<draft>Lecture 11: Tight inapproximability, Parallel Repetition and Agreement Tests

Irit Dinur

January 18, 2023

In this lecture we will talk about tight inapproximability of constraint satisfaction problems. We will define the label cover problem and outline the general approach for proving tight inapproximability results.

We will then show how the basic PCP theorem, together with a direct product theorem called "parallel repetition", shows hardness of label cover. The proof of this will turn out to involve so-called agreement tests, which we have seen in earlier lectures.

1 Hardness of Approximation

Constraint satisfaction problems are given by a collection of tuples and predicates. Examples include 3-SAT, 3-LIN, max-cut, 3-coloring, and many more. Given a CSP instance, for example a 3SAT formula φ, we define $\mathrm{val}(\varphi)$ to be the maximal fraction of satisfied clauses, ranging over all possible assignments. The problem gap-3SAT$_{c,s}$ is the problem of deciding for a given instance φ, whether

– $\mathrm{val}(\varphi) \geqslant c$, or

– $\mathrm{val}(\varphi) \leqslant s$.

The basic PCP theorem [2, 1, 4] is equivalent to

Theorem 1.1 (Basic inapproximability)**.** *There exists some $\varepsilon_0 > 0$ such that gap-3SAT$_{1,1-\varepsilon_0}$ is NP-hard.*

In fact, a much stronger statement is true:

Theorem 1.2 (Tight inapproximability, [5])**.** *For every $\varepsilon > 0$, gap-3SAT$_{1,\frac{7}{8}+\varepsilon}$ is NP-hard.*

This is optimal since one can always satisfy a 7/8 fraction of clauses for 3SAT. Similar theorems are known for many CSPs, and for many more this would follow from the unique games conjecture.

2 Label Cover

We define the label cover problem. An instance is given by a bipartite graph $G = (L, R, E)$. Each vertex v is associated with a finite label set A_v. Each edge $uv \in E$ is associated with a relation $\pi_{uv} \subset A_u \times A_v$. We will consider only relations that are functional, and sometimes called projection constraints, in which for every $a_u \in A_u$ there is exactly one $a_v \in A_v$ such that $(a_u, a_v) \in \pi_{uv}$. We write $\pi_{uv}(a_u) = a_v$.

A labeling $f = \{f_v \in A_v \ : \ v \in L \cup R\}$ is an assignment of one label per each vertex. The value of the assignment is

$$val(f) = \mathop{\mathbb{P}}_{uv\sim E}[(f_u, f_v) \in \pi_{uv}]$$

If edges in E have weights then the probability of choosing an edge is proportional to the edge weights. In this lecture we will sketch a proof for the following theorem

Theorem 2.1 (Label Cover theorem). *For every $\varepsilon > 0$, gap-Label-Cover$_{1,\varepsilon}$ is NP-hard, with projection constraints and labels of size at most $\varepsilon^{O(1)}$.*

The proof proceeds by reduction from the basic PCP theorem. As a first step, the basic PCP theorem gives a weaker result:

Lemma 2.2. *gap-Label-Cover$_{1,1-\frac{\varepsilon_0}{3}}$ is NP-hard.*

The proof is by reduction from gap-3SAT$_{1,1-\varepsilon_0}$.

Two Player Games and Parallel Repetition. A label cover instance $\mathcal{G}$ can be viewed as describing a two player game between L and R. The referee selects a question pair uv and sends u to L and v to R. The players (do not speak to each other) reply with a_u, a_v and the referee accepts if $\pi_{uv}(a_u) = a_v$. The success probability is the fraction of Accepts.

What if the players have success probability $1-\varepsilon$ and the referee wants to catch them in an unsuccessful edge? He can repeat the game k times. If he does this in parallel, we call this parallel repetition:

The referee chooses k question pairs, and sends $u_1, \ldots, u_k$ to L and $v_1, \ldots, v_k$ to R.

What is their maximal success probability now?

This simply takes a direct product of the label cover instance.

Construction of Label Cover. Given a label cover instance $\mathcal{G}$ as above, construct a new instance:

- The vertices are L^k and R^k
- An edge connects $(u_1, \ldots, u_k)$ to $(v_1, \ldots, v_k)$ if every $\{u_i, v_i\} \in E$.
- The label sets are A^k and B^k.
- For an edge $(\bar{u}, \bar{v})$ we let $\pi_{\bar{u},\bar{v}}(a_1, \ldots, a_k) = (\pi_{u_i,v_i}(a_i))_i$.

