S^3CS 2014 - Sum of Squares - Homework 3.

1 Exercises related to the 3XOR lower bound, if you haven't done them before

Exercise 1. Prove that for every $\epsilon > 0$, there exists c such that for every set of cn vectors $a_1, \ldots, a_m \in \mathsf{GF}(2)^n$, if we pick $b \in \mathsf{GF}(2)^m$ at random then with probability at least 0.99, for every $x \in \mathsf{GF}(2)^n$, there number of $i \in [m]$ such that $\langle a_i, x \rangle = b_i$ is at most $(1/2 + \epsilon)m$.

Exercise 2. Prove that for every constant $c \geq 1$, there exists a constant $\delta > 0$ and a bipartite graph $G = (L \cup R, E)$ such that every vertex in L is connected to three vertices in R, and for every set S of size at most ϵn, the size of $\Gamma(S) = \{j \in R : (i,j) \in E\}$ is at least $1.6|S|$. See footnote for hint[1]

Exercise 3. Using the lower bound for random 3LIN as a black box, prove the following statement: There exists a number $\epsilon > 0$ 3SAT formula φ with $m = O(n)$ clauses and n variables, such that every assignment to φ satisfies at most $(7/8 + 0.01)m$ of its clauses, but there exists a degree ϵn pseudo-distribution $\{x\}$ that satisfies all the constraints of φ. See footnote for hint[2]

2 Exercises related to dictionary learning

Exercise 4. Prove that

$$\|v\|_4^{16} \preceq \|v\|_8^8 \|v\|_2^8$$

(this is the case $t = 8$, $d = 4$ of the inequality used in the dictionary learning paper). See footnote for hint[3]

Exercise 5. We showed that the polynomial $P(u) = \mathbb{E}\langle x, A^\top u \rangle^d$ approximates the polynomial $\|A^\top u\|_d^d$ in the sense that the maximum value of $P(u)$ is within $o(1)$ additive factor of the maximum value of $\|A^\top u\|_d^d$. Since these values are both at least 1, this is also a $1 \pm o(1)$ multiplicative approximation of this maximum. Show that in contrast, for typical vectors, the multiplicative approximation can be very poor even for $d = 4$ and even when the rows of A are simply an orthonormal basis.

That is, when $m = n$ and A is the standard basis and $\tau = 1/\log n$, show that there exists a $(4, \tau)$ nice distribution $\{x\}$ such that with probability 0.99 over the choice of a random vector u scaled to unit norm,

$$P(u) > n^{1/10} \|A^\top u\|_d^d$$

[1] **Hint:** For k that is more than some small constant, if S is a set of size k and T is a set of size at most $1.6k$, the probability that $\Gamma(S) \subseteq T$ is small enough that we can do a union bound over all such sets S and T. For smaller sets, we can bound the expectation of the number of such sets S that violate this condition (which using Markov with decent probability we will not have much more than that) and this number will be small enough that we can remove all left vertices involved in them.

[2] **Hint:** Use the simple (but useful!) fact that zero is an even number, namely that if $a \oplus b \oplus c = 1$ then $a \vee b \vee c = 1$.

[3] **Hint:** Use repeatedly the inequality $PQ \preceq \frac{1}{2}P^2 + \frac{1}{2}Q^2$.

In some sense the power of the SOS algorithm is exactly that it is able to find a maximizer for $\|A^\top u\|_d^d$ despite only having access to this very rough approximation $P(u)$. Intuitively (though we haven't formalized it, let alone prove it), local search algorithms such as gradient descent may have trouble identifying the local maximum because on 99% of the space the behavior of $P(u)$ and $\|A^\top u\|_d^d$ is so different. [As a bonus exercise (and one that I haven't done myself) find a distribution $\{x\}$ for which the polynomial $P(u)$ will have a local maximum that is not close to any vector in the set $\{\pm a^1, \ldots, \pm a^k\}$.]

3 Exercises related to ARV

Exercise 6. Let $G = (V, E)$ be a d-regular graph, and suppose that we are given a degree 6 pseudo-distribution $\{x\}$ satisfying $\{x_i^2 = 1\}$, $\{\sum x_i = 0\}$,

$$\tfrac{1}{|E|} \sum_{(i,j)\in E} \tilde{\mathbb{E}}(x_i - x_j)^2 \leq \phi$$

1. (Done yesteday) Prove that if there exists an actual distribution $\{z\}$ over $\{\pm 1\}^n$ such that $\mathbb{E}(\sum z_i)^2 = o(n^2)$ and for every i, j $\mathbb{E}(y_i - y_j)^2 \leq \sqrt{\tilde{\mathbb{E}}(x_i - x_j)^2}$, then there exists a set S with $E(S, \overline{S}) \leq O(\sqrt{\phi}) d|S||\overline{S}|/n$.

2. Prove that if there exist subsets L, R of V with size at least $n/100$ such that $\mathbb{E}(x_i - x_j)^2 \geq \Delta$ for every $i \in L$ and $j \in R$, them there exists a set S with $E(S, \overline{S}) \leq O(\phi/\Delta) d|S||\overline{S}|/n$. See footnote for hint[4]

Exercise 7. A *metric* is a function $d : \Omega^2 \to [0, \infty)$ such that $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$ for every $x, y, z \in \Omega$. For $c, p, q \geq 1$, we say that a metric d is c-embeddable into ℓ_p^q if there exists some n and some map $f : \Omega \to \mathbb{R}^n$ such that

$$\|f(x) - f(y)\|_p^q \leq d(x, y) \leq c\|f(x) - f(y)\|_p^q$$

for every $x, y \in \Omega$, where $\|v\|_p = (\sum_{i=1}^n |f(x)_i - f(y)_i|^p)^{1/p}$. If d is 1-embeddable into ℓ_p^q then we simply say that it is an ℓ_p^q metric.

1. Prove that if d is an ℓ_1 metric then there exists some positive number c and a distribution $\{S\}$ over subsets of Ω such that $d(x, y) = c\Pr[x \in S \wedge y \notin S]$.

2. Show that if every ℓ_2^2 metric is c embedable into ℓ_1 then there is an $O(c)$ approximation algorithm for the uniform sparsest cut problem. Recall that we saw an example, based on the hypercube, showing an ℓ_2^2 graph with a constant average distance but without a good vertex separator. Does this example show that c cannot be $o(\sqrt{\log |\Omega|})$?

[4]**Hint:**Pick t at random in $[0, \Delta]$ and let S be the set of vertices i such that there exists $j \in L$ with $\tilde{\mathbb{E}}(x_i - x_j)^2 \leq t$. Use the squared triangle inequality to argue that for every edge (i, j), the probability that $i \in S$ and $j \in \overline{S}$ is at most $\tilde{\mathbb{E}}(x_i - x_j)^2/\Delta$.