Locally Testable Codes and Small-Set Expanders

Guru Guruganesh

December 8, 2014

A report based on the paper
"Making the Long Code Shorter, with applications to the Unique Games Conjecture" by Barak,Gopalan, Hastad, Meka,Raghavendra, Steurer

1 Introduction

1.1 History and Motivation

The unique games conjecture has been a very influential conjecture with many striking and profound implications. Originally proposed by Khot [Kho02], a positive resolution to the conjecture, would imply that many natural and important questions can be approximated by a basic SDP algorithm (see [Rag08]) and that no algorithm can do better (unless $P = NP$). Many important practical problems such as MAX-Cut, Kernel Clustering, Min Vertex Cover all fall under this category (see [KKMO07, KR08, RS09])

One peculiarity about this conjecture is that many of the implications of this conjecture do not extend both ways. In particular, it could be true that all the implications from UGC are still true, however the conjecture itself is false. This was a stumbling block for a while, as it was unclear what a refutation of the conjecture would imply. Furthermore, expanders which form the "hard" instances for many problems turned out to be easy for the Unique Games Conjecture (see [AKK$^+$08]).

The next observation was made by Raghvendra and Steurer (see [RS10]) who showed that the following natural combinatorial problem reduced to Unique Games. Hence, a refutation of the unique games conjecture would give improved algorithms for the small set expansion problem. Recall, the expansion of a set $S \subseteq V$ is the parameter $\Phi(S) = \frac{\mathbb{E}(S,\bar{S})}{d|S|}$.

Conjecture 1.1 ***(Small Set Expansion Conjecture)*** *For any $\varepsilon > 0$, there exists a $\delta \in (0,1)$ so that given a d-regular graph $G = (V, E)$, it is NP-Hard to*

1. *Accept if there exists $S \subseteq V$ with $|S| \leq \delta n$ have $\phi(S) \leq \varepsilon$.*
2. *Reject if all sets $S \subseteq V$ with $|S| \leq \delta n$ have $\phi(S) \geq 1 - \varepsilon$.*

The rejection instance of this problem, is usually called a (δ, ε)-small set expander. Formally,

Definition 1.2 *A graph $G = (V, E)$ is called a (δ, ε)-small set expander if for any $S \subseteq V$ with $|S| \leq \delta n$ we have $\Phi(S) \geq 1 - \varepsilon$.*

Thus, while unique games on expanders is easy, their status on small-set expanders is unknown (i.e. no polynomial time algorithm is known).

Arora, Barak, Steurer (see [ABS10]) utilized this connection to solve the unique games problem in sub-exponential time. Their chief observation was the fact that that any graph with which is an (δ, ε) small set expander can have only n^{ε} eigenvalues greater than $1 - \varepsilon$ (henceforth referred to as "top" eigenvalues). An intuitive reason for this is note that the large eigenvalues would correspond to sets that do not expand. This is the famous Cheeger's inequality (see [Alo86]). A large number of eigenvalues would imply that one can find many eigenvectors which are orthogonal to each other. They show that this is not possible unless one of those sets is very small. Their algorithm than searches for all subsets whose vector is close to the space spanned by the "top" eigenvectors.

The best known lower-bound for the number of "top" eigenvalues stemmed from the noisy hypercube which achieved $\log(n)$ eigenvalues. Showing that this was the optimal bound would have meant a tantalizing possibility as this would represent a quasi-polynomial time algorithm for the SSEproblem. This hope was crushed by Barak et al. in their paper [BGH$^+$12] who showed the existence of small set expanders with a large number of "top" eigenvalues. In this note, we will examine this result in detail.

1.2 Other Results

The paper of Barak et al. also shows that their construction can be viewed as an efficient alternative for the long code (which is ubiquitous in UG-hardness reductions). Their paper shows that many of the properties which hold for the long-code such as the invariance principle and the "majority is stablest" theorem also hold for the new short code.