The parallel repetition theorem [6, **?**, 3] shows that the value of this label cover instance goes down exponentially.

Theorem 2.3 (Parallel Repetition). *If $val(\mathcal{G}) < 1 - \varepsilon$ then $val(\mathcal{G}^k) \leqslant (1 - \varepsilon^2/16)^k$.*

We will look at a different variant which is slightly more specialized and easier to analyze. Start with a 3SAT formula φ that has n variables V and m clauses C.

- The vertices are $L' = C^k$ and $R' = V^{0.9k}$
- An edge connects $s \in L'$ to $t \in R'$ if there is an ordered choice of $0.9k$ clauses from s, and a variable from each clause, that gives t.
- The label set for s is A^k ($A = [7]$), and for t it is $\mathbb{F}_2^{0.9k}$.
- For an edge (s, t) we let $\pi_{s,t}(f)$ be the restriction of f to t.

Fom now on let us denote by $\mathcal{G}_{\varphi,k}$ the (later) label cover instance.

Simplicial Complex Perspective. Let X be a simplicial complex constructed over k copies of the ground set $X(0) = V_1 \sqcup ...V_k$, $V_i = V$. For every choice of $s' = (c_1, \ldots, c_k)$ we will have a $3k$-dimensional face. Downwards close.

Claim 2.4 (Completeness)**.** If $val(\varphi) = 1$ then for all k $val(\mathcal{G}_{\varphi,k}) = 1$.

Proof. Let $g : V \to \{0,1\}$ be an assignment that satisfies all of the clauses of φ. Label each $s \in S$ by $g|_s \in A_s$ and each $t \in T$ by $g|_t$. Clearly this satisfies all of the edge constraints and has value 1. □

Lemma 2.5 (Soundness)**.** *If $val(\varphi) < 1 - \varepsilon_0$ then $val(\mathcal{G}^k) \leqslant \exp(-k)$.*

Together Lemma 2.5 and Claim 2.4 prove Theorem 2.1.

Is the Soundness Lemma (or the parallel repetition theorem) obvious? it would be if we consider only labelings that themselves are direct product, namely consistent with some underlying assignment to the variables. Indeed, fix $h : V \to \mathbb{F}_2$, and suppose that we label every $t \in T$ with $h|_t \in \mathbb{F}_2^t$. No matter what the labeling for L' is, we can prove

Lemma 2.6. *Let (b_t) be direct product labeling. For every labeling $a = (a_s)$, $\text{val}(a, b) \leqslant \exp(-k)$.*

Proof. Assuming $val(\varphi) < 1 - \varepsilon_0$, h must falsify an ε_0 fraction of clauses. A Chernoff bound shows that nearly all $s \in S$ have at least $\frac{\varepsilon_0}{2}$ of variables that are assigned a value different from $h|_s$ (indeed $h|_s \notin A_s$ is not even a valid label). A random edge (s, t) will be inconsistent except with probability exponential in k. □

We call labelings that are consistent with a global h *direct product labelings.*

The main question is whether labelings can benefit significantly if they deviate from a direct product labeling. Naively, we would like to show that whenever a labeling has value above ε, it must be close to a direct product on some ε' fraction of the label cover nodes.

Unfortunately, this is false. Instead, we will be able to prove a weaker statement that will suffice. We will show that if a labeling has large value, it must look like a global labelling on "link".

A *restriction* is a set r of vertices, each from a distinct color class. We let S_r be the top faces containing r,

$$S_r = \{s \in S \mid \ s \supset r\}.$$

Lemma 2.7 (Global structure on restrictions)**.** *For all α there is large enough k and some $\gamma > 0$ such that if $\{f_s\}, \{g_t\}$ is a labeling for $\mathcal{G}_{\varphi,k}$ with value above $\varepsilon > (1-\gamma)^k$ then there is a restriction $r \subset V$, $|r| = 0.1k$, and a global assignment $h : V \to \{0,1\}$ such that*

$$Prob_{s \sim S_r}[h|_s \overset{\alpha k}{=} f_s] > poly(\varepsilon)$$

where the notation $x \overset{\alpha k}{=} x'$ means that x, x' differ on at most αk points, and we set $\alpha = 10^{-5}$.

