PCP and Inapproximability

Survey + Open Problems

Madhu Sudan
MIT

Acknowledgments

Thanks to Sanjeev Arora, Johan Hastad, Luca Trevisan, and David P. Williamson.

Overview

I. Abbreviated History.

II. PCP: Parameters and Connection.

III. Open Problems: different perspectives.

History: An Opening

- 1980's: Initial Forays into Complexity:
  - [Papadimitriou+Yannakakis]: MAX SNP
- 1990: Some unrelated developments:
  - [LFKN, Shamir]: IP = PSPACE.
  - [BFL]: MIP = NEXPTIME.
  - [BFLS]: Holographic proofs.
- 1991: Stunning development:
  - [FGLSS]: Inapproximability of Clique.
    Using MIP=NEXPTIME.

History: Widening and Strengthening

- 1992: Probabilistic checking of proofs.
  - [Arora+Safra]: Defn. of PCP.
    Characterization of NP.
  - [ALMSS]: NP $=$ PCP$[O(\log n), O(1)]$.
- 1992- : Widening the impact.
  - [ALMSS]: MAX SNP problems.
  - [Lund+Yannakakis.1]:
    Chromatic Number and Set Cover.
  - [Lund+Yannakakis.2]:
    Graph deletion problems.
  - [ABSS]: Lattice and coding problems.

PCP: The definition

- Verifier: Prob. poly-time algorithm V, with oracle-access to "proof".

- (r, q)-restricted:
  - On input x of length n.
  - Tosses $r(n)$ random coins.
  - Queries $q(n)$ bits from oracle.

- $L \in PCP_{c,s}[r, q]$ if:

  Completeness $x \in L \Rightarrow$
  $\exists\ \pi$, s.t. $V^{\pi}(x)$ accepts w.p. $\geq c$.

  Soundness $x \notin L \Rightarrow$
  $\forall\ \pi$, $V^{\pi}(x)$ accepts w.p. $< s$.

CSP view of PCP

Max q-CSP: Constraint Satisfaction Problems.

Instance: n Boolean variables;
m q-ary constraints $C_1, \ldots, C_m$.

Find: assgnt. max'ing # sat. constraints.
(or approximate this number).

PCP$[r, q] \equiv$ Max Q-CSP.

- Bits of proof = variables.
- V's verdict on fixed $$ = constraint.
- n = proof size; $m = 2^r$;
  $Q = q$ (say queries non-adaptive).
- Acc. prob. = # sat. constraints.

Whence Max 3SAT? Gadgets [Cook].

Refined PCP (rPCP)

- Refinements of query complexity.
- How are queries subdivided?
  - p: # probes into oracle.
  - a: length of oracle answers in bits.
- Can simulate p probes with a-bit answers using $q = pa$ queries, but not in other direction!
  - For fixed pa, better to have small p.
- Analogous class $rPCP_{c,s}[r, p, a]$.
- $rPCP_{c,s}[r, p, a] \equiv$ Max p-CSP on 2^a-ary variables.

Composition of PCPs [Arora+Safra]

Paradigm: Compose two PCP-like ingredients and obtain better PCP.

- Outer Verifier: $rPCP[r, p, A]$.
- Inner Verifier: "PCP-Gadget".
  - Semantics similar to those of gadgets.
  - Parameters + Constructions from PCPs.

    $\approx rPCP[R, p', a]$.
- Composed Verifier: $rPCP[r, p + p', a \circ A]$.

