Derandomizing Homomorphism Testing in General Groups

Amir Shpilka[*] Avi Wigderson[†]

Abstract

The main result of this paper is a near-optimal derandomization of the *affine* homomorphism test of Blum, Luby and Rubinfeld [11].

We show that for any groups G and Γ, and any *expanding* generating set S of G, the natural deramdomized version of the BLR test in which we pick an element x randomly from G and y randomly from S and test whether $f(x) \cdot f(y) = f(x \cdot y)$, performs nearly as well (depending of course on the expansion) as the original test. Moreover we show that the underlying homomorphism can be found by the natural local "belief propagation decoding".

We note that the original BLR test uses $2\log_2 |G|$ random bits, whereas the derandomized test uses only $(1 + o(1))\log_2 |G|$ random bits. This factor of 2 savings in the randomness complexity translates to a near quadratic savings in the length of the tables in the related locally testable codes (and possibly probabilistically checkable proofs which may use them).

Our result is a significant generalization of the recent result of [12], who proved such a result only for the groups $G = Z_p^m$ and $\Gamma = Z_p$. It is also an explicit version of the nonconstructive result of [18].

We use a simple combinatorial arguments and the transitivity of Cayley graphs (and this analysis gives optimal results up to constant factors). Previous techniques used the Fourier transform, a method which seems unextendable to general groups (and furthermore gives suboptimal bounds).

Finally, we provide a polynomial time (in $|G|$) construction of a (somewhat) small ($|G|^\epsilon$) set of expanding generators for *every* group G, which yield efficient testers of randomness $(1+\epsilon)\log |G|$ for G. This follows a simple derandomization of the probabilistic construction of [5], who showed that almost all logarithmic-size sets are expanding.

Our work motivates further study of similar derandomizations of other natural property testing procedures, especially those more relevant to the local testing of better codes and to PCPs.

1 Introduction

1.1 Property testers and randomness complexity

Let F be the family of all functions (from a given domain to a given range), and P a subset of these functions (those with property "P"). A tester T is a probabilistic algorithm that receives as input a (black box for) function $f \in F$, evaluates f on a set of points in the domain, and uses this information to accept or reject the input function f. Roughly speaking, T is a tester for the property

[*]The Weizmann Institute of Science, Rehovot Israel. Email `amir.shpilka@weizmann.ac.il`.

[†]Institute for Advanced Study, Princeton New Jersy, USA. Email: `avi@ias.edu`. Partially supported by NSF grant CCR-0324906.

P if every f in P is accepted with high probability, and every f which is "far" from P (in Hamming distance) is rejected with high probability.

This is the basic set up of property testing, by now a very large field[§] (see excellent surveys by Goldreich [17] and by Ron [28]). A central theme in this field is relating $error(T)$, the probability that our tester fails to give the correct output, to its "complexity" $query(T)$, measuring the number of domain samples it used, and its "accuracy" $distance(T)$, which is how far from P are the functions it rejects. The importance of this field for various applications follow numerous results giving testers for a variety of properties P, in which both *query* and *distance* depend only on *error* (and not on the size of the domain of the functions).

Central applications of this area are (the related) Locally Testable Codes (LTCs) and Probabilistically Checkable Proofs (PCPs). In these, the answers to all possible sets of queries are explicitly written down, and it is a major concern to minimize their length. This length can be seen to be directly related to (indeed, an exponential of) the number of random bits $random(T)$ used by the tester T, and so this parameter and its tradeoffs with the others have been investigated as well. Indeed such notions as "free bit complexity" and "amortized query complexity" [10, 19, 30, 20] lead to celebrated tight inapproximability results via optimizing such trade-offs in PCPs. Related are the "derandomized" amplification of hardness results [21, 31] which lead to optimal derandomization of BPP.

A recent paper of Goldreich and Sudan [18] addresses the minimization of $random(T)$ for two important testers: the homomorphism tester of Blum, Luby and Rubinfeld [11] (which was the first and motivating example of property testing of functions), and the "point vs. lines" low-degree tester of Rubinfeld and Sudan [29] (which was central in the proof of the PCP theorem). Both testers use randomness to name two random domain queries, and the motivation above of trying to bring down proof/code length from quadratic to nearly linear (in the length of the appropriate input), demands using randomness roughly sufficient for only one query. Moreover [18] showed that *nonuniformly* this is possible. Indeed, their arguments can achieve similar savings in much more general contexts of multi prover systems, but we will restrict our discussion from this point on only to the first tester for homomorphism, which is the subject of our paper.

1.2 Affine homomorphism testing

Given two finite groups G, Γ a homomorphism is a function $f : G \mapsto \Gamma$ such that for every $g_1, g_2 \in G$ we have that $f(g_1 \cdot g_2) = f(g_1) \cdot f(g_2)$. When the groups are abelian it is customary to use "+" instead of "$\cdot$", so a homomorphism is a function that for every $g_1, g_2 \in G$ satisfies $f(g_1 + g_2) = f(g_1) + f(g_2)$. This is the reason that in abelian groups homomorphisms are referred to as linear functions. In particular the famous paper of Blum et al [11] analyze homomorphism testing (which they call linearity testing) for abelian groups. An *affine* homomorphism between G and Γ is a function f such that $f(0)^{-1} \cdot f$ is a homomorphism. The BLR linearity testing can be slightly change to yield an affine version of their linearity test.

Let G and Γ be finite groups. Let F be all functions from G to Γ, let $P_{hom} \subseteq F$ be the set of all homomorphisms from G to Γ, and $P_{aff} \subseteq F$ be the set of all affine homomorphisms from G to Γ. For two functions $f, h \in F$ we have the normalized Hamming distance $dist(f, h) = Prob[f(x) \neq h(x)]$ for a uniform element $x \in G$.