We will discuss this lemma and its proof further below, and in the next lecture. First, let us see that it is useful. We show how to derive the soundness lemma from it.

Proof of Lemma 2.5. Let $\alpha = \varepsilon_0/2$. Suppose we are given a labelling $\{f_s, g_t\}$ with value above ε. By the structure lemma there is some restriction $r \subset V$, $|r| = 0.1k$, and a global assignment $h : V \to \{0,1\}$ such that

$$Prob_{s \sim S_r}[h|_s \overset{\alpha k}{=} f_s] > poly(\varepsilon).$$

h must violate at least ε_0 fraction of the clauses of φ. Except for an exponentially small fraction of $s \in S_r$, s must contain at least $\varepsilon_0/2$ violated clauses. This means that the label f_s must differ from $h|_s$ on at least that many elements, which is more than αk, a contradiction. □

3 Agreement tests

The structure lemma is really a kind of so-called agreement test. The setup for an agreement test is a simplicial complex X, and a distribution over pairs of sets in X. We also have, for every $s \in S$ a space of available local functions $L_s \subset \{0,1\}^s$

A collection of local functions $\{f_s \in L_s\}$ is a perfect collection if there is some $h : V \to \{0,1\}$ such that $f_s = h|_s$ for every $s \in S$. An agreement test tests, in the property testing sense, if a given collection is perfect. Here is the test relevant to us

Agreement test with parameter $\rho > 0$:.

- Choose $t \in X(\rho k)$.
- Choose independently $s, s' \in X(k)$ such that $s, s' \supset t$.
- Accept iff $f_s(v) = f_{s'}(v)$ for all $v \in t \subset s \cap s'$.

Typically, one wants to prove that if the test succeeds with significant probability, then $\{f_s\}$ is close to perfect. We will prove

Theorem 3.1. *Let $\rho = 0.5$ there exists $0 < \gamma < 1$ such that if $\{f_S\}$ passes the agreement test with probability at least $\varepsilon := (1-\gamma)^k$, then there exists some restriction r of size $0.1k$ and a function $h_r : V \to \mathbb{F}_2$ such that*

$$Pr_{S \supseteq r}[f_S \overset{\geqslant (1-\alpha)k}{=} h|_S] \geqslant \varepsilon^{O(1)},$$

where the notation $\overset{\geqslant (1-\alpha)k}{=}$ means that the two strings disagree on at most αk locations and one can take $\alpha = 10^{-5}$.

Note that $\{S \supset r\}$ is a very small fraction of $\{S\}$, roughly $n^{-k/10}$.

A few remarks about the theorem.

- We cannot expect a global function agreeing on many f_S when the acceptance probability is $exp(-k)$. One can easily find an example where this is not achievable. However, if the acceptance probability is $k^{-O(1)}$ then one can in fact conclude that there is a global function agreeing on $k^{-O(1)}$ fraction of the f_S [**?**].
- As we saw in the previous lecture, even this conclusion is enough to prove the parallel repetition theorem with exponential decay.

How can one find the global function h_r? Clearly, the natural strategy of defining $h_r(x)$ by the plurality value at all $\{f_S(x) \mid x \in S\}$ is not going to work. A simple counter example to this strategy is to assign a random string from $\{0^k, 1^k\}$ to f_S. Clearly, the agreement test passes with probability at least 1/2, but the plurality strategy gives a random function.

To counter such examples, the overall idea is to *zoom-in* to a small subset of $\{S\}$ such that we enjoy much stronger agreement among the sets form the subset.

3.1 Expansion of small sets in DP_ρ

It is very instructive to think of the cases when the direct product test passes with non-negligible probability but there is no global function agreeing with $\{f_S\}$. To this end, consider the folded graph DP_ρ. In this graph, there are many small sets which *do not* expand. For eg. for a fixed subset r of size $\ll k$ and consider the family of sets $\{S \mid S \supset r\}$. Since roughly with probability ρ, we are keeping an element while moving to a neighbor in the graph, with probability roughly $\rho^{|r|}$ we stay in the same set $\{S \mid S \supset r\}$. This gives a way to create a collection $\{f_S\}$ which will pass the agreement test with non-negligible probability. For every r of size ρk, take a random function g_r and set $\{f_S \mid S \supset r\}$ with respect to g_r (if not assigned previously). In this case, there is no global function correlated with $\{f_S\}$ but the agreement test passes with probability at least $\rho^{|r|}$. This is precisely because in the agreement test we end up selecting (S, S') from $\{S \mid S \supset r\}$ for some r with probability $\rho^{|r|}$. Thus, in this respect Theorem **??** is tight!