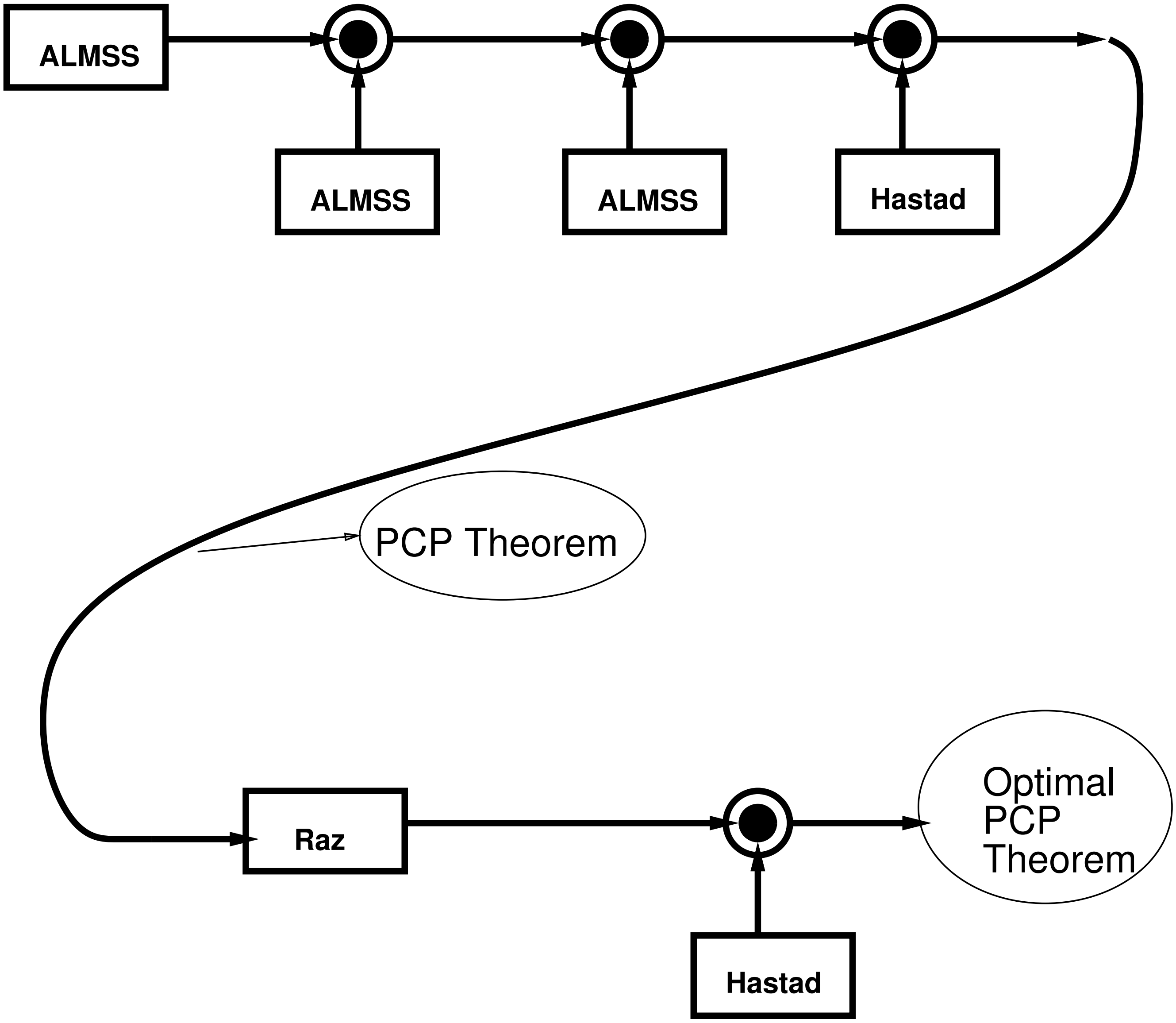
History: The Tightening

- Can PCP produce tight results?
  - [BGLR], [FK], [BS], [BGS]:
    Exploring and improving constructions.
- Meanwhile
  - [Raz]: (PCP under covers)
    The optimal outer verifier.
- PCP can produce optimal results!
  - [Feige]: Set Cover.
  - [Hastad]: Clique, Max 3SAT.
    Optimal inner verifiers.
  - [Feige+Kilian]: Chromatic Number.
  - [Samorodnitsky+Trevisan]: Max q-CSP.

Optimal PCP: Ingredients

- Composition: [Arora+Safra] $\odot$
- ALMSS-IV: (Uses BFL(S), AS, RS etc.)
  PCP with $r = \log$,mp $=$ O(1), $a = \text{poly}\log$.
- Hastad-IV: $r = 2^n, p = 3, a = 1$.
- Raz-Amplifier:
  $\text{PCP}_{1,1-\epsilon}[r, 2, a] \longrightarrow \text{PCP}_{1,(1-\epsilon')^k}[kr, 2, ka]$

Optimal PCP: Roadmap



PCP: Salient Parameters

- Query complexity? q, p, a.
  - function of c and s.
  - Fix $c = 1-, s = \frac{1}{2}+$ and minimize q.
    [Hastad]: $q = 3$ Optimal.
  - Min $\bar{q} = \lim_{k \to \infty} q_k/k$,
    s.t. $NP = PCP_{1-,2^{-k}}[\log, q_k]$.
    [Samorodnitsky+Trevisan]: $\bar{q} = 1+$.

- (Amortized) Free bit complexity f.
  (V accepts $\leq 2^f$ patterns on fixed $$.)
  - Central to Clique [FK,BS,BGS].
  - [Hastad]: $\bar{f} \to 0$.
  - $c = 1-, s = \frac{1}{2}+$, then $1 \leq f \leq 2$ (Open).

Open Problems: A Laundry List

- Vertex Cover: 2 vs. $\frac{7}{6}$.
- Max Cut: $.878$ vs. $16/17$. (Max 2SAT, Max 2-CSP).
- Coloring 3-colorable graphs: $n^{3/14}$ vs. 5!
- Graph partition problems:
  - Sparsest cut, Multi-cut: $\log$ vs. PTAS?
  - Bisection, Densest subgraph: poly vs. PTAS?
- Directed graph problems: Feedback sets.
- Geometric problems: Set cover.
- Lattices: NVP $\mathrm{poly}(n)$ approximable?
- Sequencing and scheduling:

  Max DAG, Job-shop scheduling.