[§]dealing with many other objects than functions, such as strings, distributions, graphs, etc

Fix a subset E of $G \times G$ (which may be viewed as a directed graph on G), and consider the following tester T_E. It picks uniformly a random pair $(x, y) \in E$, evaluates the input function f on the three (related) elements x, y and $x^{-1}y^{-1}$, and accepts iff it satisfies the equation $f(x)f(x^{-1}y)f(y)^{-1} = 1$. It is easy to see that if f is a homomorphism, than T_E will accept f with probability one. The interesting direction is showing that if the error of the test is small, than f is close to a homomorphism. We say that T_E is a (δ, ϵ)-test if every function that passes the test with probability at least $1 - \delta$ is at most ϵ far from having the property (either P_{hom} or P_{aff}).

The well known BLR linearity tester [11] uses (in this notation) $E = G \times G$. BLR proved that $T_{G\times G}$ is a $(\delta, 9\delta/2)$-test. However, their analysis wasn't tight and was later improved by [8, 7, 6]. Ben-Or, Coppersmith and Rubinfeld [9] extended the BLR result and showed that the test with $E = G \times G$ work for two general groups as well. The proof of Ben-Or et al is similar to the proof of the BLR.

Theorem 1.1 *[11, 9, 7, 8, 6] Let G, Γ be groups. Consider the test $T_{G\times G}$ described above for P_{hom}. For every $\delta > 0$, if f passes the test with probability $> 1 - \delta$ then there exist a homomorphism $h \in P_{hom}$ such that $dist(f, h) \leq \delta/3 + O(\delta^2)$. In other words, $T_{G\times G}$ is a $\left(\delta, \delta/3 + O(\delta^2)\right)$-test for P_{hom}.*

To save on randomness, [18] suggested to use sparser graphs E. The tester T_E obviously has $random(T_E) = \log |E|$. The maximum of this value is attained by the BLR test, $random(T_{G\times G}) = 2 \log_2 |G|^{\dagger}$. It is also easy to see that any nontrivial tester (giving any dependence between error and distance) must satisfy $random(T_E) \geq \log |G| - O(1)$. Goldreich and Sudan [18] showed that this lower bound can essentially be matched, and at negligible cost to the dependence of distance on error.

Theorem 1.2 *[18] For all but $exp(-|G|)$ fraction of all possible graphs E of size $C|G| \log |\Gamma|$ (with C an absolute constant) the following holds. For every $\delta > 0$, T_E is a $\left(\delta, \delta/3 + O(\delta^2) + exp(-|G|)\right)$-test for P_{hom}.*

On the one hand, notice that for this size of E, we have $random(T_E) = \log |G| + \log \log |\Gamma| + O(1)$. This gives $(1 + o(1)) \log |G|$ for all interesting cases ($|\Gamma| \leq |G|$). It gives the optimal $\log |G| + O(1)$ when Γ is of fixed size, which includes the important special case of linearity testing, in which $\Gamma = Z_p$ for a fixed prime p, and $G = Z_p^m$ for a large m.

On the other hand, the proof is not explicit. It uses a probabilistic argument in choosing E, which gives no clue to which graphs induce good testers. This is a major problem if one wants to use such testers in objects like PCPs. This raises a natural "derandomization" problem, of explicitly constructing good testers E (which [18] raise in their paper), or at least characterize good testers E.

This problem was answered for a special case of affine linear testing (i.e. for the property P_{aff}), by [12] who proved the following:

Theorem 1.3 *[12] Fix any $\lambda > 0$. Let S be an λ-biased set in $G = Z_p^m$, and let E denote all pairs (x, xs) for all $x \in G$ and $s \in S$. Then for every $\delta > 0$, $T_{G\times S}$ is a $\left(\delta, O\left(p^2(\delta + \lambda)\right)\right)$-test ‡ for P_{aff}.*

λ-biased sets of size poly(m/λ) can be explicitly constructed for these groups [2, 3, 14, 22, 27], which gives explicit testers T_E with near optimal randomness§ $random(T_E) \leq \log |G| + O(\log(m/\lambda))$.

†From now on all logs are taken in base 2.

‡Observe that this bound is useless unless both δ and λ are above $1/p$.

§Note that the second term is only $O(1)$ for this case in the existential result of Theorem 1.2.

[12] note that the resulting graphs E are precisely Cayley graphs over G with generating set S whose second (normalized) eigenvalue is bounded by λ. In short, Cayley expanders are good tests. This fact, as well as the fact that most graphs in Theorem 1.2 are expanders as well, one may be tempted to conjecture that *any* expander leads to a good homomorphism test for *any* group G. However [12] caution that their proof works only due to the link between the algebra of the test, and the algebraic structure of the graph, which needs be a Cayley graph over the same group $G = Z_p^m$.

Indeed, the insight that one needs a specific expander rather than an arbitrary one comes from Goldreich [16] who designed a counter example for $m = 2$. Goldreich introduced a function which is very far from any linear function from Z_p^2 to Z_p, and yet passes the test defined by the Margulis graph [25, 15] (which is a Schrier graph of an action (of some group) on $G = Z_p^2$) with high probability.

Thus the question of which graphs are good testers for general groups G (and Γ) seem more subtle. Moreover, the techniques of [12] use Fourier transforms, and seem to work only for Abelian groups. We make significant progress for characterizing good testers for general groups, which we describe next.

1.3 Our results

In brief, we show that for every domain group G, *all* expanding Cayley expanders E on the group G are good testers. Since any group G has an expanding generating set of size $O(\log|G|)$ [5], our result immediately gives a non uniform test with a near-optimal $randomness(T_E) = \log|G| + O(\log\log|G|)$. Moreover, we derandomize [5] to give a polynomial time algorithm to generate, for every group G, an expanding set of generators of size $|G|^\epsilon$, giving the randomness $(1+\epsilon)\log|G|$ explicitly and uniformly.

We note that even our non explicit result is much stronger than [18], as one can explicitly verify whether a given Cayley graph is an expander and therefore good as a test graph, while Goldreich and Sudan cannot tell which expanders are the good graphs. We note again that Goldreich gave an example showing that not every expander is good. We include this example in section 5.

