A Randomness-Efficient Sampler for Matrix-valued Functions and Applications

Avi Wigderson
Institute for Advanced Study
Princeton, NJ 08540
avi@ias.edu

David Xiao
Princeton University
Princeton, NJ 08544
dxiao@cs.princeton.edu

Abstract

In this paper we give a randomness efficient sampler for matrix-valued functions. Specifically, we show that a random walk on an expander approximates the recent Chernoff-like bound for matrix-valued functions of Ahlswede and Winter [1], in a manner which depends optimally on the spectral gap. The proof uses perturbation theory, and is a generalization of Gillman's and Lezaud's analysis of the Ajtai-Komlos-Szemeredi sampler for real-valued functions [11, 21, 2].

Derandomizing our sampler gives a few applications, yielding deterministic polynomial time algorithms for problems in which derandomizing independent sampling gives only quasi-polynomial time deterministic algorithms. The first (which was our original motivation) is to a polynomial-time derandomization of the Alon-Roichman theorem [4, 20, 22]: given a group of size n, find $O(\log n)$ elements which generate it as an expander. This implies a second application - efficiently constructing a randomness-optimal homomorphism tester, significantly improving the previous result of Shpilka and Wigderson [29]. The third is to a "non-commutative" hypergraph covering problem - a natural extension of the set-cover problem which arises in quantum information theory (e.g. [1, 16]), in which we effciently attain the integrality gap when the fractional semi-definite relaxation cost is constant.

1. Introduction

1.1. Background

The Chernoff bound [8] and its variants are among the most useful mathematical results, and in particular are extremely useful in theoretical computer science. Roughly stated, it says that if we wish to estimate the mean of a bounded real function on some domain V, the average of the values at k independent samples deviates from the true mean (by a small additive constant) only with error probability bounded by $2^{-\Omega(k)}$. Note that if every sample requires r random bits, this sampling procedure requires a total of rk random bits to achieve error $2^{-\Omega(k)}$.

A remarkable construction and analysis of Ajtai, Komlos and Szemeredi [2] suggested a way of achieving essentially the same error using only $r + O(k)$ bits. The idea is to impose a good constant degree *expander* graph G on the vertex set V, and select k (highly dependent) samples by taking a random path of length k in this graph. The analysis of this sampler due to Gillman [11], which is the first to consider sampling any bounded real function (see also [18, 21]), shows that the error is bounded by $2^{-\Omega(\varepsilon k)}$, where ε is the *spectral gap* of the random walk on the expander G. The fact that explicit families of constant degree expanders with constant spectral gap are known [10, 24, 23, 27] show that such a randomness-efficient sampler can be efficiently implemented.

This sampler has become a paramount tool in theoretical computer science. Indeed, it has found a large number of applications in such a variety of areas as deterministic amplification [9, 17], security amplification in cryptography [14], hardness of approximation [5, 3], extractor construction (e.g. see surveys [26, 13, 28]), construction of efficient error-correcting codes [30, 7], construction of ε-biased spaces [25] and much more. In algorithmic applications, including some of the ones above, often both r and k are $O(\log n)$ where $n = |V|$ is the input size of the problem, so derandomizing simply (i.e. enumerating all possible values of the random bits) the independent sampling requires quasi-polynomial time, while the AKS-sampler can be derandomized in polynomial time.

Recently, a Chernoff-like bound was introduced by Ahlswede and Winter [1] for matrix-valued random variables. Here we seek to estimate the average of a function from V to $d \times d$ complex Hermitian[1] matrices of bounded norm. The [1] generalization of the Chernoff bound states that the average of k independent points deviates significantly in norm from the mean with probability bounded by

[1] For all practical purposes the reader can think of real symmetric matrices.

$d2^{-\Omega(k)}$.

