Polynomial-Size PCPs for NP

We have constructed exponential-size PCPs for NP:

$$NP \leq PCP [\varepsilon_0 = 0, \varepsilon_0 = 0.5, \Sigma = \{0,1\}, l = \exp(n), q = O(1), r = \text{poly}(n)]$$

Our next goal is to reduce proof length to polynomial size:

**Theorem:** $NP \leq PCP [\varepsilon_0 = 0, \varepsilon_0 = 0.5, \Sigma = \{0,1\}, l = \text{poly}(n), q = \text{poly}(\log n), r = O(\log n)]$

[We will see how to further reduce $q$ to $O(1)$ towards the end of this course.]

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

Proof strategy:

1. Construct a low-degree PCP for NP
2. Construct a low-degree test
3. Low-degree PCP + low-degree test $\rightarrow$ polynomial-size PCP
Polynomial-Size PCP for Quadratic Equations

Recall the following NP-complete problem about quadratic equations over a field $\mathbb{F}$:

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

We will construct a PCP for $\text{QESAT}(\mathbb{F})$:

**Theorem:** $\text{QESAT}(\mathbb{F}) \leq \text{PCP}[\varepsilon_\text{e}=0, \varepsilon_\text{p}=0.5, \Sigma=\mathbb{F}, l=|\mathbb{F}|^{O\left(\frac{\log n}{\log \log n}\right)}, q=\text{poly}(\log n), r=O(\log n)]$

We design the PCP in several steps:

- use a small amount of randomness to reduce $m$ equations $p_1, \ldots, p_m$ to 1 equation $p$, preserving satisfiability w.h.p.
- for every possible $p$, include a proof that $p$ is satisfied by the low-degree extension of the candidate assignment
- add low-degree testing
Part 1: From m Equations to 1 Equation

**Lemma:** there is a probabilistic algorithm $T$ s.t. for $|\mathbf{F}|=\text{poly}\log(m)$

1. $T(p_1, \ldots, p_m)$ uses $O(\log m)$ random bits and outputs a quadratic equation $p(x_1, \ldots, x_n)$
2. If $\exists a, p_i(a) = \ldots = p_m(a) = 0$ then $\Pr[T(p_1, \ldots, p_m; \mathbf{r})(a) = 0] = 1$
3. If $p_1, \ldots, p_m$ are unsatisfiable then $\Pr[\exists a \ T(p_1, \ldots, p_m; \mathbf{r})(a) = 0] \leq \frac{1}{2}$.

**Idea #1:** $T$ samples $j \in [m]$ and outputs $p_j$

This uses little randomness ($\log m$ bits) but the soundness error is large $(1 - \frac{1}{m})$.

**Idea #2:** $T$ samples $r_i, \ldots, r_m \in \mathbf{F}$ and outputs $p = \sum_{j \in [m]} r_j p_j$

This has small soundness error ($\frac{1}{|\mathbf{F}|}$) but uses too much randomness ($n$ elts).

[This is essentially what we did inside the LRPC for $\text{BESAT(FF)}$.]

If we sample $r_i, \ldots, r_m \in \mathbf{F}$ the soundness error is ok ($\frac{1}{2}$) but not randomness ($n$ bits).

**Idea #3:** $T$ samples $r \in \mathbf{F}$ and outputs $p = \sum_{j \in [m]} r_j p_j$

This uses little randomness ($1$ elt) but now requires the field to be large:

the soundness error is $\frac{m}{|\mathbf{F}|}$ so we need $|\mathbf{F}| \geq \Omega(m)$. 

Lemma: there is a probabilistic algorithm $T$ s.t. for small-enough $e$:

1. $T(p_1, \ldots, p_m)$ uses $O(\log m)$ random bits and outputs a quadratic equation $p(x_1, \ldots, x_n)$
2. if $\exists \alpha$ s.t. $p_1(\alpha) = \ldots = p_m(\alpha) = 0$ then $\Pr[T(p_1, \ldots, p_m; \alpha)(\alpha) = 0] = 1$
3. if $p_1, \ldots, p_m$ are unsatisfiable then $\Pr[\exists \alpha T(p_1, \ldots, p_m; \alpha)(\alpha) = 0] \leq \frac{1}{2}$

**proof:**

Identify $[m]$ with $\mathbb{F}$ with $|\mathbb{F}| = O(\log m)$ and $e \leq \frac{\log m}{\log |\mathbb{F}|}$.