Our testing result depends on two parameters - λ which is the (normalized) second largest eigenvalue, in absolute value, of the Cayley graph of G with the generating set S, and δ which is the error of the test ($\delta = error(T_{G\times S})$). We show that if S is expanding (i.e. $\lambda < 1$) then $distance(T_{G\times S}) = O(\delta)$.

Theorem 1.4 *For every G, Γ and a subset S of G, the tester $T_{G\times S}$ surely accepts any affine homomorphism $f : G \mapsto \Gamma$, and rejects with probability at least $1 - \delta$ any $f : G \mapsto \Gamma$ which is $4\delta/(1-\lambda)$ far from being an affine homomorphism, given that $\frac{12\delta}{1-\lambda} < 1$. In other words, $T_{G\times S}$ is a $(\delta, \frac{4\delta}{1-\lambda})$-test for P_{aff}.*

The proof uses a simple combinatorial argument together with the transitivity of groups. Recent analysis of (variants of) the BLR test [6, 12] use some sort of Fourier transform on abelian groups. As we deal with non-abelian groups as well, we cannot use this approach, and so rather study what may be a natural analog – the correlation of shifts of the given function with itself. It is interesting to note that the close homomorphism is defined globally, despite the fact that the tester makes only local (neighbor) tests. We also stress that our analysis avoids what seems to be an inherent problem in the Fourier approach - the relation between the Fourier coefficients and the distance to linearity is not tight, resulting in the suboptimal bounds of Theorem 1.3 of [12]. We note however that the Fourier approach has the advantage that it extends to the case where the error is relatively large, as in [12] (list decoding regime).

Our bounds are independent of the groups at hand, and thus are meaningful for constants δ and λ. As a corollary we get that the natural decoding procedure in which group elements correct their values according to the majority of their neighbors' values, converges to a homomorphism. It is interesting that this local decoding proof needs the global consistency in homomorphism testing. In contrast for derandomized "low degree" testers, [12] derive the global consistency via iterated local decoding. It is interesting if their result has a different proof which goes along the lines of this paper.

Our testers require expanding generators for the groups at hand. As mentioned, for Abelian groups such small explicit sets we known. We next provide the first nontrivial explicit construction of expanding generating sets in every group. It is fairly weak; improving it to approach the existential bound of Alon and Roichman [5] is extremely interesting.

Theorem 1.5 *For every $\epsilon > 0$ there is a polynomial time algorithm which, on input a group G, given by its multiplication table, produces a set S of size $|G|^\epsilon$ expanding generators. More precisely,*

$$\lambda(Cay(G : S)) \leq O\left(|G|^{-\epsilon/8}\right).$$

Finally, combining the two theorems we have:

Corollary 1 *For every $\epsilon > 0$ there is a polynomial time algorithm, which given any two groups G, Γ, produces a tester of randomness complexity $(1 + \epsilon)\log|G|$. This tester accepts every affine homomorphism between G and Γ with probability one, and for every $\beta > 0$, rejects every function which is β-far from any such affine homomorphism, with probability $\geq 1 - \beta/5$.*

An alternative way to view our test is that we accept with probability 1 any homomorphism and reject with high probability any function that is far from any *affine* homomorphism.

2 Preliminaries

Definition 1 (Affine homomorphism) *Let G, Γ be finite groups. A homomorphism $\phi : G \mapsto \Gamma$ is a function with the property that for every $g, h \in G$ we have that $\phi(g \cdot h) = \phi(g) \cdot \phi(h)$. We say that ϕ is an affine homomorphism if there exists an element $\gamma \in \Gamma$ such that $\gamma \cdot \phi$ is an homomorphism. Note that in this case $\phi \cdot \gamma = \gamma^{-1} \cdot (\gamma \cdot \phi) \cdot \gamma$ is also an homomorphism.*

For two functions $f_1, f_2 : G \mapsto \Gamma$ we define

$$\mathrm{dist}(f_1, f_2) = \Pr_{g \in_R G}[f_1(g) \neq f_2(g)]$$

2.1 Expander Graphs

Let $\mathcal{G} = (V, E)$ be a graph on n vertices. Let $A_{\mathcal{G}}$ be its adjacency matrix. For two sets $A, B \subset V$ denote

$$E(A, B) = \{\ (u, v) \mid u \in A \text{ and } v \in B\ \}.$$

Let $e(A, B) = |E(A, B)|$. Denote with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ the eigenvalues of $A_{\mathcal{G}}$. In case that $\mathcal{G}$ is a d-regular graph we get that $\lambda_1 = d$. Denote

$$\lambda[\mathcal{G}] = \frac{1}{d} \cdot \max(\lambda_2, |\lambda_n|).$$

We sometime use λ instead of $\lambda[\mathcal{G}]$ when $\mathcal{G}$ is clear from the context.

The next lemma due to [1] relates the edge expansion of $\mathcal{G}$ to λ.

Lemma 1 *[1][expander mixing lemma] For any two sets $A, B \subset V$ we have that*

$$\left| e(A,B) - d \cdot \frac{|A| \cdot |B|}{n} \right| \leq \lambda \cdot d \cdot \sqrt{|A| \cdot |B|}$$

In case that $A = B^c$ we get a stronger result [32, 4].