Like the Chernoff bound, this generalization has quickly found applications. Many of them are in quantum information theory (and private quantum channels) [1, 16], where such matrices arise naturally. A notably different one is to a new proof [20, 22] of the Alon-Roichman theorem [4], showing that for every finite group of size n, choosing $O(\log n)$ random generators gives an expanding Cayley graph with high probability.

1.2. Our results

In this paper we show that the AKS-sampler works as well as independent sampling even for matrix valued functions. If one samples k points on a walk of an expander of spectral gap ε, the error probability is bounded by $d2^{-\Omega(\varepsilon k)}$, "derandomizing" [1] in complete analogy to the way [2, 11] derandomized Chernoff in the real (1-dimensional) case.

Let $G = (V, E)$ be an expander graph with spectral gap ε. Define Y_i ($0 \leq i \leq k$) to be the i'th vertex visited in a random walk on G that starts from Y_0 which is uniformly distributed in V. Let $W = (Y_1, \ldots, Y_k)$ be the random variable representing the sequence of vertices encountered on a random walk.

Let f be any function on V taking values in $d \times d$ Hermitian matrices such that the matrix norm $\|f(v)\| \leq 1$ for all $v \in V$, and let $\mathbb{E}[f]$ be the mean value of f uniformly over all vertices. Define $f(W) = \sum_{i=1}^{k} f(Y_i)$ to be the value of the random walk.

Our main theorem states the following.[2]

Theorem 1.1. *For every $1 \geq \gamma > 0$ and every $k \geq \frac{4}{\gamma}$ we have*

$$\Pr[\, \|\tfrac{1}{k} f(W) - \mathbb{E}[f]\| > \gamma] \leq d2^{-\Omega(\gamma^2 \varepsilon k)}$$

The dependence on d is linear, just as in the independent case of [1].

Note that for $\varepsilon = 1$ (i.e. a complete graph) this bound is just independent sampling and thus the Chernoff bound of [1] (we state this in Theorem 2.15). For $d = 1$ it is just the 1-dimensional AKS sampler of [11, 2, 21, 18]. For $\varepsilon = d = 1$ it is just the classical Chernoff bound. Thus our work essentially generalizes all of these (up to constant factors in the exponent).

Our proof uses perturbation theory, generalizing the proofs of [11, 21]. We also have a simpler analysis using basic linear algebra of a slightly weaker bound[3] where the dependence on ε in the exponent is close to cubic instead of linear. Unfortunately we omit the proof here for space concerns.

A simple extension of the theorem above gives rise to a randomness-efficient sampler for weighted averages of matrix-valued functions, which is useful for some of our applications.

1.3. Applications

Our main application is a complete derandomization of the Alon-Roichman theorem (which was our motivation to begin with). [4] showed that given any group H if we choose $S \subseteq H$ of size $O(\log |H|)$ at random then with high probability the induced Cayley graph is a good expander. We note that derandomizing independent sampling gave only a quasi-polynomial algorithm, and that the best previous polynomial time algorithm [29] could only produce $|H|^{\Omega(1)}$ expanding generators. Our algorithm finds $O(\log |H|)$ expanding generators deterministically in polynomial time.

Theorem 1.2. *Fix $\beta < 1$. Given an arbitrary finite group H (specified by its multiplication table), one can find in time $|H|^{O(1)}$ a symmetric generating multi-set S of size $O(\frac{1}{\beta^2} \log |H|)$ such that $\lambda_2(X(H; S)) < \beta$.*

This will immediately imply the following optimal solution to a problem of [29] (see also [15]), significantly improving their results. More details appear in Section 4.2.

Corollary 1.3. *Given an arbitrary group H, one can construct in time $|H|^{O(1)}$ a* homomorphism tester *for functions on H which uses only $\log |H| + \log \log |H| + O(1)$ random bits.*

We also derandomize a natural problem arising in [1] concerning quantum hypergraphs. Unfortunately for lack of space we omit this application in the proceedings.

