Lecture 24

Limitations of PCPs and IOPs

Foundations of Probabilistic Proofs
Alessandro Chiesa

Limits on Proof Length

A PCP verifier can treat the PCP string as an MA proof string (read it in full).

In particular, $PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, vt] \subseteq MA[\varepsilon_c, \varepsilon_s, pc = \ell \cdot \log|\Sigma|, vt' = vt + \ell \cdot \log|\Sigma|]$.

Hence, PCP strings are at least as long as MA proof strings:
they inherit the limitations on proof length of MA proof strings.

We proved that $MA[\varepsilon_c, \varepsilon_s, pc, vt] \subseteq BPTIME\left(2^{O(pc)} \cdot poly\left(\frac{1}{1-\varepsilon_c-\varepsilon_s}, vt\right)\right)$.

Hence $PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, vt] \subseteq BPTIME\left(2^{O(\ell \cdot \log|\Sigma|)} \cdot poly\left(\frac{1}{1-\varepsilon_c-\varepsilon_s}, vt\right)\right)$.

We deduce that, e.g., a PCP for 3SAT with $\ell \cdot \log|\Sigma| = o(\#\text{variables})$ violates RETH.

PCPs may have ADDITIONAL limitations on proof length.

Example: For NP relations $R = \{(x,w)\}$ we have PCPs with $\ell = poly(|x|)$.
Do there exist PCPs with $\ell = poly(|w|)$?

theorem: if $SAT \in PCP[\varepsilon_c = 0, \varepsilon_s = 1/2, \Sigma = \{0,1\}, \ell = poly(\#\text{variables}), q = O(1)]$
then $NP \subseteq coNP/poly$ (and so PH collapses)

This question is related to instance compression for NP.

Limits on Query Complexity [1/2]

Recall the PCP Theorem: $NP \subseteq PCP[\varepsilon_c = 0, \varepsilon_s = 1/2, \Sigma = \{0,1\}, \ell = poly(n), q = O(1), r = O(\log n)]$.

Q: How small can query complexity be?

Hard languages are unlikely to have one-query PCPs:

lemma: $PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q=1, r] \subseteq BPTIME(2^{O(\log|\Sigma| + \log\ell)} \cdot poly(\frac{1}{1-\varepsilon_c-\varepsilon_s}, n))$.

proof: We prove that $PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q=1, r] \subseteq IP[\varepsilon_c, \varepsilon_s, K=1, pc = \log|\Sigma|, vc = \log\ell]$.

Consider the following IP protocol:

$P_{IP}(x)$		$V_{IP}(x)$
		Sample $\rho \in \{0,1\}^r$
$\pi := P_{PCP}(x) \in \Sigma^\ell$	$\xleftarrow{i}$	Compute $i := Q_{PCP}(x,\rho) \in [\ell]$.
$a := \pi[i] \in \Sigma$	$\xrightarrow{a}$	Check that $D_{PCP}(x,\rho,a) = 1$.

The lemma then follows from $IP[\varepsilon_c, \varepsilon_s, pc, vc] \subseteq BPTIME(2^{O(pc+vc)} \cdot poly(\frac{1}{1-\varepsilon_c-\varepsilon_s}, n))$. ■

EXAMPLE: a one-query PCP for 3SAT with $\log|\Sigma| = o(\#\text{variables})$ contradicts RETH.

if $\log|\Sigma| = \#$ variables then 1 symbol can encode a candidate assignment

Limits on Query Complexity [2/2]

The situation for two-query PCPs is different.

- There are NO two-query PCPs over the binary alphabet (provided the PCP is non-adaptive):

lemma: $PCP[\varepsilon_c=0, \varepsilon_s<1, \Sigma=\{0,1\}, \ell=poly(n), q=2, r=O(\log n)] \subseteq P$

proof: View a candidate PCP string as ℓ variables $z_1,\dots,z_\ell$.
The decision of $V(x;\rho)$ is a function $\phi_{x,s}(z_1,\dots,z_\ell)$ that depends on 2 variables.
- If $x\in L$ then $\exists$ assignment $a_1,\dots,a_\ell$ s.t. $\wedge_{s\in\{0,1\}^r} \phi_{x,s}(a_1,\dots,a_\ell)=1$.
- If $x\notin L$ then $\forall$ assignment $a_1,\dots,a_\ell$ $\frac{1}{2^r}\cdot|\{s \mid \phi_{x,s}(a_1,\dots,a_\ell)=1\}| \le \varepsilon_s<1$.

