

Lecture 11

Lecturer: Irit Dinur

Scribe: Na'ama Sery

1 Overview

In the last several weeks we have been working on the proof to the PCP Theorem. We defined the following definitions:

Definition 1 (Constraint Graphs) :

A constraint graph is $G = \langle (V, E), \Sigma, C \rangle$ such that

- (V, E) is a graph
- Σ is some finite set (the alphabet)
- C gives a binary constraint on the graph's edges $C : E \rightarrow \{\phi : \Sigma \times \Sigma \rightarrow \{0, 1\}\}$

Definition 2 (unsat(G)) :

For any assignment a to the vertices of G $a : V \rightarrow \Sigma$,
the unsatisfaction value of G from a is

$$\text{unsat}_a(G) = \Pr_{e \in E} [C(e)(a(u), a(v)) = 0].$$

The unsatisfaction value of G is

$$\text{unsat}(G) = \min_a \text{unsat}_a(G).$$

We saw that the PCP Theorem follows from the following theorem:

Theorem 3 There exists Σ_0 s.t. for all Σ there is some $c > 0$ and $\gamma > 0$ s.t. given a constraint graph $G = \langle (V, E), \Sigma, C \rangle$ we can generate $G' = \langle (V', E'), \Sigma_0, C' \rangle$ that maintains:

- $\text{size}(G') \leq c \cdot \text{size}(G)$
- $\text{unsat}(G) = 0 \Rightarrow \text{unsat}(G') = 0$
- $\text{unsat}(G) > 0 \Rightarrow \text{unsat}(G') = \min\{\gamma, 2 \cdot \text{unsat}(G)\}$

Proof sketch: $G \rightarrow H_1 \rightarrow H_1^t \rightarrow H_1^t \circ \phi$

1. $G \xrightarrow{\text{preprocessing}} H_1$: In this phase we create d -regular expander graph from G at the cost of decreasing the unsat level by a constant level.
2. $H_1 \xrightarrow{\text{powering}} H_1^t$: In this phase we will increase the unsat level by a constant as large as we wish (compensating also for the decrease in unsat at the two other phases). The cost of this phase is the increase of Σ .
3. $H_1^t \rightarrow H_1^t \circ \phi$: In this phase we correct the alphabet while decreasing the unsat only by a constant factor.

Last week we finished with the first operation. This week we began working on the second phase.

2 The Second Operation: Powering - Objective and Intuition

Input:

a constraint graph $G = \langle (V, E), \Sigma, C \rangle$ that is a d -regular expander graph.

Objective:

return a constraint graph $G^t = \langle (V', E'), \Sigma', C' \rangle$ that increases the unsat in case $unsat(G) > 0$ until $unsat(G') > t \cdot unsat(G)$ for a constant t as large as we wish.

Formally:

Given a constraint graph $G = \langle (V, E), \Sigma, C \rangle$ such that G is d -regular and $\lambda(G) \leq \lambda$ compute a constraint graph $G^t = \langle (V', E'), \Sigma', C' \rangle$ in polynomial time such that:

1. The size of the graph and the alphabet will grow only by a constant: $size(G^t) \leq c_2 \cdot size(G)$
2. $unsat(G) = 0 \Rightarrow unsat(G^t) = 0$
3. $unsat(G) > 0 \Rightarrow unsat(G^t) = \min\{\gamma, 2 \cdot unsat(G)\}$

Intuition:

Let $h : V \rightarrow \Sigma$ be an assignment for G . Denote $\alpha = unsat_h(G)$.

Let $F_h \subseteq E$ be the edges that h doesn't satisfy.

Therefore, the probability that a random edge in G will not be satisfied by h is $\frac{|F_h|}{|E|}$. We would like to enhance this probability. One way, could be to choose t random edges instead of one. This way, we will calculate the probability that any one of the t edges is not satisfied by h . This probability equals: $1 - (1 - \frac{|F_h|}{|E|})^t \approx t \cdot \frac{|F_h|}{|E|}$ (which is an amplification by t of the former probability).

But, how would we formalize this intuition into constraint graphs? We could think of a constraint representing every t edges which will be satisfied if and only if all t edges are satisfied. But it will not be a binary constraint and so could not be represented by a constraint graph.