Thus, study small set expansion property in these graphs is instrumental in analyzing such tests. This is different from the global expansion property of the graph. In this case, we can argue about the global expansion by studying the eigenvalues of the associated adjacency matrix.

We have the following lemma which says something about the small set expansion in the graph DP_ρ.

Lemma 3.2. *Suppose $A, B \subseteq [n]^k$ of size at least ε then*

$$Pr_{(x,y)\in DP_\rho}[x \in A \ \&\ y \in B] \geqslant \varepsilon^{\frac{2-\sqrt{\rho}}{1-\sqrt{\rho}}}$$

Here are a few simple observations regarding the above lemma.

- if $\rho = 0$ then x and y are totally uncorrelated and hence we get that the probability of the event $x \in A$ and $y \in B$ is ε^2, as expected.
- If $\rho = 1$ then x and y are perfectly correlated and if A and B are disjoint then we do get the probability to be 0.
- when ρ is somewhere in between, say 1/2, then the lemma non-trivially says that no matter which sets A and B we take, we have a considerable chance that $x \in A$ and $y \in$B.

3.2 Proof of Theorem 3.1

Definition 3.3. A restriction r is "good" if there exists $g : r \to \{0, 1\}$ such that the set

$$Z_r^g = \{S \supset r \mid f_S|_r = g\}$$

is of size at least $\Omega(\varepsilon)$.

Claim 3.4. There are at least $\Omega(\varepsilon)$ fraction of good r, where r is distributed according to the test distribution.

Proof. This follows from a simple averaging argument. □

Consider the following distribution D_1: (r, t, S, S') - Select $r \sim B(k, 1/10)$, $|t - r| \sim B(k, 4/10)$, $v_1, \ldots, v_t$. Then choose $S \setminus t$ and $S' \setminus t$.

Definition 3.5. r_0 is β-excellent if

$$Pr_{(r,t,S,S')\sim D_1}[f_S|_r = f_{S'}|_r \text{ but } f_S|_t \overset{\geqslant\beta k}{\neq} f_{S'}|_t \mid r = r_0] < exp(-k).$$

The notation $\overset{\geqslant\beta k}{\neq}$ means that the two string disagree on at least βk locations.

In other words, r_0 is excellent if for a typical pair of sets $S, S' \in Z^g_{r_0}$ agreeing on r_0 and whose intersection is more than r_0 also agree on (most of) the remaining intersection. This property is crucial in arguing that the plurality vote from the set $\{f_S \mid S \in Z^g_{r_0}\}$ is going to be consistent with many $\{f_S \mid S \in Z^g_{r_0}\}$.

Claim 3.6. There are at least $(1 - exp(-k))$ fraction of r which are β-excellent.

Proof. Consider r_0 and consider the family of sets $\{S \supset r_0\}$. Now, based on $f_S|_{r_0}$, we can partition the sets $\{S \supset r_0\}$ into at most $exp(k)$ parts. Consider a subgraph of the graph $DP_{\rho'}$ on $\{S \supset r_0\}$ where we only consider edges whose both the end points are inside one of the partitions. For an edge (S, S') let $t = (S \cap S') \setminus r_0$. We will call an edge (S, S') good if f_S and f'_S agree t on at least $(1 - \beta)$ fraction of points. Otherwise we call the edge bad. In this picture, the excellence property precisely means that the fraction of bad edges is $exp(-k)$.