Lemma 2 *[32, 4][analog of the Cheeger constant]*

$$\frac{1-\lambda}{2} \cdot d \leq \min_{|A| \leq n/2} \frac{e(A,A^c)}{|A|} \leq 2\sqrt{1-\lambda} \cdot d$$

In particular we get

Corollary 2

- *For any A we have that*

$$\min(|A|, n-|A|) \leq \frac{2}{1-\lambda} \cdot \frac{e(A,A^c)}{d}.$$

- *If we remove $2\delta dn < \frac{1-\lambda}{6} \cdot dn$ edges from the graph, then there is a connected component of size at least $\left(1 - \frac{4\delta}{1-\lambda}\right) \cdot n$.*

Proof: The first part follows immediately from lemma 2. For the second part notice that if A and A^c are disconnected after the removal of the edges then $|e(A,A^c)| \leq 2\delta dn$. Thus if $|A| \leq \frac{n}{2}$ then by the first part we get that

$$|A| \leq \frac{2}{1-\lambda} \cdot \frac{2\delta dn}{d} = \frac{4\delta}{1-\lambda} \cdot n < n/3.$$

Therefore, after the removal of the edges, if we take the uniun of two components smaller than $\frac{n}{2}$ then the size of the uniun is smaller than $2n/3 < 1 - \frac{4\delta}{1-\lambda}$. Thus the complement of the uniun has size larger than $\frac{4\delta}{1-\lambda}$ and therefore must be of size at least $\frac{n}{2}$. It follows that the size of the uniun is smaller than $\frac{4\delta}{1-\lambda}$. By induction we get that the uniun of all components of size smaller than $\frac{n}{2}$ has size at most $\frac{4\delta}{1-\lambda}$. Hence there is a large component of size $\left(1 - \frac{4\delta}{1-\lambda}\right) \cdot n$. $\square$

Next we describe a simple dynamical process on graphs, which converges quickly in every (good enough) expander. The constants below are just some parameters which suffice for our purposes - clearly one can state a more general result along the same lines.

Definition 2 *[The infection process] Let $\mathcal{G} = (V,E)$ be a d-regular graph on n vertices. Assume that initially an adversary "infects" a subset B_0 of V of the vertices. At every subsequent time step t the infected set B_t is determined to be exactly those vertices which have at least 1/3 fraction of their neighbors in B_{t-1}. A graph is healthy if for every initial subset B_0 of size at most $n/4$, after a finite number T of steps we have $B_T = \emptyset$.*

The following is an easy consequence of the expander mixing lemma 1 above.

Corollary 3 *Assume $\lambda(\mathcal{G}) < 1/13$. Then $\mathcal{G}$ is healthy. Moreover, the convergence time T is at most $O(\log n)$.*

Proof: We will show that for every t $|B_t| \leq 0.9|B_{t-1}|$. By definition, the number of edges between B_t and B_{t-1} is lower bounded by $|E(B_t, B_{t-1})| \geq d|B_t|/3$. Applying the expander mixing lemma to these two sets gives

$$d|B_t|/3 \leq |E(B_t, B_{t-1})| \leq d|B_t| \cdot |B_{t-1}|/n + \lambda d\sqrt{|B_t| \cdot |B_{t-1}|}.$$

As $\lambda < 1/13$ and (by induction) $|B_{t-1}| \leq n/4$ we get that

$$|B_t| \leq \left(\frac{12}{13}\right)^2 |B_{t-1}| < 0.9|B_{t-1}|.$$

Iterating $O(\log n)$ times shrinks the infected set to a number smaller than 1, hence zero. □

2.2 Expanding Cayley graphs

Let G be a group. Let S be a generating set for G. That is, G is the minimal subgroup of G that contains all the elements of S. S is called symmetric if $s \in S \Leftrightarrow s^{-1} \in S$. We now define the Cayley graph of G with respect to a symmetric set of generators S.

Definition 3 *Let G be a group and S a symmetric generating set for G. We define the graph $\mathrm{Cay}(G, S)$ as follows. The vertices are the elements of G. Two elements For every $g \in G$ and $s \in S$ we put an edge (labeled s) from g to gs.*

The definition describes $Cay(G; S)$ as a directed graph, which will be one useful view of it, e.g. for describing the testers. However since S is symmetric, if there is an edge from g to h labeled s, then there is an edge from h to g labeled s^{-1}, and both can be thought of as one undirected edge. Thus $Cay(G; S)$ can also be viewed as an undirected graph (which is regular of degree $|S|$), which will be used for studying its spectral and connectivity properties.

A very nice property of Cayley graphs is their transitivity. That is, if (g_1, g_2) is an edge then so thus (gg_1, gg_2) for every g. In particular if there are k edges labelled by s between A and B then there are also k edges labelled by s between $g \cdot A$ and $g \cdot B$ for any $g \in G$.

An interesting open problem is to deterministicaly find, for a given group G, a *small* symmetric generating set S, such that $\mathrm{Cay}(G; S)$ is a good expander, in time $\mathrm{poly}(|G|)$. For $G = \mathbb{Z}_2^n$ it is easy to verify that $\mathrm{Cay}(G, S)$ is an expander with second eigenvalue λ if and only if S is an λ-biased set (see [2]). However, except for some special groups [24, 25, 26] it is not known in general how to deterministicaly find such an S. The following result of Alon and Roichman [5] guarantees that if we pick a large enough S at random then almost surely the associated Cayley graph is an expander.

Theorem 1 ([5]) *For every $\eta > 0$ there is a constant $c(\eta) > 0$ such that the following holds. Let G be a group of order n and let S be a symmetric set of $c(\eta) \log n$ random elements of G then, with probability at least $1 - \eta$,*

$$\lambda[\mathrm{Cay}(G; S)] < \eta$$

This lemma assures us that we can always find an S of size $O(\log |G|)$ such that $\mathrm{Cay}(G, S)$ is an expander.

We will show that a simple "dearndomization" of this argument leads to a deterministic construction of expanding generating sets of size $O\left(|G|^{\epsilon}\right)$ for every group G and $\epsilon > 0$. For this, the following well known estimate via the trace formula will be very useful.

Definition 4 *Fix G. For a set $S \subseteq G$ and integer m, let P_{2m} be the probability that a random word of length $2m$ in the elements of S evaluates to the identity in G.*

Proposition 1 *For every m,*

$$\lambda(Cay(G;S))^{2m} \leq nP_{2m} - 1.$$

3 Derandomized Homomorphism Testers

We first prove Theorem 1.4. Then we show that the natural local decoding procedure (namely belief propagation) converges to a homomorphism.