The transformation $T$ samples $r_i, \ldots, r_{se} \in \mathbb{F}$ and outputs

$$p := \sum_{0 \leq j_1, \ldots, j_{se} < |\mathbb{F}|} r_{j_1} \ldots r_{j_{se}} \cdot p_{j_1 \ldots j_{se}}.$$

The soundness error is

$$\frac{se \cdot |\mathbb{F}|}{|\mathbb{F}|} \leq O\left(\frac{(\log m)^2}{|\mathbb{F}|}\right) \Rightarrow \text{ok if } |\mathbb{F}| = \Omega(\log m^2).$$

The amount of randomness is:

$$|\mathbb{F}|^{se} = O\left(poly(\log m) \frac{\log m}{\log(\log m)}\right) = 2^{O(\log m)} = poly(m).$$
Part 2: Low-Degree PCP for 1 Equation

Consider this setting: \( P(a \in \mathbb{F}^n) \rightarrow \prod_{p \in \mathbb{F}[x_1, \ldots, x_n]} \left( \text{quadratic poly} \right) \) Is \( p \) satisfiable?

The challenge is that the polynomial \( p(x_1, \ldots, x_n) \) may depend on every variable.

Idea: reduce to a sumcheck problem & use (unrolled) sumcheck

Step 1: arithmeticize

- identify \( \mathbb{S}_n \) with \( H_n^{s_v} \) for a subset \( H_n \subseteq \mathbb{F} \) with \( |H_n| = O(\log n) \) and \( s_v := \frac{\log n}{\log |H_n|} \).
- satisfiability as a sum:

\[
\forall a: [n] \rightarrow \mathbb{F}, \quad p(a) = \sum_{i,j \in [s_v]} c_{ij} a_i a_j = \sum_{\alpha, \beta \in H_n^{s_v}} \hat{c}(\alpha, \beta) \cdot \hat{a}(\alpha) \cdot \hat{a}(\beta)
\]

where \( \hat{a}: \mathbb{F}^{s_v} \rightarrow \mathbb{F} \) & \( \hat{c}: \mathbb{F}^{2s_v} \rightarrow \mathbb{F} \) are the low-degree extensions of \( a: [n] \rightarrow \mathbb{F} \) & \( c: [n]^2 \rightarrow \mathbb{F} \).

The addend \( q(y_1, \ldots, y_{s_v}, z_1, \ldots, z_{s_v}) := \hat{c}(y, z) \hat{a}(y) \hat{a}(z) \) has individual degree \( \leq 2(|H_n| - 1) \leq 2|H_n| \).

We have reduced the problem to \( \sum_{\alpha, \beta \in H_n^{s_v}} q(\alpha, \beta)^2 = 0 \) for \( \hat{c}(y, z) \) known by the verifier and \( \hat{a} \) supplied by the prover.
Part 2: Low-Degree PCP for 1 Equation

**Step 1:** \( p(a) = 0 \iff \sum_{x, \beta \in \mathbb{H}^2} q(x, \beta) = 0 \) for \( q(y, z) = \hat{c}(y, z) \cdot \hat{a}(y) \cdot \hat{a}(z) \)

**Step 2:** probabilistically check the arithmetized statement

\[
P(p, a) \text{ outputs } \Pi := (\hat{a}, \Pi_{\text{sc}})
\]

- \( \hat{a} := \text{LDE}(a) \)
- \( \Pi_{\text{sc}} \) is eval table of IP prover for sumcheck claim
  \[ \sum_{x, \beta \in \mathbb{H}^2} q(x, \beta) = 0 \]

**Proof Length:**
- \( |\hat{a}| = 11F \cdot s^v \)
- \( |\Pi_{\text{sc}}| = O(11F^2 \cdot s^v \cdot 1 + H \cdot l) \)
- \( |\Pi_{\text{sc}}| = \text{poly}(n) \) if \( 11F = \text{polylog}(n) \)

**Completeness:** if \( p(a) = 0 \) then \( \Pi = (\text{LDE}(a), \Pi_{\text{sc}}) \) always convinces the verifier

**Soundness:** if \( p \) unsatisfiable then \( \forall \Pi = (\hat{a}, \Pi_{\text{sc}}) \)

**Low-degree PCP condition:** if \( \hat{a} \) is LDE of some \( a \) then
\[
\frac{e_s}{11F} < \frac{(2s^v) \cdot (2H \cdot l)}{11F} < O\left(\left(\frac{\log n}{11F}\right)^2\right)
\]
Low-Degree PCP for Quadratic Equations

We put Part 1 and Part 2 together:

\[ \mathbb{P}((p_1, \ldots, p_m), a) := \]

1. For every \( r \in \mathbb{F}^{S_v} \):
   - \( p_r = T(p_1, \ldots, p_m; r) \)
   - \( \mathcal{T}_{sc}[r] := \text{eval table for sumcheck to show } p_r(a) = 0 \)
   - output \( \mathcal{T}_{sc}[r] \)