Deciding between these two is an instance of 2SAT, which is in P. ■

- There are two-query PCPs over larger alphabets:

lemma: $\exists c\in\mathbb{N}$ $NP \subseteq PCP[\varepsilon_c=0, \varepsilon_s=1-\frac{1}{c}, \Sigma=\{0,1\}^c, \ell=poly(n), q=2, r=O(\log n)]$

proof: Apply the trivial query bundling to the PCP Theorem.

$PCP[\varepsilon_c,\varepsilon_s,\Sigma,\ell,q,r] \subseteq PCP[\varepsilon_c,\varepsilon_s'=1-\frac{1-\varepsilon_s}{q}, \Sigma'=\Sigma^q, \ell'=O(\ell+2^r), q'=2, r'=r+\log q]$ ■

Limits on Soundness

Repeating the PCP verifier reduces soundness error but also increases query complexity.

Parallel repetition (studied in another lecture) reduces soundness error while preserving query complexity, but is efficient only for $O(1)$ repetitions (due to the blow up in proof length).

Q: Can one achieve small soundness error (sub-constant) AND small query complexity (constant)?

The prevailing belief is that soundness error ε is achievable with alphabet size $\text{poly}(\frac{1}{\varepsilon})$:

SLIDING SCALE CONJECTURE:

$$\exists\, q_0 \in \mathbb{N}\ \ \forall\, \varepsilon \geq \frac{1}{\text{poly}(n)},\ \ NP \subseteq PCP\left[\varepsilon_c = 0,\ \varepsilon_s = \varepsilon,\ \Sigma = \{0,1\}^{O(\log \frac{1}{\varepsilon})},\ \ell = \text{poly}(n),\ q = q_0,\ r = O(\log n)\right]$$

(ε: sliding parameter)

Such PCPs have applications:

- shorter succinct arguments (need fewer PCP queries for the same security level)
- improved hardness of approximation (especially if the PCP is a "projection" game)

Next we study limitations on PCP soundness.

In particular, we obtain intuition for the above conjecture.

(E.g. why can't we have $\varepsilon \leq 2^{-\sqrt{n}}$ with $|\Sigma| \leq 2^{\sqrt{n}}$?)

First Attempt At Soundness Limitations

lemma: Let $L \in PCP[\varepsilon_c = 0, \varepsilon_s, \Sigma, \ell, q, r]$. Then

$$\varepsilon_s < \min\{2^{-r}, |\Sigma|^{-q}\} \longrightarrow L \in DTime(\exp(r + q \cdot \log|\Sigma|))$$

claim: If $\forall x \notin L \; \exists \rho \in \{0,1\}^r \; \forall \pi \in \Sigma^\ell \; V^\pi(x;\rho) = 0$ then $L \in DTime(\exp(r + q \cdot \log|\Sigma|))$.

proof: By perfect completeness, $\forall x \in L \; \exists \pi \in \Sigma^\ell \; \forall \rho \in \{0,1\}^r \; V^\pi(x;\rho) = 1$.

The decider is $D(x)$:= For every $\rho \in \{0,1\}^r$: if all local views in Σ^q reject then output 0. Else output 1. ■

proof of lemma:

Suppose by contradiction that $L \notin DTime(\exp(r + q \cdot \log|\Sigma|))$. By the claim we deduce that:

- $\exists x \notin L, \rho \in \{0,1\}^r, \pi \in \Sigma^\ell$ s.t. $V^\pi(x;\rho) = 1$, so that $\varepsilon \geq 2^{-r}$
- $\exists x \notin L \; \forall \rho \in \{0,1\}^r \; \exists \pi \in \Sigma^\ell$ s.t. $V^\pi(x;\rho) = 1$, so that $\varepsilon \geq |\Sigma|^{-q}$ (pick a random local view) ■

Consider the regime $|\Sigma| = poly(n)$, $q = O(1)$, $r = O(\log n)$. Hence $r + q \cdot \log|\Sigma| = O(\log n)$.