1.4. Organization of the paper

The remainder of the paper is organized as follows. In Section 2 we define the background material needed to prove our main theorem. In Section 3 we prove the main technical result, Theorem 1.1. In Section 4 we derive some applications of this sampler.

2. Preliminaries

2.1. Expander graphs

Given a connected undirected d-regular graph $G = (V, E)$ on n vertices, we define its normalized adjacency

[2] One may ask why the main theorem is interesting, as we could use a union bound to independently bound the entries of the matrices. However this loses a factor of d in the bound of the eigenvalues, which is insufficient for our purposes. Other naive approaches are similarly insufficient in our setting.

[3] When using an expander for sampling, ε is a constant and this bound simply has a different constant in the exponent.

matrix A, $A_{ij} = e_{ij}/d$ where e_{ij} is the number of edges between vertices i and j (we allow self-loops and multiple edges). It is easy to see that A is real and symmetric, hence Hermitian.

It is well-known that the set of eigenvalues (called the *spectrum*) of A is of the form $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$. The spectrum of G is the spectrum of A. Note that 1 is an eigenvalue of multiplicity 1. We will frequently refer to the unit eigenvector of eigenvalue 1 as $u = [1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$,[4] where ^{T} denotes the matrix transpose of a matrix (or vector). The *spectral gap* of A is defined as $1 - \lambda_2$. A family of graphs $\{G_i\}_{i\geq 1}$ is said to be an *expander family* if the spectral gap of each G_i is strictly greater than some fixed $\varepsilon > 0$. Recall that explicit such families with constant degree exist: we can construct arbitrarily large graphs with fixed degree such that given a node in the graph we can compute its neighbors in time $\mathrm{poly}\log$ in the size of the graph. An explicit example is the following.

Theorem 2.1 ([23, 24]). *Fix any prime p such that $p \equiv 1 \pmod 4$. Then for all primes q such that $q \equiv 1 \pmod 4$, one can efficiently construct a graph of size $q+1$ and degree $p+1$ with second-largest eigenvalue at most $2\sqrt{p}/(p+1)$.*

Cayley graphs are graphs defined on groups:

Definition 2.2. Let H be a finite group and let T be a multiset with elements in H. Let $S = T \sqcup T^{-1}$ denote the multiset containing all elements T and their inverses with appropriate multiplicity. Then we can define the Cayley graph $X(H; S) = (V, E)$ where $V = H$ and $\{h, hs\} \in E$ for all $h \in H, s \in S$, again with appropriate multiplicities.

We will also use matrix tensor products, which give us a simple language to work with block matrices. Recall that if A is a $n \times m$ matrix and B is a $p \times q$ matrix, then $A \otimes B$, the *matrix tensor product*, is the $np \times mq$ matrix given by

$$(A \otimes B)_{(i,k),(j,\ell)} = A_{i,j} \cdot B_{k,\ell}$$

The following facts about the matrix tensor product are well-known:

2.2. Perturbation Theory

The proof of Lemma 3.4, the heart of our proof of the main theorem, relies on many facts from perturbation theory. We state some of the results that we will require. We use [6] (see also [19]) as our guide. We will not state the theorems in full generality for simplicity's sake.

An *analytic perturbation* (of a matrix A_0) is a matrix-valued power series $A(t) = \sum_{i=0}^{\infty} t^i A_i$ in the variable t with matrix coefficients $(A_i)_{i\geq 0}$. Note that $A(0) = A_0$. We will only be concerned here with the case that A_0 is Hermitian and all coefficients A_i have norm at most 1.

Perturbation theory studies various matrix parameters of $A(t)$ (such as eigenvalues, eigenspaces etc.) as a function of t. More specifically, we'd like them to be convergent power series in t for some radius around $t = 0$, and perturbation theory tells us how these power series behave, as well as the dependence of the convergence radius on the coefficients of the perturbation $A(t)$.