3.1 Proof of Theorem 1.4

Let G, Γ be groups such that $|G| = n$. Let f be a given function from G to Γ. Fix $S \subseteq G$ of size $|S| = d$ and let $\lambda = \lambda(Cay(G;S))$. Consider the test that pick a random $g \in G$ and a random $s \in S$ and accept if $f(g)f(s) = f(gs)$. Let δ be the rejection probability, i.e.

$$\delta = \Pr_{y \in G,\, s \in S}[f(y)f(s) \neq f(ys)]. \tag{1}$$

Also assume that

$$\frac{12\delta}{1-\lambda} < 1.$$

Define the function

$$\phi(x) = \text{Plurality}_{y \in G} f(xy)f(y)^{-1}.$$

We will prove that for every x almost all y agree on the value of $\phi(x)$, then prove that ϕ is a homomorphism, and finally that it is close to an affine shift of f. The first of these tasks, proved in the next claim, is perhaps the most surprising, as the test guarantees (near) local consistency, and we show it implies (near) global consistency.

Claim 1 *$\forall x \in G$ we have that*

$$\Pr_{y \in G}[f(xy)f(y)^{-1} = \phi(x)] \geq 1 - \frac{4\delta}{1-\lambda}.$$

Proof: Fix $x \in G$. Note that some constructs in this proof will depend on x, and later we will use them for all values of x.

From equation 1 we have, for random $y \in G$

$$\delta = \Pr_{y \in G,\, s \in S}[f(xy)f(s) \neq f(xys)] \tag{2}$$

We now construct a subgraph of $\text{Cay}(G, S)$. Call the edge (y, ys) bad for x if either $f(y)f(s) \neq f(ys)$ or $f(xy)f(s) \neq f(xys)$. By equation 1,2 the number of bad edges is at most $2\delta dn$. Consider the subgraph H_x obtained by removing all (undirected) edges that are bad for x. By the expansion of $Cay(G;S)$, and since we remove only $2\delta dn$ edges, we get by Corollary 2 that H_x contains a connected component C_x of size $\left(1-\frac{4\delta}{1-\lambda}\right)n$.

By connectivity and consistency of remaining edges with the test, we get that for all y in that component the value $f(xy)f(y)^{-1}$ is constant. We prove it formally as it is a bit subtle.

Claim 2 *For every two distinct elements $v,u \in C_x$ we have $f(xv)f(v)^{-1} = f(xu)f(u)^{-1}$*

Proof: Let $v = v_1, v_2, \cdots, v_t = u$ be a path between v and u in C_x. Let $s_i = v_i^{-1}v_{i+1}$ be the generator labeling the ith edge (i.e. the ith edge is $(v_i, v_i s_i)$). For every i, the existence of the edge (v_i, v_{i+1}) in H_x imply by definition the existence of the edge (xv_i, xv_{i+1}) in H_x as well. Since all these edges are good for x, it follows that

$$f(v_i)^{-1}f(v_{i+1}) = f(s_i) = f(xv_i)^{-1}f(xv_{i+1})$$

for all i. Thus

$$f(v)^{-1}f(u) = f(v_1)^{-1}f(v_t) = \prod_{i=1}^{t-1} f(v_i)^{-1}f(v_{i+1}) = \prod_{i=1}^{t-1} s_i$$

$$= \prod_{i=1}^{t-1} f(xv_i)^{-1}f(xv_{i+1}) = f(xv_1)^{-1}f(xv_t) = f(xv)^{-1}f(xu)$$

By changing sides we get that $f(xv)f(v)^{-1} = f(xu)f(u)^{-1}$ as required. □

Thus, $f(xy)f(y)^{-1}$ is the same for all $y \in C_x$. As $|C_x| > |G|/2$, and we have defined ϕ using plurality over y, we get $\phi(x) = f(xy)f(y)^{-1}$ for every $y \in C_x$. Since $|C_x| \geq \left(1-\frac{4\delta}{1-\lambda}\right)n$, claim 1 follows. □

Claim 3 *ϕ is a homomorphism.*

Proof: We need to show that for every $x,y \in G$ we have that $\phi(x)\phi(y) = \phi(xy)$. Consider (like [9]) arbitrary $x,y \in G$ and the probability over $h \in G$

$$\Pr_{h\in G}[\phi(x)\phi(y) = \phi(xy)] \tag{3}$$

which is independent of h, and thus is either 0 or 1. We prove that this probability is positive and therefore 1. We lower bound it by the probability of the intersection of three events over the same random variable h chosen uniformly in G

$$\Pr_{h\in G}[\phi(x)\phi(y) = \phi(xy)] \geq \Pr_{h\in G}\left[\begin{array}{c} \phi(x) = f(xh)f(h)^{-1} \quad \text{and} \\ \phi(y) = f(h)f(y^{-1}h)^{-1} \quad \text{and} \\ \phi(xy) = f(xh)f(y^{-1}h)^{-1} \end{array}\right].$$

Now

- By claim 1 $\Pr_{h\in G}[\phi(x) = f(xh)f(h)^{-1}] \geq 1 - \frac{4\delta}{1-\lambda}$.

- Notice that

$$\Pr_{h\in G}[\phi(y) = f(h)f(y^{-1}h)^{-1}] = \Pr_{h\in G}[\phi(y) = f(y\cdot(y^{-1}h))f(y^{-1}h)^{-1}].$$

  As h is random this probability equals

$$\Pr_{h'\in G}[\phi(y) = f(yh')f(h')^{-1}]$$

  which by claim 1 is at least $1 - \frac{4\delta}{1-\lambda}$.

- Similarly we get that

$$\Pr_{h\in G}[\phi(xy) = f(xh)f(y^{-1}h)^{-1}] = \Pr_{h\in G}[\phi(xy) = f((xy)\cdot(y^{-1}h))f(y^{-1}h)^{-1}] \geq 1 - \frac{4\delta}{1-\lambda}.$$

Note that each of these events have probability at least $1 - \frac{4\delta}{1-\lambda}$, and so the probability of their intersection is at least $1 - \frac{12\delta}{1-\lambda}$ which is strictly positive, and so must be 1. □

Finally we show that f is close to some affine shift of ϕ.