Assuming $P \neq NP$, the lemma implies that for NP languages $\varepsilon_s \geq \max\{2^{-r}, |\Sigma|^{-q}\} \geq \frac{1}{poly(n)}$.

Q: What if we are not in this regime, or $\varepsilon_c > 0$?

Limitations for High-Soundness PCPs

A PCP verifier reads $q \cdot \log|\Sigma|$ bits from the PCP string.

For NP languages this is interesting when $q \cdot \log|\Sigma| \ll n$. (Because reading an n-bit witness achieves soundness error $\varepsilon_s = 0$.)

In this regime the soundness error must be $\Omega(2^{-q \cdot \log \ell})$.

theorem: Assuming the (randomized) exponential-time hypothesis, 3SAT does not have PCPs where $q \cdot (\log \ell + \log|\Sigma|) = o(n)$ and $\varepsilon = o(2^{-q \cdot \log \ell})$.

In particular, for $\ell = \text{poly}(n)$ and $q = O(1)$ we get $\varepsilon \geq \text{poly}(\frac{1}{n})$.

In this regime we cannot expect expenentially-small error (regardless of alphabet size).

The theorem follows from a generic lemma about ALGORITHMS FOR PCPs:

lemma: Let $L \in \text{PCP}[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q, r]$. Then

$$\varepsilon_s < (1-\varepsilon_c) \cdot 2^{-q \cdot \log \ell} \longrightarrow L \in \text{BPTIME}\left[2^{O(q \cdot (\log \ell + \log|\Sigma|))} \cdot \text{poly}\left(\frac{1}{(1-\varepsilon_c) \cdot 2^{-q \cdot \log \ell} - \varepsilon_s}, n\right)\right].$$

Proof has 2 steps: ① from PCP to laconic MA protocol

② from laconic MA protocol to BP algorithm

Step 1: from PCP to Laconic MA

can improve to 2^{-h} where $h :=$ "query entropy"

lemma: Let $L \in PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q, r]$. If $\varepsilon_s < (1-\varepsilon_c) \cdot 2^{-q \cdot \log \ell}$ then L has an MA proof with $\varepsilon_c' = 1-(1-\varepsilon_c)\cdot 2^{-q\cdot\log\ell}$, $\varepsilon_s' = \varepsilon_s$, and $pc = q \cdot (\log \ell + \log|\Sigma|)$.

proof: The MA protocol is as follows:

$P_{MA}(x)$

1. Compute $\Pi := P_{PCP}(x)$.
2. Sample random $Q \leftarrow \binom{[\ell]}{q}$.
3. Send $\pi := (Q, \Pi[Q])$.

$V_{MA}(x, \pi)$

1. Sample $\rho \in \{0,1\}^r$ and parse π as $(Q, a \in \Sigma^Q)$.
2. Run $V_{PCP}(x;\rho)$ and answer query $i \in Q$ with $a[i]$.

(Reject if any query outside of Q.)

Completeness: If $x \in L$ then, for $\Pi := P_{PCP}(x)$, $Pr_\rho[V_{PCP}^{\Pi}(x;\rho)=1] \geq 1-\varepsilon_c$.
With probability $\geq \binom{\ell}{q}^{-1} \geq 2^{-q\cdot\log\ell}$, P_{MA} samples the correct query set.
So $Pr_{Q,\rho}[V_{MA}(x,(Q,\Pi[Q]))=1] \geq (1-\varepsilon_c)\cdot 2^{-q\cdot\log\ell}$.

Soundness: Suppose that for $x \notin L$ there is $\pi = (Q, a \in \Sigma^Q)$ s.t. $Pr[V_{MA}(x,\pi)=1] > \varepsilon_s$.
For $\Pi =$ "equal to a on Q and arbitrary on $[\ell]\setminus Q$", we have $Pr[V_{PCP}^{\Pi}(x)=1] > \varepsilon_s$ (a contradiction).

