6.841 Advanced Complexity Theory April 22, 2009

Lecture 20

Lecturer: Madhu Sudan *Scribe: Kristian Brander*

1 Introduction

The purpose of todays lecture is to explore the landscape of PCPs. In particular three constructions of Dinur, Raz and Håstad will be surveyed.

2 Recap from previous lectures

2.1 The PCP class

A language L is in $\mathrm{PCP}_{c,s}(r,q)$ if there exists a verifier V such that

- If $x \in L$ there exists a proof π, such that V accepts with probability $c(n)$.
- If $x \notin L$ then for all proofs π, V accepts with probability at most $s(n)$.

2.2 Adaptive and non–adaptive verifiers

One distinguishes between adaptive and non–adaptive verifiers. For adaptive verifier its ith query can depend on "the past" i.e. on the previous queries. For non–adaptive verifiers, the positions in the proof to be queried should be read simultaneously.

By definition adaptive verifiers are stronger than non–adaptive ones. An adaptive verifiers decision of which position to query next (given the past), may be represented as a decision tree, and therefore by querying all nodes at once, an adaptive verifier with q queries can be converted into a non–adaptive verifier using 2^q queries.

Though less powerful, the non–adaptive verifiers have the advantage of being simpler to reason about and in fact the best known PCP constructions have non–adaptive verifiers. However the two types of verifiers really have different properties as the combination of the following results show. Håstad in [Hås01] and Guruswami, Lewin, Sudan and Trevisan in [GLST98] shows that

$$\mathrm{NP} \subseteq \mathrm{PCP}_{1,0.51}\left[\mathcal{O}(\log n), 3\right],$$

which can be compared to a result of Trevisan and Zwick

$$\mathrm{NAPCP}_{1,0.51}(\mathcal{O}(\log n, 3) \subseteq P.$$

Here $\mathrm{NAPCP}_{c,s}(r,q)$ denotes the class of languages that has non–adaptive PCP verifiers.

2.3 Generalized graph coloring

Let G with vertex set V and edges $E \subseteq V^t$ where t is an integer. Given k colors and constraints $\{\pi_e : \{1,\ldots,k\}^t \to \{0,1\}\}_{e\in E}$, the gap–generalized hypergraph coloring problem $\mathrm{GGHC}_{c,s}(t,k)$ is to find an assignment of colors $A : V \to \{1,\ldots,k\}$ satisfying the constraints. $\mathrm{UNSAT}(G)$ is the minimal fraction of unsatisfied edges for any assignment of colors.

For a given graph G, one may try to separate the two cases $\mathrm{UNSAT}(G) \leq 1-c$ and $\mathrm{UNSAT}(G) \geq 1-s$, and this establishes a connection between hypergraph coloring and PCPs: To a proof one associates a hypergraph whose vertices are the bits of the proof and the hyperedges are queries. A proof can then be thought of as a coloring of the hypergraph, and the checks done by the verifier, as checking whether certain edge constraints π_e are satisfied. One can prove that

$$\begin{aligned} \mathrm{GGHC}_{c,s}(t,k) &\subseteq \mathrm{NAPCP}_{c,s}(\mathcal{O}(\log n), t\log k) \\ \mathrm{NAPCP}_{c,s}(\log n, q) &\subseteq \mathrm{GGHC}_{c,s}(q,2) \end{aligned}$$

Thus hypergraph coloring problems may be translated to PCPs and vice versa, with appropriate care in the choice of parameters.