Claim 4 *There exist $\gamma \in \Gamma$ such that*

$$\Pr_{x\in G}[\phi(x) = f(x)\cdot\gamma] \geq 1 - \frac{4\delta}{1-\lambda}.$$

Proof: For every $x \in G$ denote with G_x the set of ("good") y's satisfying $\phi(x) = f(xy)f(y)^{-1}$ (note that the set G_x contains the set H_x of Claim 1, but may in fact be even larger). By that claim, for every x $|G_x| \geq (1 - \frac{4\delta}{1-\lambda})|G|$. It follows by averaging that there exist $y \in G$ such that

$$|\{\ x\ :\ y \in G_x\ \}| \geq \left(1 - \frac{4\delta}{1-\lambda}\right)|G|.$$

For this y we have that

$$\Pr_{x\in G}[\phi(x) = f(xy)f(y)^{-1}] \geq 1 - \frac{4\delta}{1-\lambda}.$$

Therefore

$$\Pr_{x'\in G}[\phi(x'y^{-1}) = f(x')f(y)^{-1}] \geq 1 - \frac{4\delta}{1-\lambda}.$$

As ϕ is a homomorphism we get that

$$\Pr_{x'\in G}[\phi(x') = f(x')f(y)^{-1}\phi(y)] \geq 1 - \frac{4\delta}{1-\lambda}.$$

The claim follows by defining $\gamma = f(y)^{-1}\phi(y)$. □

This completes the proof of Theorem 1.4.

3.2 Iterated local majority decoding

Recall that the close homomorphism in the proof above was defined according to a *global* majority: every group element x chose the value $\phi(x)$ according to the plurality of $f(xy)f(y)^{-1}$ over *all* group elements $y \in G$. We show that iterated *local* majority decoding, where (in each phase) every group element x updates its value according to the plurality of $f(xs)f(s^{-1})$ over its neighbors in the Cayley graph, converges to (the same) global homomorphism ϕ.

Definition 5 *[Iterated majority decoding] Let G, Γ be groups, S a subset of G and $f : G \to \Gamma$ any function. Set $f = f_0$ and for every integer t define f_t by*

$$f_t(x) = \mathrm{Plurality}_{s \in S} f_{t-1}(xs) f_0(s^{-1}).^{\dagger}$$

Theorem 3.1 *Let $G, \Gamma, S, \lambda, \delta, \gamma$ be as in Theorem 1.4 and further assume that $\lambda, \delta \leq 1/17$. Let $f : G \to \Gamma$ be such that the tester $Cay(G; S)$ accepts f with probability at least $1-\delta$. Then the iterated decoding procedure above converges to a homomorphism $\phi : G \to \Gamma$ in $O(\log |G|)$ steps. Moreover, ϕ is the homomorphism defined in the proof of Theorem 1.4, and thus is at most $4\delta/(1-\lambda)$-far from $\gamma \cdot f$.*

Proof: Let ϕ be the homomorphism guaranteed by Theorem 1.4. Let f_t be the sequence of functions defined by the iterated majority decoding procedure above, and let D_t denote the set of group elements on which $\gamma \cdot f_t$ and ϕ disagree. By our choice of parameters,

$$|D_0| \leq \frac{4\delta}{1-\lambda}|G| < |G|/4.$$

We will reduce the analysis of the local decoding to that of the infection process in 3 at the end of section 2.1.

Set $B_0 = D_0$ and apply the infection process to it, to obtain a sequence B_t. We show by induction that for every t, we have $D_t \subseteq B_t$ so the theorem follows from Corollary 3. Assume for the moment that for all but 1/6 fraction of the $s \in S$ we have $f(s) = \phi(s)$. Then every element x in step t which has a 2/3 of its neighbors in the complement of B_{t-1}, gets the same value from 1/2 of them (as $D_{t-1} \subset B_{t-1}$), and this value agrees with ϕ, namely $\gamma \cdot f_t(x) = \phi(x)$.

We now argue that for at most 1/6 fraction of $s \in S$ it must be the case that $f(s) = \phi(s)$. Fix any "bad" s for which $f(s) \neq \phi(s)$. Since $|D_0| < |G|/4$, for at least 1/2 of all the elements $x \in G$ we have both $\gamma \cdot f(x) = \phi(x)$ and $\gamma \cdot f(xs) = \phi(xs)$. All these pairs x, s are rejected by the tester, and since it rejects only a δ fraction of all such pairs, the number of bad s is at most $2\delta < 1/6$.

□

4 Explicit Expanding Generators - Proof of Theorem 1.5

In this section we give a polynomial time algorithm to find a relatively small expanding generating set in every group. We state the main technical result, which is nearly identical to Theorem 1 of Alon and Roichman, except adding the condition that the choices of the generators need not be fully

†Note that we keep using the *initial* values on S in all iterations

independent. The proof remains identical to their proof, only we'll need it with different parameters. We give the proof for completeness.

For the rest of the section we fix a group G of size n.

Theorem 4.1 *Fix any integer $m \geq 2$. Consider the following distribution on Cayley graphs on G. Draw d $2m$-wise independent samples $g_1, \cdots, g_d$ from G to form a (multi)set T, and let $S = T \cup T^{-1}$. Then the expectation of $\lambda(\mathrm{Cay}(G;S))$ is*

$$E[\lambda(Cay(G;S))] < (2n)^{1/2m}(16m/d)^{1/4}.$$

Proof: We repeat the essentials of the proof of [5], with the only difference being the limited independence of the generators. This turns out not to change the analysis. We skip easy proofs which can be obtained from their paper.

By Proposition 1 and Jensen's inequality we get that

$$E[\lambda(Cay(G;S))] < (nE[P_{2m}] - 1)^{1/(2m)}.$$

Thus, it suffices to prove that $E[P_{2m}] \leq 1/n + 2(16m/d)^{m/2}$. In order to bound P_{2m} we construct a random word of length $2m$ in three steps.