Then [6] states that an eigenvalue λ of A_0 of multiplicity m may split into as many as m distinct eigenvalues $\lambda^{(1)}(t), \ldots, \lambda^{(m)}(t)$ upon perturbation [6, Ch. 3.2], where the $\lambda^{(i)}(t)$ are continuous at $t = 0$ and furthermore $\lambda = \lambda^{(i)}(0)$ for all $1 \leq i \leq m$.

The "stability" of the perturbation of λ primarily depends on the separation of λ from the other eigenvalues of A_0. (again, we assume that all A_i have norm ≤ 1, otherwise this stability depends on these norms as well). The radius of convergence also depends on this separation, which we define below.

Definition 2.3. We call

$$\varepsilon = \min_{\lambda' \in \mathrm{Spec}(A_0),\ \lambda' \neq \lambda} |\lambda - \lambda'|$$

the *separation* of λ from the other eigenvalues of A_0.[5]

We will work with the projection onto the eigenspace of all the eigenvalues splitting from λ.

Theorem 2.4 ([6, pp. 116-117, p. 326]). *Consider a perturbation $A(t)$. Let λ be an eigenvalue of multiplicity m of the unperturbed operator $A(0) = A_0$. Consider the space $\Lambda(t)$ spanned by the eigenvectors of the eigenvalues $\lambda^{(1)}(t), \ldots, \lambda^{(m)}(t)$ splitting from λ. $\Lambda(t)$ is a space of dimension m. For each t there is an operator $P(t)$ that projects onto $\Lambda(t)$, and for all $t \leq \varepsilon/3$ the function $P(t)$ is analytic in t: there exist matrices P_i (themselves not necessarily projections) such that*

$$P(t) = \sum_{i=0}^{\infty} t^i P_i \tag{2.1}$$

projects onto $\Lambda(t)$. Here, $P(0) = P_0$ is the projection onto the eigenspace of eigenvalue λ of A_0.

We will also need a few additional facts from perturbation theory.

Lemma 2.5 ([6, p. 115]). *Let ε be the separation of λ from the other eigenvalues of A_0. Suppose additionally that $\|A_i\| \leq \frac{1}{2^{i-1}}$ for all $i \geq 1$. Then for all $t \leq \varepsilon/3$, the*

[4] This is the uniform distribution on V, normalized to have $\|u\| = 1$.

[5] Notice that the spectral gap of a graph is exactly the separation of the eigenvalue 1 of the normalized adjacency matrix of the graph from the other eigenvalues.

eigenvalues of $A(t)$ in the range $[\lambda - \varepsilon/2, \lambda + \varepsilon/2]$ all split from λ (i.e. they do not split from some other eigenvalue of A_0).

Proof. Lemma 3 of [6, p. 115] tells us that we only need to verify that

$$\sum_{i=1}^{\infty} t^i \|A_i\| < \varepsilon/2$$

for all $t \leq \varepsilon/3$. This is easily done by calculation using the fact that $\|A_i\| \leq 1/2^{i-1}$ for all $i \geq 1$. ∎

Definition 2.6 ([6, pp. 74-75]). The *reduced resolvent* S_0 of a matrix A_0 with respect to the eigenvalue λ is the pseudo-inverse of $\lambda I - A_0$. That is, its restriction on the eigenspace of the eigenvalue λ of A_0 is 0 and its restriction on the orthogonal complement is $(\lambda I - A_0)^{-1}$.

Lemma 2.7. *$\|S_0\| = \frac{1}{\varepsilon}$ where ε is the separation of λ from the other eigenvalues of A_0.*

Proof. Let the eigenvalues of A_0 be $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Since S_0 is the pseudo-inverse of $\lambda I - A_0$, the eigenvalues of S_0 are 0 and $\frac{1}{\lambda_i - \lambda}$ for all $\lambda_i \neq \lambda$. It is easy to see that S_0 is Hermitian, so it follows that $\|S_0\|$ equals its eigenvalue largest in absolute value, which is exactly $\frac{1}{\varepsilon}$. ∎

The definition of the reduced resolvent is applied in the following identity.

