Approximation Algorithms and Hardness of Approximation May 7 and 10, 2013

Lecture 20 and 21

Lecturer: Ola Svensson *Scribes: Siddhartha Brahma*

1 PCP theorems using Parallel Repetition

We will concentrate on $2CSP_W$ instances. The main PCP theorem, of which we have seen a proof, shows that there exists a constant $\epsilon > 0$ such that it is NP-Hard to distinguish between instances of $2CSP_W$ for which all the constraints can be satisfied and those for which atmost ϵ fraction of constraints can be satisfied. This can be used to show the hardness of approximation of several NP-Hard problems. However, for other problems we need "stronger" PCP theorems which have much smaller ϵ's without making W too large. The key to get such a guarantee is via Raz's Parallel Repetition Theorem.

It will be useful to define the following properties.

1. Let ϕ be an instance of a $2CSP$ problem. It is said to have the *projection property* if for each constraint $\phi_r(y_1, y_2)$ and each value of y_1 there is an unique value of y_2 such that $\phi_r(y_1, y_2) = 1$. In other words, knowing the value of y_1 is enough to know the value of y_2 such that $\phi_r(y_1, y_2) = 1$. More formally, for each constraint ϕ_r, there is a function $h : [W] \to [W]$ such that (u, v) satisfies the constraint iff $h(u) = v$.

2. A $2CSP$ instance is said to be *regular* if each variable appears in the same number of constraints. Thus, if we construct a bipartite graph with vertices representing variables on the left (y_i) and vertices representing constraints on the right ($\phi_r(y_i, y_j)$) and there is an edge from y_i to every $\phi_r(y_i, \cdot)$ or $\phi_r(\cdot, y_i)$, then the left vertices have the same degree.

Raz's Parallel Repetition theorem is the following.

Theorem 1 *For all $\epsilon > 0$, there exists $W = W(\epsilon) = poly(\frac{1}{\epsilon})$ such that given a $2CSP_W$ instance that is regular and has the projection property, it is NP-Hard to distinguish instances which are satisfiable from those in which atmost a fraction ϵ of the constraints are satisfiable. Alternatively, there exists an absolute constant $c > 1$ such that for every $t \geq 1$ $GAP2CSP_W(\epsilon)$ (for regular $2CSP$ instances satisfying the projection property) is NP-hard for $\epsilon = 2^{-t}$ and $W = 2^{ct}$.*

We will not present its proof (which is quite involved!), but try to give some intuition. Suppose ϕ either has $val(\phi) = 1$ or $val(\phi) = \epsilon$, but deciding which case holds is NP-hard. We can easily create a *powered* version of ϕ (call it ϕ^t) by taking t-tuples of variables of ϕ and defining constraints on them. Thus two variables $(y_1, \ldots, y_t)$ and $(z_1, \ldots, z_t)$ of ϕ^t has a corresponding constraint $(\phi_1(y_1, z_1), \ldots, \phi_t(y_t, z_t))$. From the verification viewpoint, this corresponds to running t instances of the verifier of ϕ in a parallel fashion. Clearly, if an assignment of ϕ satisfies ϵ fraction of constraints, then we can get an assignment of ϕ^t that satisfies at least ϵ^t fraction of constraints. However, it turns out that this is not tight and the fraction of constraints that can be satisfied in ϕ^t can be much larger than ϵ^t. However, Raz's result shows that no assignment can satisfy more than ϵ^{ct} fraction of constraints of ϕ^t, where c depends on W.

2 Håstad's 3-bit PCP

Another way to obtain stronger PCPs is by keeping the soundness requirement for the verifier around 1/2 but by reducing the number of bits of the proof that need to be read by the verifier. In the main PCP theorem, we insisted on reading $O(1)$ bits. However, the following remarkable theorem due to Håstad shows that this number can be brought down to 3 without affecting soundness and completeness of the verifier (almost)!

Theorem 2 *For every $\delta > 0$ and every language $L \in NP$, there is a PCP verifier V_H that reads 3 binary queries having completeness $1-\delta$ and soundness $\frac{1}{2}+\delta$. Moreover, the tests used by V_H are linear. That is given a proof $\tilde{\pi} \in \{0,1\}^n$, V_H chooses (i_1, i_2, i_3) and $b \in \{0,1\}$ according to some distribution and accepts iff $\tilde{\pi}_1 + \tilde{\pi}_2 + \tilde{\pi}_3 = b \pmod 2$.*

