# Some Homework Problems (Under Construction)

## Some Gap Reductions

**Problem 1.** Let MAX 3SAT($c$) denote the MAX 3SAT problem restricted to formulas in which each variable occurs at most $c$ times. In lecture, we saw that there is a (sufficiently large) constant $c$ such that MAX 3SAT($c$) is NP-hard to approximate to within some constant. Use this fact as a starting point to minimize $c$—i.e., prove that MAX 3SAT($c$) is NP-hard to approximate to within some constant, where $c$ is as small as possible.

**Problem 2.** Assuming the PCP Theorem, prove that the Vertex Cover problem is NP-hard to approximate to within some constant, even in bounded-degree graphs. How small can you make the degree bound?

**Problem 3.** Assuming the PCP Theorem, prove that the Steiner Tree problem is NP-hard to approximate to within some constant. Can you prove this even for instances that are complete graphs in which all edge costs are either 1 or 2?

[Hint: Reduce from Vertex Cover in bounded-degree graphs. Given a VC instance $G = (V, E)$, construct a Steiner tree instance on the vertex set $V \cup E$.]

**Problem 4.** Assume Feige's hardness result for Set Cover: unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$, there is no $(1 - \delta) \ln n$-approximation algorithm for Unweighted Set Cover (where $\delta > 0$ is any fixed constant).

Use this result to prove the following: unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$, there is no $(1 - 1/e + \epsilon)$-approximation algorithm for Unweighted Set Coverage (where $\epsilon > 0$ is any fixed constant). (Recall that in the Set Coverage problem you are given a set system and a budget $k$, and the goal is to cover as many elements as possible using at most $k$ sets.)

**Problem 5.** Assume Feige's hardness result to prove the following: unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$, there is no $(1 + 2/e - \epsilon)$-approximation algorithm for metric $k$-median (where $\epsilon > 0$ is any fixed constant). (Recall that in metric $k$ median you are given facilities $F$, demands $D$, a metric $c$ on $F \cup D$, and costs $f$ of facilities. The goal is to open at most $k$ facilities $A \subseteq F$ and assign each demand $j \in D$ to an open facility $i(j) \in A$ to minimize the sum of the facility costs and assignment costs: $\sum_{i \in A} f_i + \sum_{j \in D} c_{i(j),j}$.)

## Variants on the PCP Theorem and its Proof

**Problem 1.** Recall in our analysis of Gap Amplification (and the Powering step in particular), we proved that our procedure doubled the satisfiability gap *provided the gap was less than some constant.* This problem shows that the analysis is tight, in the sense that for large gaps the Powering procedure need not increase the gap by a multiplicative factor.

(a) Assume that for infinitely many constants $d$, there exist infinitely many $d$-regular graphs with $n$ vertices, second eigenvalue (in magnitude, of the adjacency matrix) at most $2\sqrt{d}$, and girth at least $\frac{2}{3} \log_d n$. (This is indeed the case.) Give each edge of such a graph an inequality constraint (over the Boolean alphabet, say). Prove that the satisfiability gap of such a constraint graph is at least $1/2 - O(\sqrt{1/d})$.

[Hint: Look up the "expander mixing lemma", e.g. in Wikipedia.]

(b) On the other hand, show that for every fixed constant $t$, if $n$ is large enough, then the output $G'$ of the Powering step has satisfiability gap at most $1/2$.

## Some Expander Problems

**Problem 1.** This problem explores expander-based error reduction. Let $G$ be a $d$-regular graph on $n$ vertices, where $d$ is an absolute constant. Let $A$ be its adjacency matrix, and let $\lambda < d$ be the second-largest magnitude of one of its eigenvalues. Let $\tilde{A} = A/d$ denote the normalized adjacency matrix.

Let $B \subseteq V$ denote a set of $\alpha n$ ("bad") vertices (where $\alpha \in (0,1)$). Note that if we independently chose $t$ random vertices of $G$ (with replacement), then the probability that all of them lie in $B$ is precisely $\alpha^t$. Let $\gamma$ denote the probability that a $t$-step random walk in $G$ (with the start state chosen uniformly at random) only visits vertices of $B$. We aim to show a comparable bound for $\gamma$, provided $\lambda$ is bounded away from $d$.

(a) Let $P$ denote the $V \times V$ projection matrix such that $Px$ zeroes out the coordinates of $x$ corresponding to $V \setminus B$. Prove that $\gamma = \|(PA)^t P x_0\|_1$, where $x_0$ denotes the uniform probability distribution and $\|\cdot\|_1$ denotes the $\ell_1$ norm.

(b) The 2-norm of a matrix $C$ is defined as the maximum factor by which $C$ can elongate a vector (with respect to the Euclidean norm): $\|C\|_2 = \max_{y \neq 0} \|Cy\|_2 / \|y\|_2$. Prove that $\gamma \leq \alpha \|(PA)^t P\|_2 \leq (\|AP\|_2)^t$.

(c) Prove that $\|AP\|_2 \leq \sqrt{\alpha + (\lambda/d)^2}$ and conclude that $\gamma \leq (\alpha + (\lambda/d)^2)^{t/2}$. (Hint: Review the proof of the Key Expander Lemma used in the analysis of the Powering step.) [Bonus: prove that $\|PAP\|_2 \leq \lambda/d + \alpha(1 - \lambda/d)$ and use this to establish the sharper upper bound of $\gamma \leq \alpha(\lambda/d + \alpha(1-\lambda/d)^t)$.]

(d) Application to RP error reduction: suppose we have an RP algorithm—that is, on NO instances the algorithm always correctly rejects, while for YES instances the algorithm accepts with probability at least $1/4$. Assume that this algorithm uses $r$ random bits. Note that naive serial repetition reduces the error probability to $(3/4)^t$ while using $rt$ random bits. Show that the same error probability can be achieved using only $O(r+t)$ random bits. [Remember that while $d - \lambda = \Omega(1)$, it might be an extremely small constant.]

(e) Combine part (c) with the FGLSS reduction. By how much can you improve the result shown in class (both in terms of the hardness factor and in the complexity assumption required)?

**Problem 2.** Prove that for a sufficiently large constant $d$, there exists a $d$-regular graph on $d^4$ vertices such that the second-largest magnitude of an eigenvalue of its adjacency matrix is at most $d/8$.

**Problem 3.** Suppose $G$ is an expander on $cn$ vertices for some $c > 1$ (according to either the combinatorial or algebraic definition). But what you really want is an expander on $n$ vertices. Can you obtain one easily from $G$, for example, by contracting some of its edges? Quantify the impact of your proposed solution on the expansion of graph.

## Some Fourier Analysis Problems

**Problem 1.** Give an alternative proof of the soundness of the BLR linearity test using the following ideas. Let $f$ be a Boolean function defined on $\ell$-bit strings. Suppose the BLR linearity test rejects with probability at most $\delta$ for $\delta$ sufficiently small. Obtain the function $g$ from $f$ as follows: for every $\ell$-bit string $x$, $g(x)$ is the most frequent result of the computation $f(x+y) - f(x)$ over all possibly shifts $y$. Prove that this works, in the sense that $g$ is a uniquely defined linear function that is $2\delta$-close to $f$.