- Pick a random word W' of length $2m$ in the alphabet $\{a_1, a_1^{-1}, \cdots, a_d, a_d^{-1}\}$
- Reduce the word over the free group on d generators to obtain the word W.
- Replace every a_i by the associated random g_i from T.

The upper bound on the expectation of P_{2m} will follow from the three probability estimates below.

Claim 5

$$\Pr[|W| < m] \leq (32/d)^{m/2}$$

Claim 6 *Call W bad if none of the d letters*[†] *appears exactly once in W. Condition on $|W| \geq m$. Then*

$$\Pr[W\ bad] \leq (16m/d)^{m/2}$$

Claim 7 *Fix any good w, and replace each a_i by g_i as above to generate the word $w(T)$ in G. Then*

$$\Pr[w(T) = 1_G] = 1/n$$

We prove only the last claim, since this is the only point where the limited independence of T could make a difference. The first two claims follow from [5] after an adjustment of the parameters.

Proof: Since w is good, there is some generator, say a_1 w.l.o.g., which occurs exactly once in w. There are at most $2m - 1$ other generators a_i in w. For each of these, expose their g_i value. Now the probability in question is the probability that g_1 equals a fixed group element determined by the exposed g_i's and w. But g_1 is completely uniform given these choices, and so that probability is precisely $1/n$. □

[†] a_i and a_i^{-1} are considered the same letter

The proof now follows as the expectation of P_{2m} is bounded by the sum of the probabilities of the events in the claims above. This concludes the proof of Theorem 4.1 $\square$

Corollary 4 *Take $d = n^{4/m}$. Then*

$$E[\lambda(Cay(G; S))] < 3m^{1/4}d^{-1/8}.$$

Finally we show how to choose a set of generators deterministically, establishing Theorem 1.5. Given $\epsilon > 0$, we set $m = 4/\epsilon$, and $d = n^{4/m} = n^{\epsilon}$. Construct a sample space of size at most $(2n)^{2m}$ of d-tuples over G which are $(2m)$-wise independent. This takes polynomial time in n (see below). For each such tuple T compute (again in polynomial time, as all we need is a reasonable approximation) the associated $\lambda(\mathrm{Cay}(G; S))$, and returns the set S for which this eigenvalue is smallest.

Here is one way to construct such a sample space efficiently. Pick a prime p in $[n, 2n]$, and identify the elements of G with some n distinct field elements $R \subseteq F_p$. For every polynomial in $F_p[x]$ of degree $< 2m$ obtain a d-tuple by evaluating it on (say) the first d points in F_p. Now remove all tuples containing an element outside R.

This concludes the proof of Theorem 1.5.

5 Not Every Expander is Good

In this section we present a construction, due to Oded Goldreich, of an expander graph on a group (but not a Cayley graph), for which the natural tester fails miserably.

Let p be a prime, and consider the (Schreier) graph H_p describing the action of the group $SL_2(p)$ on the vector space Z_p^2, with generators S being the two matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and their inverses.

More concretely, the vertices of H_p are $(x, y) \in Z_p \times Z_p$, and the four neighbors of (x, y) are $(x, y \pm x)$ and $(x \pm y, y)$. Note that H_p has two connected components - the vertex $(0, 0)$ and the rest. Thus H_p has two eigenvalues of value 1, and we denote here by $\lambda(H_p)$ the maximum absolute value of any of the other eigenvalues.

The graph H_p is a variant of the famous Margulis graph - the first explicit expander. The expansion of (the large component of) H_p follows directly from the expansion of the Cayley graph $Cay(SL_2(p); S)$, which follows from Selberg's celebrated 3/16 Theorem (see Lubotzky [23] for details).

Proposition 2 *[23] For every p, $\lambda(H_p) \leq 13/16$.*

We will consider functions from Z_p^2 to Z_p. Since the groups are Abelian, we will write them additively.

For defining the tester (and the function), it will be convenient to view each undirected edge as directed "positively". In other words every vertex $v = (x, y)$ has two directed edges emanating from it: to $u = (x, y + x)$ (labeled by $u - v = (0, x)$) and to $w = (x + y, y)$, (labeled by $w - v = (y, 0)$).

Observe that in our graph labels of edges always have 0 in one of their components. Also note that there are roughly $2p$ distinct labels, despite the graph having degree 4 - this is very different from a Cayley graph (in which the number of labels is the degree).

In this notation, the tester associated to this graph, picks uniformly at random a (directed edge) from v to u and tests if $f(u) - f(v) = f(u - v)$.

We now present the example which beats this tester. It will be very far from linear (i.e. a homomorphism), but will pass the test with probability close to 1.

Consider the function $f : Z_p \times Z_p \to Z_p$ defined as follows. $f(x, y) = x^2$ if $y = 0$, $f(x, y) = y^2$ if $x = 0$, and $f(x, y) = x \cdot y$ otherwise (with all arithmetic in Z_p).

Theorem 5.1

- *The function f is $(1 - 3/p)$-far from any affine function.*
- *The function f passes the test with probability $1 - 3/p$.*

Proof:

First we prove the 1st item in the Theorem.

Every affine function g from Z_p^2 to Z_p looks like $g(x) = ax + by + c$ for some constants $a, b, c \in Z_p$. Consider only pairs $x, y \neq 0$, as there are only a fraction $2/p$ of the pairs (x, y) which are not. We want to count the number of possible solutions to the equation $xy = ax + by + c$. For every possible value of d of x we get a (different) linear equation in y, which has at most one solution. So for every possible affine homomorphism g we have $dist(f, g) \geq 1 - 3/p$, as required.

Now we prove the 2nd item in the Theorem.

Only $6p$ of the $2p^2$ directed edges have a 0 component in either of their endpoints. Thus with probability at least $1 - 3/p$ the chosen neighboring vertices v, u have no zero component. We show that all these edges pass the test.