In fact, this theorem can be used to easily prove that it is NP-Hard to approximate MAX-3SAT with an approximation factor of $\frac{7}{8}+\epsilon$ for any $\epsilon > 0$. The remainder of the lecture will be about proving this theorem. Towards this, we will introduce the powerful tool of Fourier Analysis (a discrete analog of the classical Fourier Analysis in the continuous domain). Rather than working with functions in $GF(2)$, it will be much more convenient for us to work with functions $f : \{-1,+1\}^n \to \mathbb{R}$. This can be done by replacing the role of 0 by 1 and that of 1 by -1 in functions over $GF(2)$. Since addition modulo 2 for the binary alphabet represents parity, the corresponding operation for $\{-1,+1\}$ becomes a multiplication. Thus for $f' : \{0,1\}^n \to \{0,1\}$ and $f : \{-1,+1\}^n \to \mathbb{R}$

$$f'(x_1,x_2,x_3) = x_1 + x_2 + x_3 \text{ corresponds to } f(x_1,x_2,x_3) = x_1 \cdot x_2 \cdot x_3$$

Thus, we will be interested in the family of functions $f : \{-1,+1\}^n \to \mathbb{R}$ which can also be thought of as vectors in $\mathbb{R}^{2^n}$ via the truth table. In what follows, unless otherwise stated, x will denote the vector $(x_1, \ldots, x_n) \in \{-1,+1\}^n$. We will define a vector space $\mathcal{S}_n$ over this family with the following inner product

$$< f, g >= \mathop{\mathbb{E}}_{x\in\{-1,+1\}^n}[f(x)g(x)]$$

Also, for $S \subseteq [n] = \{1, \ldots, n\}$, let $\chi_S(x) = \prod_{i\in S} x_i$. The following lemma shows that these form an orthonormal basis for $\mathcal{S}_n$.

Lemma 3 *The set of functions $\{\chi_S : S \subseteq [n]\}$ (called* characters*) form an orthonormal basis of $\mathcal{S}_n$.*

Proof Clearly the number of characters is 2^n. We have

$$< \chi_S, \chi_S > \;=\; \mathop{\mathbb{E}}_{x\in\{-1,+1\}^n}[\chi_S(x)\chi_S(x)] = \mathop{\mathbb{E}}_{x\in\{-1,+1\}^n}\left[\prod_{i\in S} x_i^2\right] = 1$$

This shows that the characters have unit norm. Further, if $S \neq T$, then

$$< \chi_S, \chi_T >= \mathop{\mathbb{E}}_{x}\left[\prod_{i\in S} x_i \prod_{i\in T} x_i\right] \overset{(a)}{=} \mathop{\mathbb{E}}_{x}\left[\prod_{i\in S\triangle T} x_i\right] \overset{(b)}{=} \prod_{i\in S\triangle T} \mathop{\mathbb{E}}_{x}[x_i] = 0$$

Here $S\triangle T$ denotes the symmetric difference between the sets S and T. (a) is true because if $i \in S\cap T$, then the term inside the product becomes x_i^2 which is 1. (b) is true because the x_i's are chosen independently with equal probability of being -1 or +1. ■ The above lemma shows that any function in

the space $\mathcal{S}_n$ can be expresses as a linear combination of the characters.

$$f(x) = \sum_{S\subseteq[n]} \hat{f}_S \chi_S(x) \text{ where } \hat{f}_S =< \chi_S, f >$$

The $\hat{f}_S$ are called the *Fourier Coefficients* of f. As an illustration, consider space of functions $\mathcal{S}_1$. The characters are (written in the vector form) $\chi_\emptyset = [1,1]$ and $\chi_{\{1\}} = [-1,1]$. Then for the function $f = [0,1]$ the fourier coefficients are

$$\hat{f}_\emptyset =< f, \chi_\emptyset >= \frac{1}{2}, \quad \hat{f}_{\{1\}} =< f, \chi_{\{1\}} >= \frac{1}{2}$$

The following facts about fourier coefficients of f can be easily proved. These are the analogues of similar results in the continuous domain.

Lemma 4 *For any $f, g \in \mathcal{S}_n$, we have*

1. $< f, g >= \sum_{S\subseteq[n]} \hat{f}_S \cdot \hat{g}_S$
2. $< f, f >= \sum_{S\subseteq[n]} \hat{f}_S^2$

Proof

$$\begin{aligned} < f, g > &= \mathop{\mathbb{E}}_x[(\sum_S \hat{f}_S \chi_S(x))(\sum_T \hat{g}_T \chi_T(x))] \\ &= \sum_{S,T} \hat{f}_S \hat{g}_T \mathop{\mathbb{E}}_x[\chi_S(x)\chi_T(x))] \\ &= \sum_S \hat{f}_S \hat{g}_S \end{aligned}$$

■ Using this machinery, we are actually now in a position to give a simple proof of the linearity test that was used in the proof of the PCP theorem before.

Theorem 5 *For each function $f : \{-1, +1\}^n \to \{-1, +1\}$ that satisfies*

$$\mathbb{P}_{x,y}[f(xy) = f(x)f(y)] \geq \frac{1}{2} + \delta$$

there exists $S \subseteq [n]$ such that $\hat{f}_S \geq 2\delta$.