Instead, we will look at a random walks of length t in the graph. The probability that at least one of the edges in the path is in F_h is at least $constant \cdot t \cdot \frac{|F_h|}{|E|}$ where the constant depends on G 's degree and expansion.

We will start by describing a first construction and then will prove the theorem on a better, slightly different, construction.

3 Definition of First Construction

Following the above ideas, our first construction of G^t will involve constraints that instead of relating to two vertices, will relate to two balls with radius t .

- Because we want that: $unsat(G) = 0 \Rightarrow unsat(G^t) = 0$, we will not invent new constraints. We will only have compositions of the former constraints.
- Still, we also want that: $unsat(G) > 0 \Rightarrow unsat(G^t) = \min\{\gamma, 2 \cdot unsat(G)\}$. So, we will try to increase the percentage of the constraints that are violated.

Formal definition of the construction of the constraint graph G^t :

- G' Vertices: $V' = V$
- G' Edges E' are defined by the adjacency matrix A^t when A is the G adjacency matrix and the power function is the regular power function over matrices. This definition means that in E' there are exactly k parallel edges between u and v if the number of walks between u and v in G of

length t is k . This is the part that will be the difference between this construction and the final construction.

- G' Alphabet $\Sigma' = \Sigma^S$. S is an upper bound on the number of vertices in a ball of radius t . S equals $d^0 + d^1 + \dots + d^t$ because in a d -regular graph, the number of vertices with distance i from a certain vertex is at most d^i . Denote: $B(v, t) = \{u \mid \text{dist}_G(v, u) \leq t\}$. The reason we define $\Sigma' = \Sigma^S$ is because we want each $\sigma \in \Sigma'$ to be an assignment $\sigma : B(v, t) \rightarrow \Sigma$. For each $v \in V'$ the $\sigma \in \Sigma'$ will be a vector of length $|B(v, t)|$ where each entry $\sigma_i \in \Sigma$ means the "opinion" v has on its i 'th neighbor.
- G' Constraints C' are defined for each $(u, v) \in E'$ to "check as much as you can". $C'(u, v)$ will satisfy on $a, b \in \Sigma'$ if and only if these two conditions are true:
 1. u and v agree in opinions on their mutual relatives. Let us denote the union of v 's and u 's relatives by: $B = B(u, t) \cup B(v, t)$. Then, formally, the condition means that there exists a function $f : B \rightarrow \Sigma$ so that:
 - f agrees with a on $B(u, t)$: $\forall w \in B(u, t) f(w) = a(w)$
 - f agrees with b on $B(v, t)$: $\forall w \in B(v, t) f(w) = b(w)$
 2. all the original constraints in the area of the two vertices satisfy. Formally: for the same function f , $\forall e = (w_1, w_2) \in E \cap (B \times B) : C(w_1, w_2)$ is satisfied by $f(w_1), f(w_2)$

Notice that from this definition of C' it follows that every two parallel edges have the same constraint. That is because, when defining the constraint on edge (u, v) we depended only on u and v and not on the path between them that the edge represents. It is still important for us to have these parallel edges because they enhance the percentage of the unsatisfied constraints when they are violated.

4 Correctness of First Construction (Intuition)

We required that given a d -regular expander constraint graph $G = \langle (V, E), \Sigma, C \rangle$, our construction return a constraint graph $G^t = \langle (V', E'), \Sigma', C' \rangle$ such that:

1. The size of the graph and the alphabet will grow only by a constant: $\text{size}(G^t) \leq c_2 \cdot \text{size}(G)$
2. $\text{unsat}(G) = 0 \Rightarrow \text{unsat}(G^t) = 0$
3. $\text{unsat}(G) > 0 \Rightarrow \text{unsat}(G^t) = \min\{\gamma, 2 \cdot \text{unsat}(G)\}$

and that our construction will take polynomial time.

Let us analyze the attributes of G^t in our construction:

- size and construction time:
 - $|V'| = |V|$ because $V' = V$
 - G^t is a d^t regular graph - because in a d -regular graph G , for every $v \in V$ the number of vertices with distance t from v are d^t .
 - $|E'| = |V| \cdot \frac{d^t}{2}$ because G^t is a d^t regular graph (and since d is a constant that is ok)
 - $|\Sigma'| = |\Sigma|^S$ but since we assume Σ is of constant size and S is a function of d which is constant, that is ok.
 - C' has constraints much more complex than C but, again, it grows by factors dependent only on constants (so its ok).