Theorem 2.8 ([6, p. 156]). *If $A(t), P(t)$ are defined as above, then*

$$(A(t) - I)P(t) = \sum_{i=1}^{\infty} t^i Z^{(i)} \tag{2.2}$$

where

$$Z^{(i)} = -\sum_{k=1}^{i} \sum_{\substack{\mu_1+\ldots+\mu_k=i \\ \sigma_1+\ldots+\sigma_{k+1}=k-1 \\ \mu_j \geq 1, \sigma_j \geq 0}} S_0^{(\sigma_1)} A_{\mu_1} S_0^{(\sigma_2)} \ldots A_{\mu_k} S_0^{(\sigma_{k+1})}$$

The $S_0^{(\sigma)}$ is shorthand, where $S_0^{(0)}$ is the projection $P(0) = P_0$, and for $\sigma \geq 1$ we define $S_0^{(\sigma)} = -(-S_0)^{\sigma}$, where S_0 is the reduced resolvent of A with respect to the eigenvalue λ. This series is convergent for $t \leq \varepsilon/3$.

Remark 2.9. For a full discussion of this expression see [6]. The curious reader will note that in our statement there is no constant term in the series $\sum_{i=1}^{\infty} t^i Z^{(i)}$. This is because $A(t)$ is diagonalizable and so the constant term[6] is zero. He or she will also note that the summation in the definition of $Z^{(i)}$ is over $\sigma_j \geq 0$, which is different from the statement in [6]. This also follows from the fact that $A(t)$ is diagonalizable.

[6] This is the *eigennilpotent* of $A(t)$.

2.3. Probability theory of matrix-valued random variables

We will write I to be the identity, or I_d when the dimension d is not clear.

Theorem 1.1 is stated in terms of the *matrix 2-norm*, which is defined for any $d \times d$ matrix A as $\|A\| = \max_{x \in \mathbb{C}^d} \|Ax\|/\|x\|$. If A is Hermitian then $\|A\| = |\lambda_{\max}|$, where $\lambda_{\max}$ is the eigenvalue of A with largest absolute value.

To prove Theorem 1.1, we will work with a different though related partial ordering of Hermitian matrices. A Hermitian matrix is *positive semi-definite* (p.s.d.) if all its eigenvalues (which are real) are non-negative. Note that non-negative linear combinations of p.s.d. Hermitian matrices are also p.s.d. Hermitian, i.e. Hermitian p.s.d. matrices form a real cone. We define that $A \geq 0$ if A is p.s.d., and $A \geq B$ if $A - B \geq 0$. The interval $[A, B]$ is defined as all Hermitian X such that $A \leq X \leq B$. Note one can test whether $A \geq B$ in polynomial-time by finding all the eigenvalues of $A - B$.

Remark 2.10. For intervals of the form $[-\alpha I, \alpha I]$, saying A is in this interval is equivalent to saying $\|A\| \leq \alpha$. Note that this means the probability bounded in Theorem 1.1 is exactly the probability $\Pr[\frac{1}{k} f(W) - \mathbb{E}[f] \notin [-\gamma I, \gamma I]]$

[1] develops a theory of probability inequalities for Hermitian matrices, including analogues of the traditional Markov, Chebyshev, and Chernoff inequalities. We state some of the theorems from [1] here without proof.