Proof Since the domain of f is $\{-1, +1\}$, we have

$$\mathop{\mathbb{E}}_{x,y}[f(xy)f(x)f(y)] = \mathbb{P}_{x,y}[f(xy) = f(x)f(y)] - \mathbb{P}_{x,y}[f(xy) \neq f(x)f(y)] \geq (\frac{1}{2} + \delta) - (\frac{1}{2} - \delta) = 2\delta$$

On the other hand

$$\begin{aligned} \mathop{\mathbb{E}}_{x,y}[f(xy)f(x)f(y)] &= \mathop{\mathbb{E}}_{x,y}[(\sum_S \hat{f}_S \chi_S(xy))(\sum_T \hat{f}_T \chi_T(x))(\sum_U \hat{f}_U \chi_U(y))] \\ &= \sum_{S,T,U} \hat{f}_S \hat{f}_T \hat{f}_U \mathop{\mathbb{E}}_{x,y}[\chi_S(xy)\chi_T(x)\chi_U(y)] \\ &= \sum_{S,T,U} \hat{f}_S \hat{f}_T \hat{f}_U \mathop{\mathbb{E}}_{x,y}[\chi_S(x)\chi_S(y)\chi_T(x)\chi_U(y)] \\ &= \sum_{S,T,U} \hat{f}_S \hat{f}_T \hat{f}_U \mathop{\mathbb{E}}_x[\chi_S(x)\chi_T(x)] \mathop{\mathbb{E}}_y[\chi_S(y)\chi_U(y)] \\ &\overset{(a)}{=} \sum_S \hat{f}_S^3 \leq \hat{f}_{\max} \sum_S \hat{f}_S^2 = \hat{f}_{\max} \end{aligned}$$

Here $\hat{f}_{\max}$ is the highest fourier coefficient of f and (a) follows from the fact that $\mathop{\mathbb{E}}_x[\chi_S(x)\chi_T(x)]$ is non-zero only when $S = T$ and $\mathop{\mathbb{E}}_y[\chi_S(y)\chi_U(y)]$ is non-zero only when $S = U$. ■

3 Proof sketch of Håstad's PCP

We will look at $2CSP_W$ instances with m constraints and n variables, satisfying the projection property. A proof will be an assignment to the variables from the alphabet $[W]$. We need a verifier V_H that satisfies the requirements of the theorem by reading very few bits of the proof. As with the proof of the main PCP theorem, we need to encode the proof in a certain way such that it is possible to achieve good soundness even by reading only 3 proof bits. To this end, we will use a coding method called *long code.* An assignment to the i-the entry of the proof $\pi(i) = w$ is encoded by the function $f : \{-1, +1\}^W \to [W]$ such that $f(x_1, \ldots, x_W) = x_w$. In other words $f = \chi_{\{w\}}$, the corresponding codeword being the truth table of the function that has 2^W entries. Such functions are also called *dictator functions.* Note that the set of functions has size only W (and thus requires only $\log W$ bits for representation) but we are representing it with a codeword of length 2^W.

In the case of a correct proof of ϕ where $\pi(i) = w$, the verifier V_H will thus expect the codeword corresponding to the truth table of $\chi_{\{w\}}$. In other words, it will expect a function f such that the fourier coefficient corresponding to $\chi_{\{w\}}$ i.e. $\hat{f}_{\{w\}}$ is significant. The decoding of long codes is done in a manner similar to Walsh-Hadamard codes. Recall that the decoding in WH codes was done through a randomized linearity test. However, to achieve good soundness using only three proof bits it will be necessary to introduce *noise* and slightly change the test. This will decrease the completeness guarantee from 1, but not by much.

More concretely, let γ be a small constant in $(0, 1)$. Let $z \in \{-1, +1\}^W$ be a random vector where the i-th coordinate is chosen independently as

$$z_i = \begin{cases} +1 \text{ with probability } 1 - \gamma \\ -1 \text{ with probability } \gamma \end{cases}$$

The verifier selects z according to the distribution above and x, y uniformly at random from $\{-1, +1\}^W$ and accepts iff $f(x)f(y) = f(xyz)$. Intuitively speaking, the introduction of noise affects a function f of the correct form (i.e. $\chi_{\{w\}}$) differently from a function of the form χ_S where $|S|$ is large. This is because $\mathbb{E}_z(\chi_S(z)) = \prod_{i\in S} \mathbb{E}[z_i] = (1 - 2\gamma)^{|S|}$. Thus the noise has the effect of depressing contribution from $\hat{f}_S$ for large S, allowing us to conclude that the small S's must contribute a lot.