This allows them to that the Max Cut reduction can be reduced to an $n \cdot qpoly(k)$ instance of the usual $n2^{O(k)}$ that comes from the long code reduction. They also improve existing Sherali Adams +SDP for the unique games problem from poly(log(log(n))) rounds to exp(poly(log(log(n)))) rounds. While these results are important and interesting in their own right, we will only focus on the connection small set expanders from locally testable codes.

2 Locally Testable Codes

A locally testable code (LTC) is any code which has an algorithm that given a string x, can determine if x is a valid codeword or very far from a code by querying a small number of locations in x. These tests are probabilistic and we only require them to be always accept actual codewords in C and reject elements that are "far away" from C with high probability. We make this precise in the definition below.

Definition 2.1 *Let $\mathcal{C}$ be a $[N, K, D]_2$ linear code that is a K-dimensional subspace of $\mathbb{F}_2^n$. We say that $\mathcal{C}$ is an LTC if there exists a verifier which makes q queries to a given word x and accepts with probability 1 if $x \in C$ and rejects with probability $\geq s(k)$ all words x with distance at least k.*

Definition 2.2 *A canonical tester for a code C is a distribution $\mathcal{T}$ over $\mathcal{C}^\perp$ such that the tester accepts if $\Pr_{q\sim\mathcal{T}}[q \cdot x = 0]$ and rejects otherwise. For an $x \in \mathbb{F}_2^n$, we say that $s_\mathcal{T}(x) := \Pr_{q\sim\mathcal{T}}[q \cdot x = 1]$ We define the soundness curve of this tester as*

$$s_\mathcal{T}(k) := \min_{\substack{\alpha\in\mathbb{F}_2^n \\ \Delta(\alpha,\mathcal{C})\geq k}} s_\mathcal{T}(\alpha)$$

It was shown in [BSHR05] that every verifier for a linear code can be converted into a canonical tester. For the remainder, we will assume that every verifier will be of this form.

LTCs have received a lot of attention in the computer science literature. They are considered a "combinatorial counterparts" to the celebrated PCP theorem. Two distinct regimes have been well studied. In this note, we are particularly interested in the high rate regime. Our codes will have a constant distance, and high rate; forcing the dual code to have a small number of words. We say a tester is ε-smooth, if $\Pr_{q\sim\mathcal{T}}[q_i = 1] = \varepsilon$, (i.e. the coordinates are queried in an unbiased fashion.) A tester is said to be ε-2-smooth if in addition, for any two distinct coordinates $i \neq j$, $\Pr_{q\sim\mathcal{T}}[q_i \neq q_j = 1] = \varepsilon^2$.

Lemma 2.3 *If $\mathcal{T}$ is a ε-smooth canonical tester, then $s_\mathcal{T}(\alpha) \leq \Delta(\alpha, \mathcal{C})\varepsilon$ for all $\alpha \in \mathbb{F}_2^n$. Furthermore, if $\mathcal{T}$ is ε-2-smooth, then $s_\mathcal{T}(\alpha) \geq (1-\gamma)\cdot\Delta(\alpha, C)\varepsilon$ where $\Delta(\alpha, C)\varepsilon \leq \gamma$.*

Proof: Let $x \in \mathbb{F}_2^n$ and $k = \Delta(x, C)$. By renaming the coordinates, we can say that Pr probability of rejection $= s_\mathcal{T}(x) \leq \Pr_{q\sim\mathcal{T}}[q_1 = 1] + \Pr_{q\sim\mathcal{T}}[q_2 = 1] + \cdots + \Pr_{q\sim\mathcal{T}}[q_k = 1]$. By the ε-smoothness of the tester $\mathcal{T}$, one can conclude that $s_\mathcal{T}(x) \leq k\varepsilon$.