Lemma 2.11 (Markov's inequality [1]). *Let Y be a matrix-valued random variable taking value in the Hermitian, p.s.d. matrices of dimension d. Let $M = \mathbb{E}[Y]$ and let A also be a Hermitian p.s.d. matrix. Then we have that*

$$\Pr[Y \not\leq A] \leq \mathrm{Tr}(MA^{-1})$$

We will apply Bernstein's trick (taking the exponential generating function and then applying Markov) on this lemma to get an exponential bound. This uses the *matrix exponential*:

Definition 2.12. $\exp(A) = I + A + A^2/2 + \ldots = \sum_{i=0}^{\infty} A^i/i!$

This series is convergent for all A. Also, if X is Hermitian then so is $\exp(X)$, and $\exp(X) \geq 0$ for any Hermitian X. In general $\exp(A + B)$ is not necessarily equal to $\exp(A)\exp(B)$. However, the Golden-Thompson inequality gives a relationship between the *traces* of $\exp(A + B)$ and $\exp(A)\exp(B)$:

Theorem 2.13 ([12, 31]). *For A, B Hermitian matrices we have*

$$\mathrm{Tr}(\exp(A + B)) \leq \mathrm{Tr}(\exp(A) \cdot \exp(B))$$

We can use the definition of matrix exponential to apply Bernstein's trick to Lemma 2.11 and get the following.

Lemma 2.14 ([1]). *If Y is a matrix-valued random variable and B is a constant matrix, both taking value in the Hermitian matrices of the same dimension, then for every $t > 0$*

$$\Pr[Y \not\preceq B] \leq \mathrm{Tr}(\mathbb{E}[\exp(t(Y - B))])$$

[1] uses this and Theorem 2.13 to get a Chernoff bound, similar to Theorem 1.1 but with true independent samples. We state only a special case of their bound.

Theorem 2.15 (Chernoff bound, [1]). *Let $Y_1, \ldots, Y_k$ be independent, identically distributed random variables taking value in the Hermitian matrix interval $[-I, I]$ with mean 0. Suppose $1 \geq \gamma > 0$. Then $\Pr[\|\frac{1}{k}\sum_{i=1}^{k} Y_i\| > \gamma] \leq 2de^{-\gamma^2 k/(2\ln 2)}$.*

The constant is better than what we are able to achieve in Theorem 3.1 but qualitatively the bound achieves the same effect.

3. Randomness-efficient sampling of matrix-valued functions

In Section 3.1 we prove Theorem 1.1 and finally in Section 3.2 we derive the randomness-efficient and derandomized samplers.

3.1. Expander walks for Matrix-Valued Random Variables

In this section we prove the main theorem. This will involve applying perturbation theory akin to that of [11, 21] to prove Lemma 3.4. Note in the d-dimensional case there is an extra factor of d in both the independent sampling Chernoff bound of Theorem 2.15 and in our expander walk Chernoff bound Theorem 1.1. This is because by bounding a $d \times d$ Hermitian matrix, we are in some sense bounding d variables (the eigenvalues) simultaneously, and so the d falls out of a union bound.

The d-dimensional case is delicate for several reasons. First, because matrices do not necessarily commute, the matrix exponential does not behave as the real exponential, which is why we need Theorem 2.13. Second, [11, 21] study the perturbation of the largest eigenvalue of the normalized adjacency matrix A of the graph, which has multiplicity 1. Although we also study a similar eigenvalue, it will have multiplicity d instead of 1. Because of this, the techniques of [11, 21] do not apply in the obvious way.

Recall the setting of the main theorem. We have a random walk $W = (Y_1, \ldots, Y_k)$ on an expander $G = (V, E)$, where Y_i is the i'th vertex visited in the walk. The spectral gap of G is ε. For simplicity of notation in the proof we will only prove Theorem 3.1 below. For any f such that $\|f(v)\| \leq 1$ for all v, we can simply shift and scale f to fit the hypotheses of Theorem 3.1, changing only constants in the bound. Thus our Main Theorem 1.1 follows immediately from Theorem 3.1 and Remark 2.10.

