Probabilistically Checkable Proofs

3-SAT

- Given a Boolean formula, does there exist an assignment which satisfies it?

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (x_2 \vee x_4 \vee x_5) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_4)$$

- YES instance if there exists an assignment which satisfies it, else NO instance.

- If SAT instance is a YES instance, the satisfying assignment suffices to “prove” this.

- Given a polynomial sized proof, a Verifier (Turing machine) can verify in polynomial time that the instance is a YES instance.

Proof

$$(x_1 \vee \neg x_2 \vee x_3) \wedge (x_2 \vee x_4 \vee x_5) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_4)$$

- Proof is "11101", i.e., $x_1 = 1,\ x_2 = 1,\ x_3 = 1,\ x_4 = 0,\ x_5 = 1.$
- Other proofs "10011",…

- A verifier (Turing Machine) can verify in polynomial time that this formula is satisfiable.

- Does the verifier need to read the whole proof, or can the verifier make a decision after reading only $O(1)$ bits of the proof?

Verifier

- Verifier can use at most r random bits and read q locations in the proof.
- Verifier's decision should be correct with "good" probability.

- Prover writes down a "proof" X.
- Verifier uses the r independent random bits to decide upon the q random locations $l_1, l_2, \dots, l_q$ of the proof to query.
- Verifier computes $g\left(X_{l_1}, \dots, X_{l_q}\right)$ and accepts if it evaluates to 1 and rejects if it evaluates to 0.
  - $g: \{0,1\}^q \to \{0,1\}$ is some fixed function depending on the setting.

Verifier

- If YES instance, then there should exist a proof that the verifier accepts with probability at least c.
- If NO instance, ∀ proofs verifier should accept with probability at most s.

- $\mathrm{PCP}_{c,s}(r, q)$ = class of languages which have a probabilistically checkable proof with these parameters.

- We would like $q = O(1), r = O(\log n)$. Polynomial length proof.
- We would like $c - s$ to be “large”.

Assignment to Variables as a Proof

- Prover gives a “satisfying assignment” to the 3-SAT instance as proof.
- Verifier:
  - Use random bits to sample a uniform random clause from the 3-SAT instance.
  - If the assignment given by the prover satisfies this clause, then Accept, else Reject.

- Since there are m clauses, choosing a random clause requires $r = O(\log m)$.
- Since checking only one clause, $q = 3$.

Assignment to Variables as a Proof

- If 3-SAT instance is a YES instance and the prover gives a satisfying assignment of the instance,

$$c = \Pr[\text{verifier accepts}] = 1$$

- If 3-SAT instance is a NO instance, then for any proof given by the prover

$$s = \Pr[\text{verifier accepts}] \leq \frac{m-1}{m}$$

- Therefore, $c - s \neq \Omega(1)$.

$\mathrm{PCP}_{c,s}\big(\mathrm{O}(\log n), O(1)\big) \subseteq NP$

Fix $L \in \mathrm{PCP}_{c,s}\big(\mathrm{O}(\log n), O(1)\big)$. Given a x and a proof P_x that x belongs to L.

- Enumerate all 2^r possible values of the random bits and check whether $\geq c$ fraction make the verifier accept or whether $\leq s$ fraction make the verifier accept.
- Running time is $\mathrm{poly}(n)$ since $r = O(\log n)$.
- Therefore $L \in NP$.

PCP Theorem

- $\mathrm{PCP}_{1,s}\big(\mathrm{O}(\log n), O(1)\big) = NP$ for some absolute constant $s < 1$.

  [Arora, Safra – 92, Arora, Lund, Motwani, Sudan, Szegedy - 92],[Dinur - 04]

Serial repetition

- For $\mathrm{PCP}_{1,s}(r,q)$, repeat the protocol k times.
  - If verifier answered Yes in all of them, then output Yes.
  - If verifier answered No in even one of them, then output No.

- Number of random bits needed = $k \cdot r$
- Number of locations queried = $k \cdot q$
- Completeness = 1
- Soundness = s^k (H.W.)

- Therefore, $\forall k \in \mathbb{Z}^+$, $\mathrm{PCP}_{1,s}(r,q) \subseteq \mathrm{PCP}_{1,s^k}(kr,kq)$.

Hardness of Approximation of Max-g

- Let $g: \{0,1\}^q \to \{0,1\}$ be a fixed function.
- Max-g: Given a set of n variables and a set of m constraints $\left\{g\left(X_{l_1^{(i)}}, X_{l_2^{(i)}}, \dots, X_{l_q^{(i)}}\right): i \in [m]\right\}$, compute an assignment to the variables that maximizes the fraction of constraints satisfied.

- Theorem: If $3\text{SAT} \in \text{PCP}_{c,s}\left(O(\log n), O(1)\right)$ with g as the test function, then NP-hard to obtain an approximation factor better than s/c for this problem.