Using principle of inclusion-exclusion we can say that $s_\mathcal{T}(x) \geq \sum_{i=0}^n \Pr_{q\sim\mathcal{T}}[q_i = 1] - \sum_{i,j} \Pr_{q\sim\mathcal{T}}[q_i = q_j = 1]$. This is at least $k\varepsilon - k^2\varepsilon^2 \geq k\varepsilon(1 - k\varepsilon)$. ■

3 Constructing Small Set Expanders

We define a graph $\mathrm{Cay}_\mathcal{C} = (\mathcal{C}^\perp, \mathcal{T})$ whose vertex set consists of all $x \in \mathcal{C}^\perp$. We connect x to $x + q$ and assign it weight $\Pr[q \sim \mathcal{T}]$ for all $q \sim \mathcal{T}$. Due to the special structure of this graph, we can calculate the eigenvectors exactly and they will correspond to characters (a fact that is true for all Cayley graphs constructed from abelian groups).

The central observation is to notice that the eigenvalues of this graph are related to the soundness of the canonical tester $\mathcal{T}$.

3.1 Cayley Graphs

Lemma 3.1 *The eigenvectors of* $Cay_{\mathcal{C}}$ *are denoted by* $\chi_x(v) = -1^{\langle x,v\rangle}$*. Furthermore, two eigenvectors* χ_x *and* χ_y *are the same iff* $x + y \in \mathcal{C}$*. Lastly, the eigenvalue* λ_x *associated with* χ_x *is* $\lambda_x = 1 - 2s_{\mathcal{T}}(x)$*.*

Proof: Firstly, we will show that χ_x (as defined above) is an eigenvector. To show this we simply multiply χ_x by the adjacency matrix of Cay_{RM} (denoted as A).

$$\begin{aligned}
&(A \cdot \chi_x)(v) \\
&= \sum_q \Pr[q \sim \mathcal{T}] \cdot \chi_x(v+q) \\
&= \mathbb{E}_{q\sim\mathcal{T}}[\chi_x(q+v)] \\
&= \mathbb{E}_{q\sim\mathcal{T}}[\chi_x(q) \cdot \chi_x(v)] \\
&= \mathbb{E}_{q\sim\mathcal{T}}[\chi_x(q)]\chi_x(v) \\
&= \mathbb{E}_{q\sim\mathcal{T}}[-1^{\langle x,q\rangle}]\chi_x(v)
\end{aligned}$$

Observing that $\mathbb{E}[-1^{\langle x,q\rangle}] = 1 - 2\Pr[\langle x, q\rangle = 1]$.

$$\begin{aligned}
&= \big(1 - 2\Pr_{q\sim\mathcal{T}}[\langle x, q\rangle = 1]\big)\chi_x(v) \\
&= \big(1 - 2s_{\mathcal{T}}(x)\big)\chi_x(v)
\end{aligned}$$

From the previous calculation, one can deduce that χ_x is an eigenvector and has eigenvalue $\lambda_x = 1 - 2s_{\mathcal{T}}(x)$.

Suppose we have that χ_x and χ_y be the same value. Then note that $-1^{\langle\alpha,x\rangle} = -1^{\langle\alpha,y\rangle}$ for all $\alpha \in \mathcal{C}^{\perp}$. This implies that $\langle\alpha, x - y\rangle = 0$ for all $\alpha \in \mathcal{C}^{\perp}$. Since this is the exact characterization of the dual, we can say that $x - y \in \mathcal{C}$. Hence for each coset $\mathbb{F}_2^n/\mathcal{C}$, we generate a unique eigenvector. ■

Since there is one eigenvector for each coset, we will choose the minimum weight (non-zero positions) to be the unique representative for this coset.

3.2 Large Number of "Top" Eigenvalues

If the original code has distance at least 1, then note that any two characters corresponding to χ_x will be different. We will now show that they also correspond to vectors that have large eigenvalues.