2. output \( \hat{A} : \mathbb{F}^{S_v} \rightarrow \mathbb{F} \)
   \([\text{LDE } A : [n] \rightarrow \mathbb{F}]\)

\[ \mathbb{V}((p_1, \ldots, p_m)) := \]

1. Sample \( r \in \mathbb{F}^{S_v} \) and compute
   \( p_r := T(p_1, \ldots, p_m; r) \)

2. run sumcheck to check that
   \[ \sum_{\alpha, \beta \in H_v^{S_v}} \hat{C}_r(\alpha, \beta) \hat{A}(\alpha) \hat{A}(\beta) = 0 \]

**Completeness:** if \( p_1(a) = \ldots = p_m(a) \) then \( \forall r \in \mathbb{F}^{S_v} \) \( p_r(a) = 0 \) and so \( \sum_{\alpha, \beta \in H_v^{S_v}} \hat{C}_r(\alpha, \beta) \hat{A}(\alpha) \hat{A}(\beta) = 0 \)

**Soundness:** if \((p_1, \ldots, p_m)\) is unsatisfiable then, except w.p. \( \leq O \left( \frac{\log^2 m}{11F} \right) \), so is \( p_r \).

Hence, \( \forall \hat{A} \) that is LDE, \( \sum_{\alpha, \beta \in H_v^{S_v}} \hat{C}_r(\alpha, \beta) \hat{A}(\alpha) \hat{A}(\beta) \neq 0 \). So, \( \forall \mathcal{T}_{sc} \), the sumcheck accepts w.p. at most \( O \left( \frac{\log^2 m}{11F} \right) \). So \( 11F = \Omega \left( \text{poly}(\log m, \log n) \right) \) suffices.
Low-Degree Testing [statement only]

Lemma [to be proved in the next lecture]

there exists a ppt oracle machine $V_{LDT}$ s.t. $\forall f : \mathbb{F}^n \rightarrow \mathbb{F}$

1. completeness: if $f$ has total degree at most $d$ then $\mathbb{P} \left[ V_{LDT}^f(\mathbb{F},n,d)=1 \right] = 1$

2. soundness: if $f$ is $\frac{1}{10}$-far from all functions of total degree at most $d$ then $\mathbb{P} \left[ V_{LDT}^f(\mathbb{F},n,d)=1 \right] \leq \frac{1}{2}$

3. efficiency: $V_{LDT}(\mathbb{F},n,d)$ makes poly$(\|f\|,n,d)$ queries

Why total degree test?

It is simpler and we can make do with it [see next slide].

Also, there is a generic way to "lift" a total degree test to an individual degree test.

Remark: the requirement that $f$ is defined on $\mathbb{F}^n$ rather than $D^n$ for $D \subseteq \mathbb{F}$ comes from the LDT [this can be relaxed somewhat but is not easy]
At Last: PCP for Quadratic Equations

**Theorem:** $\text{QESAT}(\mathbb{F}) \leq \text{PCP}[\epsilon_c = 0, \epsilon_s = 0.5, \Sigma = \mathbb{F}, l = |\mathbb{F}|^{O(\frac{\log n}{\log \log n})}, q = \text{poly}(\log n), r = O(\log n)]$

$I((p_1, \ldots, p_m), a) :=$
1. For every $r \in \mathbb{F}^n$:
   - $p_r = T(p_1, \ldots, p_m; r)$
   - $\Pi_{sc}[r] := \text{eval table}$ for sumcheck to show $p(a) = 0$
   - Output $\Pi_{sc}[r]$
2. Output $\hat{a} : \mathbb{F}^n \rightarrow \mathbb{F}$
   - LD gadget $a : [m] \rightarrow \mathbb{F}$

$V((p_1, \ldots, p_m)) :=$
1. Sample $r \in \mathbb{F}^n$ and compute $p_r := T(p_1, \ldots, p_m; r)$
2. Run sumcheck to check that $\sum_{(\alpha, \beta) \in H_n} \mathcal{C}_r(\alpha, \beta) \hat{a}(\alpha) \hat{a}(\beta) = 0$
3. Run low-degree test on $\Pi a$
   - $V_{\text{ldt}}(\mathbb{F}, S_n, S_{1\mathbb{F}})$

1. If we can only ensure that total degree of $\hat{a}$ is $S_{1\mathbb{F}}$ then the soundness error of the term $O(S_{1\mathbb{F}})$ increases to $O(S_{1\mathbb{F}}^{2/3})$. That's ok.
2. If $\hat{a}L$ is $\frac{1}{10}$-far from LD then $V_{\text{ldt}}$ accepts w.p. $\leq \frac{1}{2}$. If $\hat{a}L$ is $\frac{1}{10}$-close to some $\hat{a}$, then we don't need self-correction! $V_{\text{sc}}$'s 2 queries are random, so pay $2 \cdot \frac{1}{10}$ in error.