- Let X be the proof provided by the prover. Each value of the random coins $R \sim \{0,1\}^r$ gives q locations $l_1^{(R)}, l_2^{(R)}, \dots, l_q^{(R)}$. Consider the set of tests $\left\{g\left(X_{l_1^{(R)}}, X_{l_2^{(R)}}, \dots, X_{l_q^{(R)}}\right): R\right\}$ performed by the verifier.

- The probability of the verifier accepting is equal to the fraction of tests satisfied by the proof.
- Therefore, prover's task can be viewed as finding an assignment to the "variables" X such that the fraction of satisfied constraints in $\{g(X_{l_1^{(R)}}, X_{l_2^{(R)}}, \dots, X_{l_q^{(R)}}) : R\}$ is maximized.

- If 3SAT instance is a YES instance, then verifier accepts with probability at least c.
- Therefore, there exists an assignment to X which satisfies at least c fraction of the constraints in $\{g(X_{l_1^{(R)}}, X_{l_2^{(R)}}, \dots, X_{l_q^{(R)}}) : R\}$.

- If 3SAT instance is a NO instance, then verifier accepts with probability at most s.
- Therefore, any assignment to X will satisfy at most s fraction of the constraints in $\{g(X_{l_1^{(R)}}, X_{l_2^{(R)}}, \dots, X_{l_q^{(R)}}) : R\}$.

Hardness of Approximation

- Therefore, for $\{g(X_{l_1^{(R)}}, X_{l_2^{(R)}}, \dots, X_{l_q^{(R)}}) : R\}$, it is NP-hard to determine whether there is an assignment which satisfies at least c fraction of the constraints, or whether all assignments will satisfy at most s fraction of the constraints.

- Therefore, for Max-g, it is NP-hard to obtain any approximation algorithm with approximation factor better than s/c.

- $\text{Gap}g_{c,s}$: Given an instance of Max-g promised to be one of the two cases
  1. YES: $\exists$ an assignment to the variables satisfying $\geq c$ fraction of constraints
  2. NO: $\forall$ assignments to the variables satisfy $\leq s$ fraction of constraints

Determine which case the instance is.

Max 3-SAT

- [Hastad 01]: For every $\epsilon > 0$, and every $L \in NP$, there is a PCP with $q = 3$, $c \geq 1 - \epsilon$ and $s \leq \frac{1}{2} + \epsilon$. Moreover, the verifier chooses indices $(l_1, l_2, l_3) \sim [m]^3$ and $b \sim \{0,1\}$ according to some distribution and checks whether $X_{l_1} + X_{l_2} + X_{l_3} = b \ (mod\ 2)$.

- For any $\epsilon > 0$, $\mathrm{GapE3LIN}_{1-\epsilon,\frac{1}{2}+\epsilon}$ is NP-hard.
- A random assignment gives $\frac{1}{2}$ approximation (verify).
- A reduction from Max-E3LIN to MaxE3SAT shows that for any ϵ, $\mathrm{GapE3SAT}_{1-\epsilon,\frac{7}{8}+\epsilon}$ is NP-hard. [Hastad 01] also proved $\mathrm{GapE3SAT}_{1,\frac{7}{8}+\epsilon}$ is NP-hard.

FGLSS Graph

(Feige, Goldwasser, Lovasz, Safra, Szegedy)

- For $\mathrm{PCP}_{1,s}(r, q)$ construct a graph as follows.
- Vertices
  - $R = 2^r$ rows, each containing $\leq 2^q$ vertices. Total number of vertices $N \leq 2^{r+q}$.
  - Vertices in each row correspond to views on the queried bits that make the verifier accept.
- Edges
  - Add edges between vertices that represent consistent partial assignments.

- Any row is an independent set. Two rows that query different set of indices have a complete bipartite graph.

- Lemma: If the graph has a clique of size M, then the probability of accepting is at least $M/2^r$ (H.W.).
  - The partial assignment of any two vertices in the clique are consistent with each other.

- There is no $1/s$ approximation for max-clique assuming $3\text{SAT} \in \text{PCP}_{1,s}(r, q)$.
- There is no $1/s^k$ approximation for max-clique assuming $3\text{SAT} \in \text{PCP}_{1,s^k}(kr, kq)$.

- Theorem: $\forall \epsilon > 0$, there is no $1/2^{\log^{1-\epsilon} N}$ approximation for max-clique assuming $\text{NP} \not\subseteq \cup_{t>0} \text{DTIME}\left(2^{\log^t n}\right)$.
- Choosing $k = \log^t n$, $N \approx 2^{k(r+q)} = 2^{O\left(\log^{t+1} n\right)}$.
- Gap $= 1/s^k = 1/2^{O\left(\log^t n\right)} = 1/2^{O\left(\log^{\frac{t}{t+1}} N\right)}$.