2.4 Hypergraph coloring and coloring of (ordinary) bipartite graphs

Generalized hypergraph coloring can be reduced to coloring ordinary (bipartite) graphs, at the expense of having to deal with more colors, more specifically

$$\mathrm{GGHC}_{1,1-\varepsilon}(t,k) \leq \mathrm{GeneralizedGraphColoring}_{1,1-\frac{\varepsilon}{t}}(k^t).$$

This is done as follows: For a t–regular hypergraph $G=(V,E)$ with constraints π_e, construct the ordinary bipartite graph G' as follows: Let the left vertex set of G' be V and the right vertex set be E and connect a pair of vertices (v,e) in G' if $v \in e$ in the original graph G. The constraint on G' are now on k^t colors and are saying that the coloring of G' should be "consistent" with the constraints on G. More specifically, an assignment of colors will satisfy the constraint $\pi'_{(v,e)}$ for an edge $(v=v_i, e=(v_1,\ldots,v_t))$ if the color of v_i and the i–th component of the color of e agree, and if the coloring of e satisfies the constraint π_e from the original graph. A coloring of the hypergraph implies a coloring of the associated bipartite graph, and so the transformation is consistency preserving. Checking the claim on the soundness is left as an exersice. The transformation first appeared in [Fortnow, Rompel and Sipser].

3 Dinur's construction

The main ingredients in Dinur's proof of the PCP theorem [Din06] are the following two lemmas

Lemma 1 (Amplification Lemma) *$\forall C \exists K, \varepsilon$ such that there is a linear time transformation from*

$$\mathit{GeneralizedGraphColoring}(k) \to \mathit{GeneralizedGraphColoring}(K),$$

mapping G to $\tilde{G}$ such that

- *$\mathit{UNSAT}(G) = 0$ implies $\mathit{UNSAT}(\tilde{G}) = 0$.*
- *$\min\{\varepsilon, C \cdot \mathit{UNSAT}(G)\} \leq \mathit{UNSAT}(\tilde{G})$*

For comprehensibility, one may ignore the appearance of ε in the above lemma, and simply read it as a statement saying that *there is a linear time transformation which amplifies gaps.*

Lemma 2 (K–reduction Lemma) *$\exists \delta \forall K$ there is a transformation from $\mathit{GeneralizedGraphColoring}(K)$ to $\mathit{GeneralizedGraphColoring}(k)$ mapping G to $\tilde{G}$, such that*

$$\delta \cdot \mathit{UNSAT}(G) \leq \mathit{UNSAT}(\tilde{G}).$$

With these two lemmas at hand, Dinur's proof consists of combining the parameters (and quantifiers) in the above statements.

There exists "stronger" versions of the PCP theorem, where more attention is paid to the ingoing constant. The current state of the art is [Moshkovitz, Raz], which shows that

$$\mathrm{SAT} \in \mathrm{PCP}_{1-\varepsilon,1/2}((1+o(1))\log_2 n, 3).$$

We will not go into the details of this result, but instead survey other constructions by Raz and Håstad.

4 Raz's construction

Raz's PCP theorem may be stated as a transformation between two instances of the generalized graph coloring problem for bipartite graphs.

Theorem 3 *$\forall s, k \exists \bar{s}$, such that $\forall t$ it holds that*

$$BipartiteGeneralizedGraphColoring_{c,s}(k) \leq BipartiteGeneralizedGraphColoring_{c^t,\bar{s}^t}(k^t).$$

Furthermore, if $s < 1$ then $\bar{s} < 1$.

Let $G = (V = L \cup R, E)$ be a bipartite graph, where L and R denotes the left and right vertex set of G respectively. The proof of the theorem relies on repeating the graph "t–fold", by constructing a new bipartite graph $\tilde{G}$ with left vertex set L^t and right vertex set R^t and connecting two nodes $(u_1, \ldots, u_t) \in L^t$ to $(v_1, \ldots, v_t) \in R^t$ if u_i and v_i are connected in G, for all i. The edge constraints in the new graph are conjunctions of the edge constraints from G:

$$\pi_{((u_1,\ldots,u_t),(v_1,\ldots,v_t))} = \bigwedge_i \pi_{(u_i,v_i)}.$$

This is what gives the exponent t on the consistency parameter in the theorem. The hard part is to bound the soundness, details of the analysis can be found in a survey by Holstein.