Before specifying the verifier in more detail, we will introduce a slight restriction on the encoding functions. Notice that the dictator functions are *odd* i.e. for all $v \in \{-1, +1\}^W$, $f(-v) = -f(v)$. In PCP parlance they are called *bifolded* functions. Thus V_H will also assume that the encodings are that of bifolded functions. This makes sure that the function corresponding to $\chi_\emptyset$ is not a codeword. Also, for a constraint $\phi_r(i, j)$ let $h : [W] \to [W]$ be the function describing ϕ_r. Let $h^{-1}(u)$ denote the set $\{w \in W : h(w) = u\}$. For $y \in \{-1, +1\}^W$ define $H^{-1}(y)$ to be a string in $\{-1, +1\}^W$ such that for every $w \in [W]$, the w-th bit of $H^{-1}(y)$ is $y_{h(w)}$. We are in a position to define Håstad's verifier V_H.

1. Pick a random constraint $\phi_r(i, j)$ in the $2CSP_W$ instance.

2. Let f and g be the functions in the proof for the two variables in ϕ_r represented by the projection function h.

3. Pick $x, y \in \{-1, +1\}^W$ uniformly at random.

4. Select the noise vector $z \in \{-1, +1\}^W$ by choosing each z_i independently to be $+1$ with probability $1 - \gamma$ and -1 with probability γ.

5. Test whether $f(x)g(y) = f(H^{-1}(y)xz)$.

Out of $n2^{W-1}$, only three bits are accessed in the proof, and yet we have the following remarkable theorem!

Theorem 6 *If the $2CSP_W$ instance ϕ is satisfiable, then there is a proof that V_H accepts with probability $1-\gamma$. On the other hand, if $val(\phi) \leq \epsilon$ then V_H accepts no proof with probability greater than $\frac{1}{2}+\delta$, where $\delta = \sqrt{\frac{\epsilon}{\gamma}}$.*

For the completeness part, if ϕ is satisfiable we will show that V_H accepts a long code encoding of a satisfying assignment with probability $1-\gamma$. If $f = \chi_{\{w\}}$ and $g = \chi_{\{u\}}$ are the long codes of two integers w, u which satisfy $h(w) = u$, then for $x, y \in \{-1,+1\}^W$

$$f(x)g(y) = x_w y_u$$

and

$$f(H^{-1}(y)xz) = H^{-1}(y)_w x_w z_w = y_{h(w)} x_w z_w = y_u x_w z_w$$

Thus $f(x)g(y) = f(H^{-1}(y)xz)$ whenever $z_w = 1$, which happens with probability $1-\gamma$, which is thus the completeness guarantee. We will not prove the soundness part here but give the main claims that lead to the proof. The key insight is that if (f,g,h) is accepted by V_H with probability significantly more than $\frac{1}{2}$, then f, g must be correlated. We have the following definition.

$$h_2(S) = \{u \in [W] : |h^{-1}(u) \cap S| \text{ is odd}\}$$

The following lemma holds which implies the soundness part of the theorem.

Lemma 7 *The probability of acceptance of V_H is given by*

$$\mathop{\mathbb{E}}_{f,g,h}\left[\frac{1+\sum_{S\subseteq[W]} \hat{f}_S^2 \hat{g}_{h_2(S)}(1-2\gamma)^{|S|}}{2}\right]$$

where the expectation probability is over the randomly picked constraints f, g, h. Further, if this probability is greater than $\frac{1}{2}+\delta$, then

$$\mathop{\mathbb{E}}_{f,g,h}\left[\sum_{S\subseteq[W]} \frac{\hat{f}_S^2 \hat{g}_{h_2(S)}^2}{|S|}\right] \geq \gamma\delta^2$$

Finally, given a proof $\tilde{\pi}$, we give a randomized strategy to get an assignment π of the variables in ϕ such that the probability that a constraint is satisfied is at least $\gamma\delta^2$. We can think of $\tilde{\pi}$ providing a $f_i : \{-1,+1\}^W \to \{-1,+1\}$ for each variable $i \in [n]$. The algorithm is as follows.

1. For variable $f = f_i$, the square of the fourier coefficients $\hat{f}_S$ for $S \subseteq [W]$ defines a distribution because $\sum_S \hat{f}_S^2 = 1$. Pick S with probability $\hat{f}_S^2$.
2. Pick random $w \in S$ and assign $\pi[i] = w$.

With the above assignment, it can be show that the probability that a randomly picked constraint (f,g,h) is satisfied is no less than

$$\sum_{S\subseteq[W]} \frac{\hat{f}_S^2 \hat{g}_{h_2(S)}^2}{|S|} \geq \gamma\delta^2.$$

Indeed, the probability that we sample the label for f from a set S and the label for g from a set T such that $T = h_2(S)$ is $\sum_{S\subseteq[W]} \hat{f}_S^2 \hat{g}_{h_2(S)}^2$; and after picking $v \in h_2(S)$, with chance at least $1/|S|$ we pick a $u \in S$ such that $h(u) = v$.