Lemma 3.2 *There are at least* $\frac{n}{2}$ *eigenvectors whose eigenvalues are greater than* $1 - 4 \cdot \varepsilon$*.*

Proof: Consider the characteristics that correspond to the dictators or $e_i \in \mathbb{F}_2^n$. In particular, $\lambda_i = 1 - 2\mathrm{Pr}_{q \sim \mathcal{T}}[q_i = 1]$. Note that we know that we have a ε-smooth tester which means

$$\sum_{i=1}^{n} \mathrm{Pr}_{q \sim \mathcal{T}}[q_i = 1] \leq \varepsilon n$$

By Markov's inequality, we can say that at least $\frac{n}{2}$ of these will have eigenvalue at most $\mathrm{Pr}_{q \sim \mathcal{T}}[q_i = 1] \leq 2\varepsilon$. Hence, proving the claim. ∎

3.3 Expansion

Expansion of this graph stems from the hypercontractivity of the eigenspace spanned by the top eigenvectors. In an intuitive sense, hypercontractivity is a bound on the 4-norm with respect to the 2-norm. The main idea is that if a function is very spiky or concentrated on a few co-ordinates, then its 4-norm is much higher than its 2-norm. However, if a function has a subspace where the 4-norm is a constant factor away from 2 norm, it must be fairly spread out. In other words, its support cannot be very sparse. It is important to understand that these values only make sense as the size of the subspace grows to infinity and the hypercontractivityj constant is independent of the size of the subspace.

The second connection is that given any small set expander $S \subseteq V$, its indicator vector will be close to the top eigenspace. This is by a higher order analogue of Cheeger's inequality, we know that every set with small expansion must have a large eigenvalue. However, if the top eigenspace is hypercontractive, then the indicator vector of any small set cannot be in it. Hence, we can conclude that the small sets will expand. We will make these arguments precise below:

Lemma 3.3 *For every vertex subset $S \subseteq \mathcal{C}^{\perp}$ for the graph $(\mathcal{C}^{\perp}, \mathcal{T})$, and every $k < \frac{D}{5}$ we have $\Phi(S) \geq 2s_{\mathcal{T}}(k) - 3^k \mu(S)^{\frac{1}{2}}$.*

Proof: Consider the set of cosets whose eigenvectors have minimum weight at most k where we will pick $k < \frac{D}{4}$. Given any set S let v be the indicator vector of this set and let $\mu(S)$ be the size of this set. We know that $\phi(S) = 1 - \mathrm{Pr}_{q \sim \mathcal{T}}[(x+q) \in S \mid x \in S]$. Therefore we say that

$$\begin{aligned} &\mu(S)(1 - \Phi(S)) \\ &= \mathbb{E}_{x \in S, y \in N(S)}[v(x)v(y)] \\ &= \langle v, Av \rangle \end{aligned}$$

Noting that v can be written as $v = \sum_{\alpha} \hat{v}_{\alpha} \chi_{\alpha}$ and the fact that χ_x are orthonormal vectors

$$= \sum_{\alpha \in \mathbb{F}_2^n / \mathcal{C}} \lambda_{\alpha} \hat{v}_{\alpha}^2$$

$$= \sum_{\substack{\alpha \in \mathbb{F}_2^n/\mathcal{C} \\ |\alpha| \le k}} \lambda_\alpha \hat{v}_\alpha^2 + \sum_{\substack{\alpha \in \mathbb{F}_2^n/\mathcal{C} \\ |\alpha| > k}} \lambda_\alpha \hat{v}_\alpha^2$$

Noting that the $1 - 2s_{\mathcal{T}}(k)$ is monotonic and bounded by 1

$$\le \sum_{\substack{\alpha \in \mathbb{F}_2^n/\mathcal{C} \\ |\alpha| \le k}} \hat{v}_\alpha^2 + (1 - 2s_{\mathcal{T}}(k))\|v\|_2^2$$

Rewriting this, we get that $\phi(S) \ge 2s_{\mathcal{T}}(k) - \frac{1}{\mu(S)}\left\|v^{\le k}\right\|_2^2$. Hence any bound on $\left\|v^{\le k}\right\|_2^2$ would give us a lower bound for the expansion of any set S. Let $V^{\le k}$ be the space spanned by χ_α where $|\alpha| \le k$ and define $V_{p \to q}^{\le k} := \max_x \frac{\left\|P_{V^{\le k}} x\right\|_q}{\|x\|_p}$.