By the way we translated PCPs into graph coloring problems, where the verifier probes a (small) number of random edges, decreasing soundness is like increasing UNSAT. If $s < 1$, increasing t makes the soundness go down exponentially, and therefore one may think of Raz's result as a parallel to Dinur's gap amplification theorem. The main differences are:

- In Dinur's construction the absolute constant ε bounds how UNSAT increases with repeated transformations. In contrast, with Raz's construction the soundness can be made arbitrarily close to 0.
- In Dinur's transformation the size of the resulting graph is $\mathcal{O}(n)$, whereas for Raz's construction it grows exponentially with t.

Therefore Dinur's gap amplification and Raz's transformation, are not equivalent, but are similar, with different tradeoffs between soundness and size.

5 Håstad's construction

Håstad's PCP theorem can again be viewed as a result on generalized graph coloring.

Theorem 4 *For all $\delta > 0$ there exists $\varepsilon > 0$ such that*

$$GeneralizedGraphColoring_{1,\delta}(K) \leq GGHC_{1-\varepsilon,1/2}(3, 2).$$

The result is a translation from coloring with many (K) colors to coloring with only 2 colors and thus it may be viewed as a parallel to Dinur's K–reduction lemma. The theorem implies that

$$\mathrm{NP} \subseteq \mathrm{PCP}_{1-\varepsilon,1/2+\varepsilon}(\mathcal{O}(\log n), 3).$$

The proof of the theorem consists of giving a reduction from bipartite graph coloring with K colors (and so–called projective constraints) to colorings of 3–regular hypergraphs using only 2 colors. A constraint is said to be projective if $\pi_(u, v)$ only depends on v, i.e. if it is *function* of this vertex.

Let $f : \{1, \ldots, K\} \to \{-1, 1\}$ be a function (here $\{-1, 1\}$ should be thought of as a bit, and this representation is chosen to facilitate the analysis using Fourier analysis). The value of the function $-f$ at any point is just

minus the value of f at the same point. Therefore f and $-f$ may be considered equivalent, and modulo this "equivalence relation" there are 2^{K-1} functions from $\{1,\ldots,K\}$ to $\{-1,1\}$.

From G a new bipartite graph G' is constructed by expanding each vertex v (both on the left and the right side) to a cloud of 2^{K-1} vertices, indexed by (u,f) where f is a function of the above type. One should think of each cloud U as a function table holding values $U[f] = f(\mathrm{Color}(u))$, except at the positions with errors. The verifier will then be checking relations between the functions given by these tables.

The verifier will be asking three questions of the form "what is the value of $f(\mathrm{Color}(u))$?". It will ask three such questions on an "edge" between one vertex on the left vertex set and two vertices on the right vertex set. The edge will be chosen randomly, by first taking (u,f) randomly from the left vertex set, then (v,g) randomly from the right vertex set and then finally $(v, f\circ\pi_e)$, again from the right set, where π_e is the edge constraint. If the chosen vertices are error–free it will hold that

$$\mathrm{Color}(v,g)\cdot\mathrm{Color}(u,f) = \mathrm{Color}(v,(f\circ\pi_e)\cdot g),$$

and the verifier could use this relation as its "check" on the edge. However, these checks are not quite strong enough for the verifier. Instead checks of the form

$$\mathrm{Color}(v,g)\cdot\mathrm{Color}(u,f) = \mathrm{Color}(v,(f\circ\pi_e)\cdot g\cdot\eta),$$

are used, where $\eta : \{1,\ldots,K\}\to\{-1,1\}$ is some noisy function such that $\eta(i)=1$ with high probability $1-\varepsilon$ and $\eta(i)=-1$ with probability ε. This scheme suffices to give the result stated in the theorem. The analysis of the consistency and soundness parameters relies on Fourier analysis of boolean functions.

References

[Din06] Irit Dinur. The PCP theorem by gap amplification. In *Proc. 38th ACM Symp. on Theory of Computing*, pages 241–250, 2006.

[GLST98] Venkatesan Guruswami, Daniel Lewin, Madhu Sudan, and Luca Trevisan. A tight characterization of NP with 3 query PCPs. In *FOCS*, pages 8–17, 1998.

[Hås01] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.