Alternate way of choosing D_1 is first select t form the appropriate binomial distribution $B(k, 5/10)$ and then select r as a subset of t. According to this distribution given that the event $f_S|_t \overset{\geqslant\beta k}{\neq} f_{S'}|_t$ occurs, the probability that $f_S|_r = f_{S'}|_r$ is $2^{\Omega(-\beta k)}$. This is because while choosing r we will have to miss every βk elements from t where $f_S, f_{S'}$ disagree. Thus,

$$Pr_{(r,t,S,S')\sim D_1}\left[f_S|_r = f_{S'}|_r \text{ but } f_S|_t \overset{\geqslant\beta k}{\neq} f_{S'}|_t\right] \leqslant 2^{\Omega(-\beta k)}.$$

Therefore, there are at most η fraction for r_0 such that

$$Pr_{(r,t,S,S')\sim D_1}\left[f_S|_r = f_{S'}|_r \text{ but } f_S|_t \overset{\geqslant\beta k}{\neq} f_{S'}|_t \mid r = r_0\right] \geqslant \frac{2^{\Omega(-\beta k))}}{\eta}.$$

Rest of the r_0 are excellent, setting $\eta = exp(-k)$ proves the claim. □

We have a following simple corollary.

Corollary 3.7. *There are at least* $poly(\varepsilon)$ *fraction of* r *which are good and excellent.*

The following claim finishes the proof of the direct product theorem.

Claim 3.8. If r is good (large Z^g_r) and excellent then there exists $h_r : V \to \{0, 1\}$ such that

$$Pr_{S\sim Z^g_r}[f_S \overset{\geqslant\alpha k}{\neq} h_r|_S] \leqslant \varepsilon^{O(1)}$$

We define the function h_r on $x \in [n] \setminus r$ by taking plurality of $\{f_S(x) \mid S \in Z^g_r, x \in S \setminus r\}$.

We will give a proof sketch here. For more rigorous proof see [**?**]. Before proceeding, let us see why it should work. The reason why plurality works is because we are in the *high acceptance regime* inside Z^g_r. In other words, inside the set Z^g_r, if we look at a pair of sets whose intersection is more then r then with high probability (w.p close to 1) they agree on most of the intersection. This is precisely the excellence property! Thus, once we zoom-in to Z_r^g, we have a direct product test (a slight variation as we are only considering whether they *mostly* agree or not inside the intersection instead of a *total* agreement) which accepts with probability close to 1.

In order to use the excellence property, it is desirable to consider the graph $DP_{\rho'}$ where $\rho' = 5 \cdot \rho$. In this graph, we can label edges (S, S') as 'good' if $f_S|_{S \cap S'} \overset{\geqslant(1-\beta)\rho' k}{=} f_{S'}|_{S \cap S'}$, and 'bad' otherwise. Since r is excellent, there are many 'good' edges. These good edges will contribute towards showing $f_S \overset{\geqslant(1-\alpha)k}{=} h_r|_S$, provided there are many 'good' edges inside Z_r^g. This is because f's opinion on only Z_r^g is considered while defining h_r. A priori, it is not clear why it should be the case that many 'good' edges are inside Z_r^g. This is where we use Lemma 3.2. Thus, using this lemma, there are many good edges inside Z_r^g and the *plurality decoding* works.

Proof. (Sketch) Suppose the claim in not true, this means for a random $S \in Z_r^g$, f_S and $h_r(S)$ disagree on at least αk locations with probability $\varepsilon^{O(1)}$. Select a random set $e \subseteq S \setminus r$ of size $0.4k$. Then, by simple application of Chernoff bound, we get that $h_r(e)$ and $f_S(e)$ disagree on at least $\alpha/2$ fraction of the locations with high probability. However, since we define h_r by taking the plurality vote, for a random e, $h_r(e)$ should agree with at least $\Omega(\varepsilon)$ fraction of $f_S|_e$ on at least $\Omega(1)$ fraction of locations. These two contradict the excellence property. The starting assumption claims that for a random e and S containing e, $h_r(e)$ and $f_S|_e$ disagree on many locations, whereas the plurality condition would imply that $h_r(e)$ and $f_S|_e$ should agree on $\Omega(1)$ fraction of points. Both these properties imply that for a random e and S, S' containing e in Z_r^g, $f_S|_e$ and $f_{S'}|_e$ disagree on many locations, contradicting the excellence property of r. □

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and intractability of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. 1

[2] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. 1

[3] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 624–633, 2014. 2

[4] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. *Journal of the ACM*, 43(2):268–292, 1996. 1

[5] Johan Hastad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001. 1

[6] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998. 2