- The time that it takes to compute E' and C' is polynomial.
- $unsat(G) = 0 \Rightarrow unsat(G^t) = 0$: This is true because given an assignment for G $h : V \rightarrow \Sigma$ that satisfies every edge in G , we can construct an assignment for G^t $H : V \rightarrow \Sigma'$ that satisfies every edge in G^t . In this assignment vertex v 's "opinion" on a vertex $u \in B(v, t)$ will be exactly $h(u)$. So, obviously, every constraint is satisfied by H .
- What is left to show is that $unsat(G) > 0 \Rightarrow unsat(G^t) = \min\{\gamma, 2 \cdot unsat(G)\}$. We will only give an intuition of why this is true here and give a full proof on the next construction. For simplicity, we will describe how $unsat(G) > 0 \Rightarrow unsat(G^t) = \alpha \cdot unsat(G)$ when α is a function of t .
 - For every assignment for G $h : V \rightarrow \Sigma$ we can construct an assignment for G^t $H : V \rightarrow \Sigma'$ as we described in the previous point. Then it holds that: $unsat_H(G^t) \geq const \cdot t \cdot unsat_h(G)$. This is because $unsat_H(G^t)$ is the probability that a random edge in G^t is not satisfied. A random edge in G^t is a random walk of length t in G . To decide whether the constraint on edge (u, v) in G^t should be satisfied, we observe the two balls with radius t around the vertices at the ends of the random walk in G : u, v . Because of the way H was defined, u and v will agree on their "opinions" on their mutual relatives ($H(u)_{(w)} = h(w) = H(v)_{(w)}$). Therefore, the constraint on the edge will be violated if and only if there is an edge in $E \cap (B \times B)$ that is violated by h . Within the union of the two balls there are the edges of the random walk from u to v . If one of these edges are violated then surely the constraint on edge (u, v) in G^t is violated. Therefore, the probability of a random edge being violated in G^t is at least the probability of an edge being violated in a random walk of length t in G . The last probability is equivalent also to the probability that a random walk of length t will "meet" the group of edges F_h . It equals $const \cdot t \cdot unsat_h(G)$ when $const$ depends on d and G 's expansion.
 - Of course, this argument is insufficient since, perhaps, the optimal assignment H for G^t doesn't agree with any assignment h for G completely. We will show, that for the optimal assignment H for G^t , i.e. $unsat_H(G^t) = unsat(G^t)$, there is an assignment h for G , such that $unsat_H(G^t) \geq c \cdot t \cdot unsat_h(G)$ where c is a constant independent of t . From there will follow that $unsat(G^t) = unsat_H(G^t) \geq c \cdot t \cdot unsat_h(G) \geq c \cdot t \cdot \min_a unsat_a(G) = c \cdot t \cdot unsat(G)$. As mentioned before, we will not give a full proof in this section, because this construction achieves an amplification in the scale of \sqrt{t} . A full proof will be provided for the more complex final construction which achieves an amplification in the scale of t . We will only give some more intuition on how the proof will work:
 - The assignment H might not have an assignment h for G that it totally agrees with because it is possible for vertices not to agree on their "opinions" on their mutual relatives. So, we will define h such that for every $w \in V$ $h(w)$ is the "common" opinion of w 's relatives on him. The "common" opinion could be the majority opinion or something more complex that takes into account the closeness of that relative. For now, we will consider the most common opinion amongst my relatives opinions while I consider all the opinions uniformly. Later, we will define a different distribution.
 - We want to show that an edge in G^t , which is a path of length t in G , has a good chance of being violated by H if there is an edge in the path that is violated by h .
 - Consider an edge $(u, v) \in G$ that is violated by h . Then what is the probability of a path $v_0 \rightsquigarrow v_t$ that includes (u, v) to be violated by H ? If, in H , v_t agrees with the majority's opinion on v and v_0 agrees with the majority's opinion on u , then the constraint on (v_0, v_t) in G^t is violated by H . This is because $H(v_0)_u = h(u)$ and $H(v_t)_v = h(v)$ and $(u, v) \in E \cap (B \times B)$ and so the consistency of (u, v) will be checked by (v_0, v_t) 's constraint.
 - We want to calculate the percentage of unsatisfied edges in G^t . From the previous point, that percentage is greater than the probability that an edge $(v_0, v_t) \in E'$ will have a violated edge $(u, v) \in E$ in the path from v_0 to v_t and that $H(v_0)_u = h(u)$ and $H(v_t)_v = h(v)$. So we ask:

given a random edge in G and a random path of length t that includes this edge, what is the probability that this will occur?

- Given a random edge $(u, v) \in G$, we will pick a random path by picking a random path of length i from u and a random path of $t-1-i$ from v . Together and with the (u, v) edge this consists of a path of length t . Now, we would like to calculate the probability that v_0 has the common opinion on u and v_t has the common opinion on v . This depends on the way we define the common opinion. We should define the common opinion in a way that would be most beneficial for this proof. We will define a distribution on the opinions so that this probability will be constant. That way, if (u, v) is violated by h , then with a constant probability the random path will be violated by H . Then, we will use the fact that the path holds t edges. So, the probability that any edge in the path is violated is approximately t times the probability that one specific edge is violated (derives from the expansion of G). This is the main idea of the proof, and will be expanded in the next construction.

5 Definition of Final Construction

This construction for G^t is similar to the previous construction, only we are going to change the definition of E' a bit. We are following a paper of Charanjit S. Jutla, that can be found in

<http://eccc.hpi-web.de/eccc-reports/2006/TR06-121/index.html>.

Let us assume that we use only $t \in \mathbb{N}$ divisible by 8.

Let us define: $T_2 = [-\frac{t}{2}, \frac{t}{2}]$, $T_4 = [-\frac{t}{4}, \frac{t}{4}]$ and $T_8 = [-\frac{t}{8}, \frac{t}{8}]$ groups $\subseteq \mathbb{Z}$.

We will define the distribution D on T_2 by the random process:

1. choose a random number j_1 from T_4 by the uniform distribution
2. choose another random number j_2 from T_4 by the uniform distribution
3. return $j_1 + j_2 \in T_2$

Claim 4 $\forall \ell \in T_2 \Pr_D(\ell) = \frac{\frac{t}{2} + 1 - |\ell|}{(\frac{t}{2} + 1)^2}$

Proof

$$\begin{aligned}
\Pr_D(\ell) &= \Pr_{i, j \in T_4} [\ell = i + j] \\
&= \sum_{i \in T_4} \Pr(i) \cdot \Pr_{j \in T_4} [\ell = i + j \mid i] \\
&= \sum_{i \in T_4} \frac{1}{|T_4|} \cdot \Pr[|\ell - i| \leq \frac{t}{4} \wedge |j| \leq \frac{t}{4} \wedge j = \ell - i] \\
&= \sum_{i \in T_4} \frac{1}{|T_4|} \cdot \Pr_{i \in T_4} [|\ell - i| \leq \frac{t}{4}] \cdot \Pr_{j \in T_4} [j = \ell - i] \\
&= \sum_{i \in T_4} \frac{1}{|T_4|} \cdot \frac{|[\ell - \frac{t}{4}, \ell + \frac{t}{4}] \cap T_4|}{|T_4|} \cdot \frac{1}{|T_4|} \\
&= \frac{1}{|T_4|^2} \sum_{i \in T_4} \frac{\frac{t}{2} + 1 - |\ell|}{|T_4|} \\
&= \frac{1}{(\frac{t}{2} + 1)^2} \cdot [\frac{t}{2} + 1 - |\ell|] \\
&= \frac{\frac{t}{2} + 1 - |\ell|}{(\frac{t}{2} + 1)^2}
\end{aligned}$$

■

Observe that, by distribution D , the more ℓ is closer to 0, the more ℓ is probable.