Using this new definition, we can bound $\left\|v^{\le k}\right\|_2^2 \le (V_{4/3 \to 2}^{\le k})^2 \|v\|_{4/3}^2 \le (V_{4/3 \to 2}^{\le k})^2 \mu(S)^{\frac{3}{2}}$. Finally, we use the well known fact in boolean analysis that for any subspace R we have that $R_{4/3 \to 2} \le R_{2 \to 4}$. All that remains is to bound $V_{2 \to 4}^{\le k}$.

Let v be any arbitrary vector and let $\hat{v}_\alpha$ be its Fourier coefficients. Define the function $g : \{0,1\}^n \to \mathbb{R}$ as

$$g := \sum_{\substack{\alpha \in \mathbb{F}_2^n/\mathcal{C} \\ |\alpha| \le k}} \hat{v}_\alpha \chi_\alpha$$

Since the original code $\mathcal{C}$ had distance D, we know that the dual code is $D-1$-wise independent. Since $k < D/5$, we know that $\mathbb{E}_{x \sim \{0,1\}^n}[g(x)^4]^{\frac{1}{4}} = \mathbb{E}_{x \sim \mathcal{C}^\perp}[g(x)]^{\frac{1}{4}} = \left\|v^{\le k}\right\|_4$. By the hypercontractive inequality applied to g, we can now say that

$$\begin{aligned}
\left\|v^{\le k}\right\|_4 &= \mathbb{E}_{x \sim \{0,1\}^n}[g(x)^4]^{\frac{1}{4}} \\
&\le \sqrt{3}^k \mathbb{E}_{x \sim \{0,1\}^n}[g(x)^2]^{\frac{1}{2}} \\
&\le \sqrt{3}^k \left\|v^{\le k}\right\|_2 \\
&\le \sqrt{3}^k \|v\|_2
\end{aligned}$$

Thus we can conclude that $(V_{2 \to 4}^{\le k})^2 \le 3^k$. Substituting this into the original bound, we get the required statement. ■

3.4 Reed Muller Codes and Continuous Random Works

So far, we have shown a mechanism that given a code and a canonical tester with small query complexity and with good soundness parameters, produces graphs that are small set expanders yet contain a large number of "top" eigenvalues.

In this section, we will instantiate this with Reed Muller (RM) codes , and fine tune the parameters. Our RM code $\mathcal{C}$ will consist of all n bit polynomials with degree at most $n-d+1$ and distance 2^{d+1}. Note that the dual code is also a RM code with max degree d. Hence we note that the $\mathcal{C}^{\perp}$ has $\sum_{i=0}^{d}\binom{n}{i}$ codes. The graph will consist of $\text{Cay}_{RM} = (\text{RM}_{n,d}, \mathcal{T})$ where the tester queries words which are products of d-affine variables.

In [BKS+10], show that there exists an efficient tester which has a good soundness curve. Their tester samples a minimum weight code with equal probability. It is known that minimum weight codes correspond to products of d-affine variables. These code words have weight 2^{n-d}. The number of large eigenvalues is related to the weight of these words, we can say that $\varepsilon = 2^{-d}$. The following theorem outlines the soundness parameters of the graph.

Theorem 3.4 *[?] There exists a constant ν_0 such that for all n, d, k, $k < \nu_0 \cdot 2^d$ the tester $\mathcal{T}_{RM}$ which queries only εN positions and has soundness $s(k) \geq \frac{k}{2}2^{-d}$.*

Substituting theorem 3.4 into lemma 3.3, and letting $k = \frac{\nu_0}{2}2^d$ (note that assuming $\nu_0 < \frac{1}{2}$, this also satisfies the requirement that $k < \frac{D}{5}$), we see that the expansion of all sets is at least $\Phi(S) \geq \frac{\nu}{2} - O(\frac{1}{n})$. Now we can "power this up" to get a graph with near perfect expansion by taking a random walk. One can run into potentially tricky issues regarding the number of steps to take and other discretization errors. To get around this, the authors propose to take a continuous random walk and reweigh the graph edges accordingly.