Theorem 3.1. *Let $f : V \to [-I, I]$ and $\mathbb{E}[f(v)] = 0$. Let $f(W) = \sum_{i=1}^{k} f(Y_i)$. Then for every $1 \geq \gamma > 0$ and every $k \geq \frac{4}{\gamma}$, we have the two following bounds:*

$$\Pr[\tfrac{1}{k}f(W) \not\preceq \gamma I] \leq de^{-\gamma^2\varepsilon k/60}$$
$$\Pr[\tfrac{1}{k}f(W) \not\succeq -\gamma I] \leq de^{-\gamma^2\varepsilon k/60}$$

Proof of Theorem 3.1. Note that the lower bound follows immediately from the upper bound by replacing f with $-f$, thus we only prove the first inequality.

We reduce the problem of computing the probability bound to bounding the largest eigenvalue of a perturbation matrix. Then in the proof of the Main Lemma 3.4, we use perturbation theory to bound the norm of this perturbed operator, which in turn implies the theorem.

First apply Lemma 2.14 to the expression, then bring out γI:

$$\begin{aligned}\Pr[\tfrac{1}{k}f(W) \not\preceq \gamma I] &\leq \mathrm{Tr}\mathbb{E}[\exp(t(f(W) - k\gamma I))] \\ &\leq e^{-\gamma k t}\mathrm{Tr}\mathbb{E}[\exp(tf(W))]\end{aligned}$$

Applying Theorem 2.13 and the fact that trace and expectation commute, we can write that this is at most

$$\begin{aligned}&\leq e^{-\gamma k t}\mathbb{E}\mathrm{Tr}\left[\exp\left(t\left(\sum_{i=1}^{k} f(Y_i)\right)\right)\right] \\ &\leq e^{-\gamma k t}\mathbb{E}\mathrm{Tr}\left[\prod_{i=1}^{k}\exp(tf(Y_i))\right] \\ &\leq e^{-\gamma k t}\mathrm{Tr}\mathbb{E}\left[\prod_{i=1}^{k}\exp(tf(Y_i))\right]\end{aligned}$$

It is important to note here that the $\exp(tf(Y_i))$ do not commute so the product notation means the product in the order $\exp(tf(Y_k))\exp(tf(Y_{k-1}))\ldots\exp(tf(Y_1))$.

Let A be the normalized adjacency matrix of G and let $\tilde{A} = I_d \otimes A$. One can visualize this as A but where each entry is $A_{i,j}I_d$ instead of just $A_{i,j}$. Define, $\tilde{D}_t$, which is the $dn \times dn$ block diagonal matrix with $d \times d$ blocks where the i'th diagonal block is $\exp(tf(i))$. Define $\tilde{u}$ to be the $dn \times d$ matrix $I_d \otimes u$ where $u = [1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$ is the unit uniform column vector. This is in some sense a "unit eigenvector of the eigenvalue 1 of $\tilde{A}$".

Claim 3.2. *We have that* $\mathbb{E}\left[\prod_{i=1}^{k}\exp(tf(Y_i))\right] = \tilde{u}^T(\tilde{D}_t\tilde{A})^k\tilde{u}$

Proof of Claim 3.2. We may view the expectation on the LHS to be taken over all walks on G. Let $y = (y_1, \ldots, y_k)$ be a walk, y_i the i'th vertex visited of the walk, p_y be the probability of y, and $f(y)$ the value of the walk. Then

$$\mathbb{E}\left[\prod_{i=1}^{k} \exp(tf(Y_i))\right] = \sum_y p_y \prod_{i=1}^{k} \exp(tf(y_i))$$

We interpret the expression on the RHS as follows. We initialize the value of the walk to I, then take a random walk starting from a random start vertex, and at each vertex y_i we encounter, we multiply the value of the walk on the left by $\exp(tf(y_i))$. Thus a calculation yields that the RHS is $\tilde{u}^T(\tilde{D}_t\tilde{A})^k\tilde{u}$. ∎

Now note that $\mathrm{Tr}(\tilde{u}^T(\tilde{D}_t\tilde{A})^k\tilde{u}) = \sum_{i=1}^{d}\langle(e_i \otimes u), (\tilde{D}_t\tilde{A})^k(e_i \otimes u)\rangle \leq d\|(\tilde{D}_t\tilde{A})^k\|$. The final inequality follows from applying Cauchy-Schwarz, since $\|e_i \otimes u\| = 1$.

