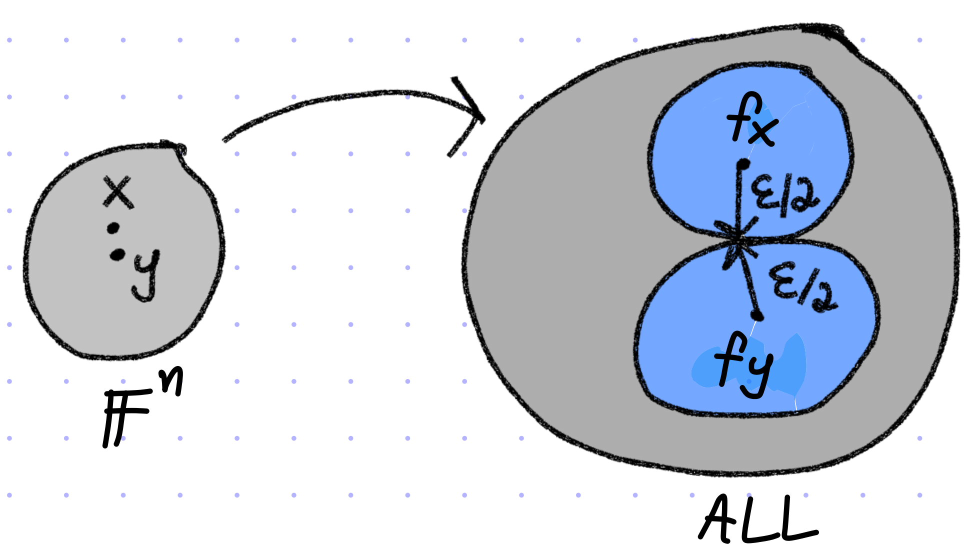
$ALL = \{ f : \mathbb{F}^n \to \mathbb{F} \}$ $\quad |ALL| = |\mathbb{F}|^{|\mathbb{F}|^n}$

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

The subspace LIN constitues the Hadamard code.

But how do we encode a message $x \in \mathbb{F}^n$ to a codeword $f_x \in LIN$?

<u>Def:</u> $\begin{cases} f_x : \mathbb{F}^n \to \mathbb{F} \\ f_x(z) = \langle x, z \rangle \end{cases}$

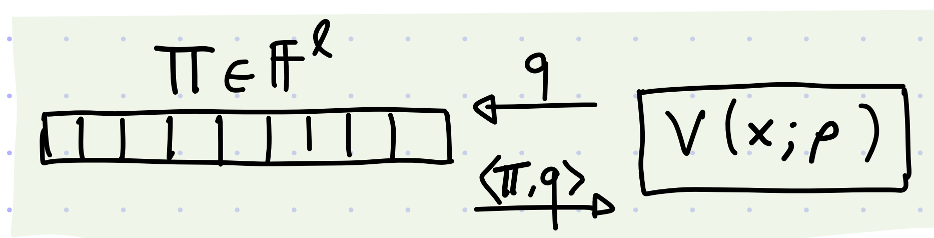


Linear PCPs

A linear PCP is a PCP where:

1. the honest proof is a linear function
2. we only consider malicious proofs that are linear functions

Given a field $\mathbb{F}$ and vector $\pi \in \mathbb{F}^\ell$, $f_\pi : \mathbb{F}^\ell \to \mathbb{F}$ is the function $f_\pi(x) := \langle \pi, x \rangle$.



def: We say that (P,V) is a LPCP system for L (over $\mathbb{F}$) if

1. completeness: $\forall x \in L$, for $\pi := P(x) \in \mathbb{F}^\ell$, $\Pr_\rho\left[V^{f_\pi}(x;\rho)=1\right] \geq 1-\varepsilon_c$
2. soundness: $\forall x \notin L$ $\forall \tilde{\pi} \in \mathbb{F}^\ell$ $\Pr_\rho\left[V^{f_{\tilde{\pi}}}(x;\rho)=1\right] \leq \varepsilon_s$.

We use similar class notation as for PCP: $\mathrm{LPCP}[\varepsilon_c, \varepsilon_s, \ell, q, r, \dots]$.

theorem: $\mathrm{NP} \subseteq \mathrm{LPCP}[\varepsilon_c = 0, \varepsilon_s = 0.5, \Sigma = \{0,1\}, \ell = O(n^2), q = O(1), r = O(n)]$

Quadratic Equations are NP-Complete

A system of m quadratic equations in n variables over $\mathbb{F}$ is a list of polynomials $p_1, \ldots, p_m \in \mathbb{F}[X_1, \ldots, X_n]$ where each p_i has total degree ≤ 2.