Lemma 3.5 *Given a graph G with adjacency matrix A_G, with eigenvectors $x_1, \ldots, x_n$ and eigenvalues $u_1, \ldots u_n$. The eigenvectors of the matrix $e^{t\cdot A}$ are $x_1, \ldots, x_n$ and have eigenvalues $e^{t\mu_1}, \ldots, e^{t\mu_n}$.*

Proof: Expanding e^{tA} as $I + tA + t^2A^2...$, and noting that all powers of A commute (hence share common eigenbasis), we note that they must have the same eigenvectors. A simple calculation shows that the associated eigenvalues are $e^{t\mu_i}$. ■

Let $A_{\text{Cay}_{RM}}$ denote the adjacency matrix of Cay_{RM}. By $\text{Cay}_{RM}(t)$ we denote the graph formed by the adjacency matrix $e^{-t(I-A_{\text{Cay}_{RM}})}$. The following technical lemma is useful in analyzing the precise structure of the graph and reasoning about the eigenvalue profile. We omit the proof as it is presented clearly in [BGH+12](see lemma 4.13).

Lemma 3.6 *Let $t = \varepsilon \cdot 2^{d+1}$ for $\varepsilon > 0$ and $\rho = e^{-\varepsilon}$. Let λ_α denote the eigenvalues of $Cay_{RM}(t)$ then there exist absolute constants μ_0, δ_0, such that if $deg(\alpha) = k$ then $\lambda_\alpha \leq \max\left(\rho^{k/2}, \rho^{\mu_0 2^d}\right)$*

Theorem 3.7 *For any $\varepsilon, \eta > 0$, there exists a graph G with $(2^n)^{\frac{1}{d}}$ eigenvalues larger than $1-4\varepsilon$ where $d = log(1/\varepsilon) + \log(\log(1/\eta)) + O(1)$ and where every set $S \subseteq G$ and has size $\mu(S) \leq \delta$ has expansion $\Phi(S) \geq 1 - \eta - 3^{O(\frac{\log(1/\eta)}{\varepsilon})}\sqrt{\mu(S)}$.*

Proof: Fix $l = \frac{c_1}{\varepsilon}\log(\frac{1}{\nu})$ so that

$$\exp(-\varepsilon\frac{l}{2}) = \eta \tag{1}$$

Next, we fix $d = \log(l) + O(1)$ so that it satisfies

$$l \le \min(\mu_0 2^{d+1}, 2^d/5) \tag{2}$$

where μ_0 is from the previous lemma.

Let $\mathrm{Cay}_{RM}(t)$ be the graph at time $t = \varepsilon 2^{d+1}$ and consider the eigenvalues associated with the eigenvectors α whose minimum weight 1.Therefore, it has eigenvalue $s_{\mathcal{T}}(1) \ge 2^{-d-1}$. Here the corresponding eigenvalue in $\mathrm{Cay}_{RM}(t)$ is

$$\begin{aligned}\lambda_\alpha &= \exp(-t(1-\mu_\alpha)) \\ &\ge \exp(-t(1-(1-2(2^{-d-1})))) \\ &\ge \exp(-t(2^{-d})) \\ &\ge \exp(-2\varepsilon) \\ &\ge 1-2\varepsilon\end{aligned}$$

Thus we can conclude there are n eigenvectors with eigenvalue at least $1-2\varepsilon$. Since we have about $O(n^d)$ vertices and there are about $n/2$ eigenvectors whose eigenvalues are high. We can conclude that there are about $2^{\log(|G|)^{\frac{1}{d}}}$ eigenvalues larger than $1-2\varepsilon$.