In the last construction G^t 's edges, E' , were all the paths in G of length t . This time, we will define a distribution on G 's paths according to D . Let P^* be the group of all paths in G (P^* is a infinite group because we didn't bound the paths' length). We will define the distribution A on $P^* \times T_4$ by the random process:

1. choose $\ell \in T_2$ by the D distribution
2. choose a random walk \vec{p} on G of length $t + \ell$
3. mark $(\ell + T_4) = [\ell - \frac{t}{4}, \ell + \frac{t}{4}]$ and choose a random $s \in (\ell + T_4) \cap T_4$ uniformly
4. return (\vec{p}, s)

That is, the most probable paths are of length t and we also get paths shorter or longer then t by at most $\frac{t}{4}$. So, we defined a distribution on G 's paths, but we need to define a group for the edges of G^t .

Claim 5 *There exists a multiset P in which*

1. $|P| = O(|E|)$ (we count the size of P with the repetitions)
2. *The next distribution is identical to A :*
 - (a) choose a random $\vec{p} \in P$ uniformly
 - (b) mark $\ell = |\vec{p}| - t$ and choose a random $s \in (\ell + T_4) \cap T_4$ uniformly
 - (c) return (\vec{p}, s)

Proof The distribution D assigns the multiplicity of the paths in P . We wish that P will be such that choosing a random path $\vec{p} \in P$:

$$\Pr(|\vec{p}| = k) = \Pr_D(\ell = k - t) = \frac{\frac{t}{2} + 1 - |k - t|}{(\frac{t}{2} + 1)^2}$$

. So we will assign the multiplicity of a path of length k accordingly.

$$\begin{aligned} \Pr(|\vec{p}| = k) &= \frac{\text{the number of paths of length } k \text{ in } P}{|P|} \\ &= \frac{\{\text{multiplicity of a path of length } k \text{ in } P\} \cdot \{\text{the number of paths of length } k \text{ in } G\}}{|P|} \\ &= \frac{\{\text{multiplicity of a path of length } k \text{ in } P\} \cdot \{|V| \cdot d^{k-1}\}}{|P|} \end{aligned}$$

\Downarrow

$$\text{multiplicity of a path of length } k \text{ in } P = \frac{\Pr(|\vec{p}|=k) \cdot |P|}{|V| \cdot d^{k-1}} = \frac{\Pr_D(\ell=k-t) \cdot |P|}{|V| \cdot d^{k-1}}$$

and

$$|P| = \sum_{k=\frac{t}{2}}^{\frac{3t}{2}} \{\text{multiplicity of a path of length } k \text{ in } P\} \cdot \{\text{the number of paths of length } k \text{ in } G\}$$

which is $|V|$ times some constant that depends on d and t only. ■

We will define G^t as before with the change that $E' = \{(u, v) | (u, \dots, v) \in P\}$. Each multiplicity of an edge in E' has the same constraint as defined in the previous section. Now, choosing a random edge in E' is equivalent to choosing a random path in P which is equivalent to choosing a random path by the A distribution.

6 Correctness of Final Construction

Reminder:

We required that given a d -regular expander constraint graph $G = \langle (V, E), \Sigma, C \rangle$, our construction return a constraint graph $G^t = \langle (V', E'), \Sigma', C' \rangle$ such that:

1. The size of the graph and the alphabet will grow only by a constant: $\text{size}(G^t) \leq c_2 \cdot \text{size}(G)$
2. $\text{unsat}(G) = 0 \Rightarrow \text{unsat}(G^t) = 0$
3. $\text{unsat}(G) > 0 \Rightarrow \text{unsat}(G^t) = \min\{\gamma, 2 \cdot \text{unsat}(G)\}$

and that our construction will take polynomial time.

Let us analyze the attributes of G^t in this construction:

- size and construction time:
 - $|V'|$, $|\Sigma'|$ and $|C'|$ as in last construction
 - $|E'| = O(|E|)$ because $|E'| = |P|$ and the last claim showed that $|P| = O(E)$
 - The time that it takes to compute E' and C' is polynomial.
- $\text{unsat}(G) = 0 \Rightarrow \text{unsat}(G^t) = 0$: the argument is the same as in the last construction
- what is left to show is that $\text{unsat}(G) > 0 \Rightarrow \text{unsat}(G^t) = \min\{\gamma, 2 \cdot \text{unsat}(G)\}$. For simplicity, we will describe how $\text{unsat}(G) > 0 \Rightarrow \text{unsat}(G^t) = \alpha \cdot \text{unsat}(G)$ when α is a function of t .