For example:

$$\begin{aligned} p_1 &: X_1X_3 + X_2^2 + X_6 \\ p_2 &: X_1 + X_7 - 1 \\ p_3 &: X_1X_4 + 5X_2X_3 - 7 \end{aligned}$$

def: $\mathrm{QESAT}(\mathbb{F}) = \{(p_1, \ldots, p_m) \mid \exists a_1, \ldots, a_n \in \mathbb{F} \text{ s.t. } \forall i \in [m] \;\; p_i(a_1, \ldots, a_n) = 0\}$.

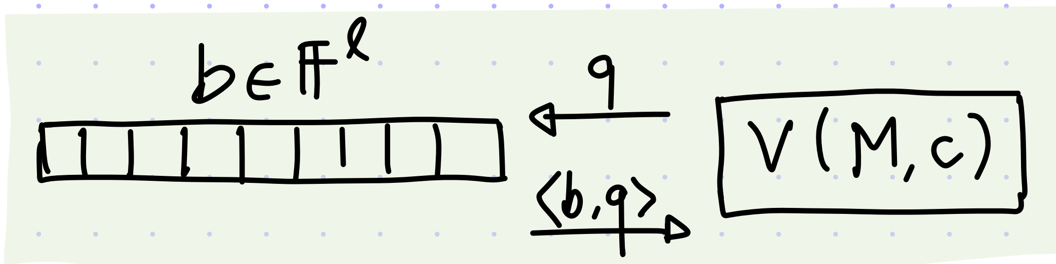
lemma: For any finite field $\mathbb{F}$, $\mathrm{QESAT}(\mathbb{F})$ is NP-complete.

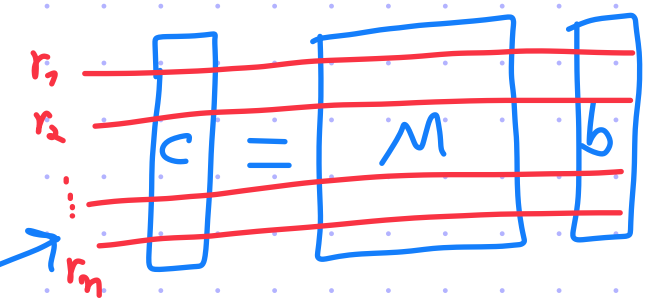
proof: Reduce from boolean circuit satisfiability (recall $\{0,1\}$ is a subset of every field):

- assign each wire a variable name: $\underbrace{X_1, \ldots, X_{n_{in}}}_{\text{inputs}}, \underbrace{X_{n_{in}+1}, \ldots, X_{n-1}}_{\text{internal}}, \underbrace{X_n}_{\text{output}}$
- use equations to enforce gates: $X_k = \mathrm{NAND}(X_i, X_j) \mapsto X_k - (1 - X_i \cdot X_j)$
- enforce booleanity: $\forall i \in [n_{in}]$, $X_i(1 - X_i) = 0$.

Warm Up 1: Linear PCP for Linear Equations

Let $M \in \mathbb{F}^{m \times \ell}$, $b \in \mathbb{F}^{\ell}$, and $c \in \mathbb{F}^{m}$, and consider this setup:





The verifier wishes to check the condition $c \stackrel{?}{=} Mb$ via linear queries.

Idea: use random linear combinations (which are linear queries)

That is: $V(M,c) :=$ sample $r \in \mathbb{F}^m$, query $b \in \mathbb{F}^{\ell}$ at $q := M^T r \in \mathbb{F}^{\ell}$ and check that $\langle c, r\rangle = \langle b, q\rangle$

Completeness: if $c = Mb$ then $\forall r \in \mathbb{F}^m$ $\langle c, r\rangle = \langle Mb, r\rangle = \langle b, M^T r\rangle = \langle b, q\rangle$.

Soundness: if $c \neq Mb$ then

Polynomial Identity Lemma applied to the non-zero poly $p(X_1, \ldots, X_m) := \sum_{i \in [m]} (c - Mb)_i X_i$

$$\Pr_r\left[\langle c, r\rangle = \langle b, q\rangle\right] = \Pr_r\left[\langle c, r\rangle = \langle b, M^T r\rangle\right] = \Pr_r\left[\langle c - Mb, r\rangle = 0\right] = \Pr_r\left[\sum_{i \in [m]} (c - Mb)_i r_i = 0\right] \leq \frac{1}{|\mathbb{F}|}. \quad \square$$

From linear to quadratic equations

Problem: How to extend the linear approach to quadratic terms?

e.g., $p(x_1, x_2, x_3) = x_1 + 2x_2 + 5x_3 + \underbrace{x_1 x_2 + 2x_2 x_3 + 2x_1 x_3 + x_1^2 + 3x_2^2 + x_3^2}$

Idea: The prover can provide the value of each quadratic monomial $z_{ij} = x_i \cdot x_j \quad \forall i, j$.