Let $v = (x, y)$. There are 2 similar cases. First take $u = (x, y + x)$. Then

$$f(u) - f(v) = x(y + x) - xy = x^2 = f((u - v)).$$

Now take $w = (x + y, y)$. Then

$$f(w) - f(v) = (x + y)y - xy = y^2 = f((u - v)).$$

□

Acknowledgements

We thank Eli Ben-Sasson, Oded Goldreich and Salil Vadhan for many illuminating discussions and for reading and commenting earlier versions of the manuscript. We thank Oded also for kindly allowing us to include his counterexample here.

References

[1] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks, *Discrete Mathematics*, pages 15–19, Vol. 72, 1988.

[2] N. Alon and O. Goldreich and J. Hastad and R. Peralta. Simple Constructions of Almost $k-$wise Independent Random Variables, *Journal of Random Structures and Algorithms*,3:3 (1992), pp 289–304

[3] N. Alon, Y. Mansour. ϵ-Discrepancy sets and their applications for interpolation of sparse polynomials. *Information Processing Letters*, 54:337-342 (1995).

[4] N. Alon and V. D. Milman. Eigenvalues, Expanders and Superconcentrators (Extended Abstract). *In 25th Annual Symposium on Foundations of Computer Science*, pages 320-322, Singer Island, Florida, 24-26 October 1984.

[5] N. Alon, Y. Roichman. Random Cayley Graphs and Expanders. *Rand. Str. Alg.* vol. 5 (1994), 271–284

[6] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, M. Sudan. Linearity testing over characteristic two, *IEEE Transactions on Information Theory* vol. 42(6), pp 1781–1795, November 1996.

[7] M. Bellare, S. Goldwasser, C. Lund, A. Russell. Efficient probabilistic checkable proofs and applications to approximation. In *Proceedings of the Twenty-Fifth ACM Symposium on the Theory of Computing (STOC)*, pages 294–304, San Diego, California, 16-18 May 1993.

[8] M. Bellare, M. Sudan. Improved non-approximability results. *Proceedings of the Twenty Sixth Annual ACM Symposium on Theory of Computing*, pages 184-193, Montreal, Quebec, Canada, 23-25 May 1994.

[9] M. Ben-Or, D. Coppersmith and R. Rubinfeld. Non-abelian homomorphism Testing. Manuscript, December 2003.

[10] M. Bellare, O. Goldreich, M. Sudan. Free bits, PCP and non-approximability - towards tight results. In *SIAM Journal on Computing*, 27(3): 804-915, June 1998. Preliminary version in *Proceedings of the 36th FOCS*, pages 422-431, Milwaukee, Wisconsin, 23-25 October 1995.

[11] M. Blum, M. Luby and R. Rubinfeld. Self-Testing/Correcting with Applications to Numerical Problems. *JCSS*, Vol. 47, No. 3, pages 549–595, 1993.

[12] E. Ben Sasson and M. Sudan and S, Vadhan and A. Wigderson. Randomness-efficient Low degree tests and short PCPs via Epsilon-Biased Sets, *In Proceedings of the 35th STOC*, pages 612-621, 2003.

[13] F. R. K. Chung. Diameters and eigenvalues, *Journal of the AMS*, pages 187–196, Vol. 2(2), 1989.

[14] G. Even, O. Goldreich, M. Luby, N. Nisan, B. Velickovic Approximations of General Independent Distributions, *STOC'92*, pages 10-16, 1992.

[15] O. Gabber and Z. Galil. Explicit constructions of linear size superconcentrators. *In proceedings of the 20th Annual Symposium Foundations of Computing Science*, IEEE, New York, pages 364–370, 1979.

[16] O. Goldreich. Private Communication, June 2002.

[17] O. Goldreich. Combinatorial Property Testing. *DIMACS series in Discrete Mathematics and Theoretical Computer Science*, pages 45–59, Vol. 43, 1998.

[18] O. Goldreich, M. Sudan. Locally Testable Codes and PCPs of Almost-Linear Length *Electronic Colloquium on Computational Complexity*, Report TR02-050.

[19] J. Håstad. Some Optimal Inapproximability Results. *Journal of the ACM*, pages 798–859, volume 48, 2001.

[20] J. Håstad, A. Wigderson. Simple Analysis of Graph Tests for Linearity and PCP. To appear in *Random Structures and Algorithms*.

[21] R. Impagliazzo and A. Wigderson. P=BPP unless E has subexponential circuits: derandomizing the XOR lemma, *In Proceedings of the 29th STOC*, pages 220–229, 1997.

[22] N. M. Katz. An Estimate for Character Sums. *J. AMS* 2, (1963) pages 197-200.

[23] A. Lubotzky. Discrete Groups, Expanding Graphs and Invariant Measures. Progress in Math. 125, Birkhaüser Verlag, Basel 1994.

[24] A. Lubotzky, R. Phillips, P. Sarnak. Ramanujan graphs, *Combinatorica*, pages 261–277, Vol. 8, 1988.

[25] G. A. Margulis. Explicit constructions of concentrators. *Problemy Peredachi Informatsii*, pages 71–80, 1973.

[26] M. Morgenstern. Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *Journal of Combinatorial Theory B*, 62(1):44-62, 1994.

[27] J. Naor, M. Naor. Small Bias Probability Spaces: Efficient Constructions and Applications. *22nd STOC* 1990, pages 213-223.

[28] D. Ron. Property testing (a tutorial). *Handbook of Randomized Computing*, (S. Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolim editors), Kluwer Press (2001).

[29] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, Vol. 25 (2), pages 252–271, 1996.

[30] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. *In Proceedings of the 32nd STOC*, pages 191–199, 2000.

[31] M. Sudan and L. Trevisan and S. Vadhan. Pseudorandom generators without the XOR Lemma. *Journal of Computer and System Sciences*, 62(2): 236–266, March 2001.

[32] R.M. Tanner. Explicit concentrators from generalized N-gons, *SIAM Journal on on Algebraic Discrete Methods*, pages 287–293, Vol. 5(3), 1984.