Theorem 6 *There exist constants c and γ dependent on d and Σ such that*

$$\text{unsat}(G^t) \geq \min(\gamma, \text{unsat}(G) \cdot t \cdot c)$$

Proof

Let $H : V \rightarrow \Sigma'$ be the best assignment to V . That is, $\text{unsat}_H(G^t) = \text{unsat}(G^t)$.

Let $F' \subseteq E'$ be the edges that H violates.

We will define $h : V \rightarrow \Sigma$ such that

$$\forall v \in V \quad h(v) = \arg \max_{a \in \Sigma} \Pr(H(w)_v = a)$$

when w is chosen by the distribution:

1. choose a random $\ell \in [\frac{t}{2} - \frac{t}{8}, \frac{t}{2} + \frac{t}{8}]$
2. choose a random walk of length ℓ in G
3. mark the last vertex in the walk as w

Notice: This is a deterministic definition of h .

Notice: In contrast to the sketch of the proof for the first construction where $h(v)$ was the majority of the opinions on v , here, closer relatives' opinions count more than distant ones because the chance of choosing them is greater.

Let $F \subseteq E$ be the edges that h violates. We will show that in high probability, a path that includes an edge in F is in F' .

For any $i \in T_8$ we define an event A_i on the probability space $P \times T_4$.

Let $\vec{p} \in P$ $\vec{p} = (v_1, \dots, v_{t+\ell})$ and $s \in T_4$.

Let us mark $u = v_{i+s+\frac{t}{2}}$ and $v = v_{i+s+\frac{t}{2}+1}$.

Then, $A_i(\vec{p}, s)$ occurs if and only if:

1. the edge (u, v) is in F
2. $H(v_1)_u = h(u)$
3. $H(v_{t+\ell})_v = h(v)$

We have already seen in the proof sketch from before that if $A_i(\vec{p}, s)$ is true then the edge $(v_1, v_{t+\ell}) \in E'$ is in F' . That is because:

- $(u, v) \in E \cap (B \times B)$ because u 's distance from v_1 and v 's distance from $v_{t+\ell}$ is less than t .
- therefore, the constraint on (u, v) will be checked by (v_0, v_t) 's constraint
- since $H(v_0)_u = h(u)$, $H(v_t)_v = h(v)$ and h violates (u, v) that constraint will also be violated.

Let $x_i(\vec{p}, s)$ be the indicator random variable of A_i .

Define a random variable $N(\vec{p}, s) = \sum_{i \in T_s} x_i$.

For a path $\vec{p} \in P$ $N(\vec{p}, s) > 0$ if any one of the edges around $(s + \frac{t}{2})$ in the path causes the edge in G^t representing \vec{p} to be in F' .

Claim 7 $unsat(G^t) = \Pr_{e \in E'}(e \in F') \geq \Pr_{(\vec{p}, s) \in A}(N(\vec{p}, s) > 0)$

Proof The probability that $N(\vec{p}, s) > 0$ is the probability that a random path $\vec{p} \in P$ will include an edge that makes the edge \vec{p} represents be violated by H . This probability is smaller than the probability that a random edge $e \in E'$ be violated by H . ■

Claim 8 $\Pr_{(\vec{p}, s) \sim A}(N(\vec{p}, s) > 0) \geq \min(\alpha_0, c \cdot t \cdot unsat(G))$

Proof This week we only presented the sketch of the proof that will be extended next week:

1. N is a nonnegative random variable since it is a sum of indicator random variables.
2. We will prove the second moment method: If N is a nonnegative random variable then

$$\Pr[N > 0] \geq \frac{(\mathbb{E}N)^2}{\mathbb{E}[N^2]}$$

3. We will calculate $\mathbb{E}N$ and show that it is approximately $c \cdot t \cdot unsat(G)$.
4. We will calculate $\mathbb{E}[N^2]$ and show that it is approximately $\mathbb{E}N$.
5. We will conclude that $\Pr[N > 0]$ is approximately $c \cdot t \cdot unsat(G)$.

■ ■