Point of view: Approximability

Fact: Basic levels of approximability. P, FPTAS, PTAS, APX, $\log$-APX, poly-APX.

State of affairs:

- Know how to rule out P, FPTAS, PTAS.
- $\Omega(\log n), \Omega(n^{\epsilon})$ bounds?
- Set Cover, Clique, Chromatic Number.
- Self-improvability: Lands in $2^{\log^{1-\epsilon} n}$ land.
- Need new reductions!
- Problems of interest:
  FES, ATSP, Sparsest Cut, Job Shop.

Point of view: CSP [KSTW]

Instance: m constraints on Boolean variables.
Goal: Find assgnmt.

1. Sat. Max. # constraints (Max CSP)
2. Unsat. Min. # constraints (Min CSP)
3. Sat. all; Max. # 1's. (Max Ones)
3. Sat. all; Min. # 1's. (Min Ones)

If constraints from family $\mathcal{F}$, then obtain
(Max/Min) (CSP/Ones) $(\mathcal{F})$.

– Obtain wide variety of natural problems.
– Can talk about all uniformly.
– Compact results about coarse approx.
$(1+\alpha(n) \leq$ approx $\leq 1+O(\alpha(\text{poly}(n)))$.
– Infinitely many problems, finitely many independent open ones.

CSP list of open problems

- Min UNCUT: Partition graph into two parts so as to minimize # uncut edges [GVY]. Reduces to Multicut.
- Min 2CNF Deletion: Delete minimum number of clauses in 2CNF formula to make it satisfiable [KPRT]. Similar to FES.
- Nearest Codeword Problem: Given generator matrix of linear code and recieved word, find nearest codeword. [ABSS].
- Min Horn deletion: (Analogous to Min 2CNF Deletion, with Horn formulae)

Point of view: PCP

Current state of ignorance:

- PCP well-understood for $q \geq 3$.
- What if $q \leq 2$?
- What if $p \leq 2$, and a large?
- What if $f \leq 1$? (V's constraints are 2CNF formulae in such case.)
- For all the above, can we study $c, s = 1 - o(1)$?

Scope of these questions

- Vertex Cover $\equiv$ 1 free bit.
  [Hastad] $NP \subseteq FPCP_{1,\frac{1}{2}}[n, 1]$.
- Max CUT, Max 2SAT $\approx$ 2 query bits.
- $o(\log n)$-coloring 3-colorable graphs $\approx$ 2 query trits.
- Min UnCut $\approx$ $q = 2, s = 1 - o(1))$.
- 2CNF deletion $\approx$ $q = 2, s = 1 - o(1))$.
- Nearest codeword $\approx$ 2 query ($n$its).

Covering Soundness PCPs

- New notion of PCP.
- Usual (r, q)-restriction.
- Usual completeness.
- Unusual soundness.
  - What is the smallest k s.t.
    $\exists \pi_1, \ldots, \pi_k$ s.t.
    $\forall$ random R, $\exists i$ s.t.
    $V^{\pi_i}(x; R)$ accepts.
- [Hastad]: PCP with
  $r = \log, q = 4, c = 1, k = \omega(1)$.

Non-adaptive PCP on 3 queries

[Hastad]: $NP = PCP_{1-,\frac{1}{2}}[\log, 3]$.

[Trevisan,Zwick]: $PCP_{1,\frac{5}{8}}[\log, 3] = P$.

[Hastad]: $NP = PCP_{1,\frac{3}{4}+}[\log, 3]$.

Open: $\frac{5}{8} \leq s \leq \frac{3}{4}$.

Point of view: Anti-PCP

- Are PCPs necessary?
  - For Max q-CSP, Clique: Yes!
- For many opt. problems - No!
  - [GKRSY]: Max. Edge-disjoint paths.
  - [Khanna]: Longest Path.
  - [Guruswami+Khanna]: 3 vs. 5 coloring.
- Open: Set Cover, Metric-TSP, NVP.
- Simplify (damn) proof of PCP theorem.

Point of view: Limitations of PCP/CSP

- PCP and CSP go hand in hand.
  - CSP problems well-understood via PCP.
  - CSP problems are almost the only ones studied via PCP.

- Issues not addressed by CSP framework:
  - Global constraints: Bisection.
  - Rational objective (related to above).
  - Input-restrictions: Topology, geometry may restrict nature of inputs. Sometimes this makes problems easier: e.g. TSP [Arora].
  - Permutations vs. Combinations (ordered vs. unordered structures); sequencing vs. assignments. Real hope for progress here.

– Non-Boolean CSPs: Usual machinery can work, but technically we can't handle it. Time for another look.
– Framework for studying scheduling problems. Taxonomy exists, but does not relate to PCPs.

Conclusions

Wish list:

- New Analysis Tools: 2 query PCPs.
- New reductions: $\Omega(\log n)$ barrier.
- New framework: Sequencing problems.
- New assumptions: (E.g.
  "CVP is $\mathrm{poly}(n)$-inapproximable
  unless SZK† = BPP?")

† SZK = Statistical Zero Knowledge.