We note that hypercontractivity follows as the eigenvectors for $\mathrm{Cay}_{RM}(t)$ are the same as Cay_{RM}. We note that the proof follows verbatim and hence we can conclude that the space formed by the eigenvectors of weight l satisfy $V^{\le l}_{2\to 4} \le 3^l$. By applying Lemma 3.6, we can say that if the eigenvalues of all values are bounded by $\exp -\varepsilon\frac{l}{2} \le \eta$. Hence, we can say that for all sets we have $\phi(S) \ge 1-\eta-3^{\frac{c_1}{\varepsilon}\log(\frac{1}{\eta})}\sqrt{\mu(S)}$. ■

4 The Converse and Open Problems

In a follow up work by Gopalan, Vadhan, Zhou [GVZ14], the authors show that there is a converse relationship to the above construction. In particular, the previous construction demonstrated a method of generating Cayley graphs whose eigenvalues are similar to the boolean hypercube from locally testable codes. They complement this result by showing that a Cayley graph with some spectral properties can produce a locally testable code with good soundness and smoothness.

Let us say that a locally testable code has soundness δ if $\Pr_{q\sim\mathcal{T}}[\langle\alpha, q\rangle = 1] \ge \delta\Delta(r, \mathcal{C})$.

Note that given any Cayley Graph $(\mathcal{A} = \mathbb{F}_2^h, \mathcal{D})$ where $\mathcal{D}$ is a distribution on the $\mathcal{A}$. Let $b \in \mathcal{A}^*$ where $\mathcal{A}^*$ is the set of linear functions of the form $\mathcal{A} \to \mathbb{F}_2$. Then the characters of graph are defined by $\chi_b(\alpha) = -1^{b(\alpha)}$ for every $\alpha \in \mathcal{A}$ and $b \in \mathcal{A}^*$.

Definition 4.1 *Given a cayley graph $(\mathcal{A}, \mathcal{D})$ be a Cayley graph on the group $\mathcal{A} = \mathbb{F}_2^h$. Let $\mu, \xi \in [0,1]$ and $d \in [n]$. Let $B^* = b_1, \ldots, b_n$ be a set of d-wise independent set of generators for $\mathcal{A}^*$ of cardinality n. We say that $\mathcal{B}^*$ is a (μ, ξ) spectrum generator if satisfies the following properties:*

- ***Large Eigenvalues.*** *$\lambda_b \geq 1 - \mu$ for all $b \in \mathcal{B}^*$.*

- ***Spectral Decay.*** *For $a \in \mathcal{A}^*$, $\lambda_a \leq 1 - \xi \cdot rank_{\mathcal{B}^*}(a)$.*

Theorem 4.2 *Let $Cay(\mathcal{A} = \mathbb{F}_2^h, \mathcal{D})$ be a Cayley graph and let $\mathcal{B}^* = \{b_1, \ldots, b_n\}$ be a (μ, ξ)-spectrum generator for it. Then we can construct a $\mathcal{C} = [n, n-h, d]_2$ linear code and $\mathcal{D}$ can be used to generate a $\mu/2$-smooth tester and soundness $\xi/2$.*

This work shows that there is a natural equivalence between Cayley graphs that are small set expanders and locally testable codes with good soundness and smoothness properties.

While this work closes the bound between small-set expanders with many top eigenvalues, there is still a gap. Closing this gap would improve the current understanding on algorithmic challenges to solving the Unique Games Conjecture.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 563–572. IEEE, 2010.

[AKK+08] Sanjeev Arora, Subhash A Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K Vishnoi. Unique games on expanding constraint graphs are easy. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 21–28. ACM, 2008.

[Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[BGH+12] Boaz Barak, Parikshit Gopalan, Johan Hastad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 370–379. IEEE, 2012.

[BKS+10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of reed-muller codes. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 488–497. IEEE, 2010.

[BSHR05] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3cnf properties are hard to test. *SIAM Journal on Computing*, 35(1):1–21, 2005.

[GVZ14] Parikshit Gopalan, Salil Vadhan, and Yuan Zhou. Locally testable codes and cayley graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 81–92. ACM, 2014.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775. ACM, 2002.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. ODonnell. Optimal inapproximability results for maxcut and other 2variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254. ACM, 2008.

[RS09] Prasad Raghavendra and David Steurer. Integrality gaps for strong sdp relaxations of unique games. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 575–585. IEEE, 2009.

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764. ACM, 2010.