Prover communication: $|\pi| = |Q| + |\Pi[Q]| = q\cdot\log\ell + q\cdot\log|\Sigma|$. ■

Concluding the Proof of the Lemma

lemma: Let $L \in PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q, r]$. Then

$$\varepsilon_s < (1-\varepsilon_c)\cdot 2^{-q\cdot\log\ell} \longrightarrow L \in BPTIME\left[2^{O(q\cdot(\log\ell+\log|\Sigma|))}\cdot poly\left(\frac{1}{(1-\varepsilon_c)\cdot 2^{-q\cdot\log\ell}-\varepsilon_s}, n\right)\right].$$

proof: The lemma is a direct implication of the two inclusions below.

① [TODAY] from PCP to laconic MA protocol:

$\varepsilon_s < (1-\varepsilon_c)\cdot 2^{-q\cdot\log\ell}$

↳ $PCP[\varepsilon_c, \varepsilon_s, \Sigma, \ell, q, r] \subseteq MA[\varepsilon_c' = 1-(1-\varepsilon_c)2^{-q\cdot\log\ell}, \varepsilon_s, pc = q\cdot(\log\ell+\log|\Sigma|), vr = r]$

② [BEFORE] from laconic MA protocol to BP algorithm:

$MA[\varepsilon_c, \varepsilon_s, pc, vr] \subseteq BPTIME\left(2^{O(pc)}\cdot poly\left(\frac{1}{1-\varepsilon_c-\varepsilon_s}, n\right)\right)$ ∎

The Case of IOPs: Proof Length

First we consider proof length.

An IOP verifier can, in particular, read each IOP prover message in full:

$$\mathrm{IOP}[\varepsilon_c, \varepsilon_s, k, \Sigma, \ell, vt] \subseteq \mathrm{IP}[\varepsilon_c, \varepsilon_s, k, pc = \ell \cdot \log|\Sigma|, vt' = vt + \ell \cdot \log|\Sigma|]$$

Hence, proof length of a private-coin/public-coin IOP $\geqslant$ prover-to-verifier communication of a private-coin/public-coin IP.

In particular IOPs inherit the relevant limitations (wrt communication) of IPs.

Unlike PCPs, some relations $R = \{(x,w)\}$ have IOPs where $\ell = \mathrm{poly}(|w|)$ rather than $\ell = \mathrm{poly}(|x|)$.

Ex: the IOP for QESAT built from LDEs and the sumcheck protocol.

But not all relations have such IOPs:

theorem: if $\mathrm{CSAT} \in \mathrm{IOP}[\varepsilon_c = 0, \varepsilon_s = 1/2, k = O(1), \Sigma = \{0,1\}, \ell = \mathrm{poly}(\#\text{inputs}), q = O(\log \#\text{gates})]$ then "some plausible conjecture about CSAT is false".

The Case of IOPs: Soundness

Can we hope for significantly smaller soundness error via IOPs compared to PCPs?

The answer is NO (to a first order).

This is because we can design similarly efficient ALGORITHMS FOR IOPs.

An IOP verifier reads $q \cdot \log|\Sigma|$ bits from the IOP strings.

For NP languages this is interesting when $q \cdot \log|\Sigma| \ll n$. (Because reading an n-bit witness achieves soundness error $\varepsilon_s = 0$.)

In this regime the soundness error must be $\Omega(2^{-q \cdot \log \ell})$.

The generic technical lemma is as follows:

lemma: Let $L \in \mathrm{IOP}[\varepsilon_c, \varepsilon_s, k, \Sigma, \ell, q, r]$ (with public-coin). Then

$$\varepsilon_s < (1-\varepsilon_c) \cdot 2^{-q \cdot \log \ell} \longrightarrow L \in \mathrm{BPTIME}\left[2^{O(q \cdot (\log \ell + \log|\Sigma|))} \cdot \mathrm{poly}\left(\frac{k}{(1-\varepsilon_c) \cdot 2^{-q \cdot \log \ell} - \varepsilon_s}, n\right)^k\right].$$

Proof has 2 steps: ① from (public-coin) IOP to laconic (public-coin) IP protocol

② from laconic (public-coin) IP protocol to BP algorithm