Problem: Check consistency beteen linear and quadratic terms.

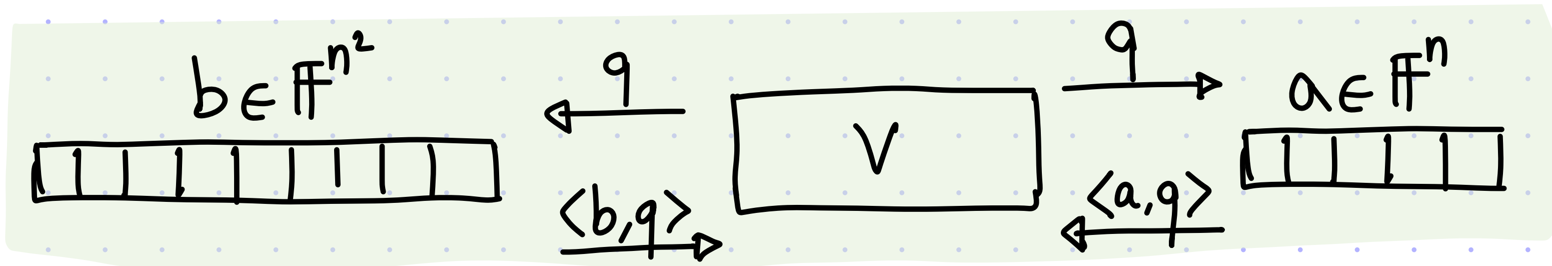
Idea: use tensor structure!

Recall that for $a \in \mathbb{F}^n$, we have $(a \otimes a)_{ij} = a_i \cdot a_j$.

Denote by $\text{flat}(a \otimes a)$ the concat. of $a \otimes a$'s rows.

Warm Up 2: Linear PCP for Tensor Structure

Let $a \in \mathbb{F}^n$ and $b \in \mathbb{F}^{n^2}$ and consider this setup:



The verifier wishes to check the condition $b \stackrel{?}{=} \mathrm{flat}(a \otimes a)$ via linear queries.

$V :=$ sample $s, t \in \mathbb{F}^n$, query b at $\mathrm{flat}(s \otimes t)$, query a at s & t, and check that $\langle b, \mathrm{flat}(s \otimes t) \rangle = \langle a, s \rangle \cdot \langle a, t \rangle$.

Completeness: if $b = \mathrm{flat}(a \otimes a)$ then $\forall\, s, t \in \mathbb{F}^n$

$$\langle b, \mathrm{flat}(s \otimes t) \rangle = \langle \mathrm{flat}(a \otimes a), \mathrm{flat}(s \otimes t) \rangle = \sum_{i,j} a_i a_j s_i t_j = \Big(\sum_i a_i s_i\Big)\Big(\sum_j a_j t_j\Big) = \langle a, s \rangle \langle a, t \rangle.$$

Soundness: if $b \neq \mathrm{flat}(a \otimes a)$ then (there is i^*, j^* s.t. $b_{i^* j^*} \neq a_{i^*} a_{j^*}$ so)

$$\Pr_{s,t}\big[\langle b, \mathrm{flat}(s \otimes t) \rangle \neq \langle a, s \rangle \cdot \langle a, t \rangle\big] = \Pr_{s,t}\Big[\sum_{ij} (b_{ij} - a_i a_j) s_i t_j \neq 0\Big] = \Pr_{s,t}\Big[\sum_i \Big(\sum_j (b_{ij} - a_i a_j) t_j\Big) s_i \neq 0\Big]$$

$$= \Pr_{s,t}\Big[\sum_i p_i(t) s_i \neq 0\Big] = \Pr_{s,t}\Big[\exists i \text{ s.t. } p_i(t) \neq 0 \ \& \sum_i p_i(t) s_i \neq 0\Big] \geq \left(\frac{|\mathbb{F}| - 1}{|\mathbb{F}|}\right)^2 \Rightarrow \text{soundness error} \leq \frac{2|\mathbb{F}| - 1}{|\mathbb{F}|^2}.$$

Linear PCP for Quadratic Equations

<u>theorem</u>: $\mathrm{QESAT}(\mathbb{F}) \in \mathrm{LPCP}\left[\varepsilon_c = 0, \varepsilon_s = \frac{2|\mathbb{F}|-1}{|\mathbb{F}|^2}, \Sigma = \mathbb{F}, \ell = n^2+n, q = 4, r = m+2n\right]$

Let $p_1, \dots, p_m \in \mathbb{F}[X_1, \dots, X_n]$ be an instance of $\mathrm{QESAT}(\mathbb{F})$.

The LPCP verifier expects a proof $\pi = (a, b) \in \mathbb{F}^{n+n^2}$ and works as follows:

$V^{f_\pi}(p_1, \dots, p_m) :=$

[$\mathrm{vec}(p_i) \in \mathbb{F}^{n+n^2}$ are non-constant coefficients of polynomial p_i]

1. sample $r \in \mathbb{F}^m$ and $s, t \in \mathbb{F}^n$
2. let $M \in \mathbb{F}^{m \times (n+n^2)}$ and $c \in \mathbb{F}^m$ be $M := \begin{bmatrix} \text{---}\ \mathrm{vec}(p_1)\ \text{---} \\ \text{---}\ \mathrm{vec}(p_2)\ \text{---} \\ \vdots \\ \text{---}\ \mathrm{vec}(p_m)\ \text{---} \end{bmatrix}$ & $c := \begin{bmatrix} \mathrm{const}(p_1) \\ \mathrm{const}(p_2) \\ \vdots \\ \mathrm{const}(p_m) \end{bmatrix}$
3. queries: $a \| b$ at $M^T r$, b at $s \otimes t$, and a at $s \,\&\, t$.
4. check that $\langle c, r \rangle = \langle a \| b, M^T r \rangle$, $\langle b, s \otimes t \rangle = \langle a, s \rangle \langle a, t \rangle$

<u>Completeness:</u> Suppose $p_1(a) = \dots = p_m(a) = 0$ and set $b := a \otimes a$. Then:

(i) $b = a \otimes a \Rightarrow$ tensor check passes w.p. 1 (ii) $M \begin{bmatrix} a \\ b \end{bmatrix} = M \begin{bmatrix} a \\ \mathrm{flat}(a \otimes a) \end{bmatrix} = c \Rightarrow$ linear check passes w.p. 1

<u>Soundness:</u> If $p_1, \dots, p_m$ have no solution then $\forall \pi = (a, b)$ either

(i) $b \neq a \otimes a \Rightarrow$ tensor check passes w.p. $\frac{2|\mathbb{F}|-1}{|\mathbb{F}|^2}$ (ii) $b = a \otimes a$ and $M \begin{bmatrix} a \\ b \end{bmatrix} \neq c$ $\Rightarrow$ linear check passes w.p. $\leq \frac{1}{|\mathbb{F}|}$

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:

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

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

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>

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

Second Attempt at the Lemma

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

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

$P_{PCP}(x) :=$ • compute $\pi := P_{LPCP}(x) \in \mathbb{F}^{\ell}$
[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 $\tilde{\Pi}(a)$

- <u>Completeness</u>: if $x \in L$ then $V_{PCP}^{\Pi}(x) = V_{BLR}^{\Pi} \wedge V_{LPCP}^{f_\pi}(x)$ accepts w.p. $\geq 1-\varepsilon_c$
- <u>Soundness</u>: 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 $\tilde{\Pi}$, and note that $\hat{\Pi}$ is unique because the distance between any two linear functions is $\geq 1 - \frac{1}{|\mathbb{F}|}$

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

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

← 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

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

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

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$(v_1, \dots, v_t)$

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

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

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

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

$P_{\mathrm{PCP}}(x) :=$ • compute $\pi := P_{\mathrm{LPCP}}(x) \in \mathbb{F}^\ell$

[same as before] • output $\Pi := \{\langle \pi, a\rangle\}_{a\in\mathbb{F}^\ell} \in \mathbb{F}^{\mathbb{F}^\ell}$

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

self-correction

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

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

- **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 $\rightarrow V_{\mathrm{BLR}}^{\tilde{\Pi}}$ rejects with probability at least $\frac{1}{16}$
  - $\tilde{\Pi}$ is $\frac{1}{8}$-close to LIN $\rightarrow$ let $\hat{\Pi} = f_\pi \in \mathrm{LIN}$ be closest to $\tilde{\Pi}$

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

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