Thus we have

$$\Pr[\tfrac{1}{k}f(W) \not\preceq \gamma I] \leq de^{-\gamma kt}\|(\tilde{D}_t\tilde{A})^k\| \qquad (3.1)$$

The proof requires a bound on $\|(\tilde{D}_t\tilde{A})^k\|$.

Definition 3.3. $\tilde{A}(t) = \tilde{D}_{t/2}\tilde{A}\tilde{D}_{t/2}$

Note that $\tilde{A}(0) = \tilde{A}$ and $\tilde{D}_t\tilde{A}$ is similar $\tilde{A}(t)$. We will apply perturbation theory to $\tilde{A}(t)$ to get the Main Lemma:

Lemma 3.4 (Main Lemma). *$\|\tilde{A}(t)\| \leq 1 + (7.5/\varepsilon)t^2$ for all $t \leq \varepsilon/15$.*

The intuition behind the Main Lemma is that $\tilde{A}(t)$ is close to $\tilde{A}$ for small t. In particular, the spectral gap of $\tilde{A}$ is large so the largest eigenvalue of $\tilde{A}(t)$ is close to the largest eigenvalue 1 of $\tilde{A}$. Note interestingly that d, the dimension of the blocks in the matrices we work with, does not appear at all in the above lemma. Intuitively, this is because the spectral behavior of $\tilde{A}$ depends only on its spectral gap between 1 and λ_2, not its size, even though 1 and λ_2 are of multiplicity d.

Before we prove the Main Lemma, we use it to derive Theorem 3.1. We will fix $t = \gamma\varepsilon/15$ later. Thus, since $\|\tilde{D}_{t/2}\| \leq e^{t/2}$ and $\|\tilde{D}_{-t/2}\| \leq e^{t/2}$, we have

$$\begin{aligned}\|(\tilde{D}_t\tilde{A})^k\| = \|\tilde{D}_{t/2}(\tilde{A}(t))^k\tilde{D}_{-t/2}\| &\leq e^t\|\tilde{A}(t)\|^k \\ &\leq e^t(1 + (7.5/\varepsilon)t^2)^k\end{aligned}$$

which is at most $e^{t+(7.5k/\varepsilon)t^2}$ by the fact that $1 + \alpha \leq e^\alpha$ for all $\alpha \in \mathbb{R}$. So from Equation 3.1 we have

$$\Pr[\tfrac{1}{k}f(W) \not\preceq \gamma I] \leq de^{-\gamma kt+(7.5k/\varepsilon)t^2+t}$$

We fix $t = \gamma\varepsilon/15$, which along with the fact that $k \geq \frac{4}{\gamma}$ gives us that

$$\Pr[\tfrac{1}{k}f(W) \not\preceq \gamma I] \leq de^{-\gamma^2\varepsilon k/60}$$

∎

Now we turn to the proof of the Main Lemma:

Proof of Lemma 3.4. $\tilde{A}(t) = \tilde{D}_{t/2}\tilde{A}\tilde{D}_{t/2}$ is an analytic perturbation of the form $\tilde{A}(t) = \sum_{i=0}^{\infty} t^i\tilde{A}_i$ where $\tilde{A}(0) = \tilde{A}_0 = I_d \otimes A$, and where the other coefficients are given by the following.

Claim 3.5.

$$\tilde{A}_i = \frac{1}{i!}\frac{1}{2^i}\sum_{j=0}^{i}\binom{i}{j}\tilde{\Delta}^i\tilde{A}\tilde{\Delta}^j$$

Here $\tilde{\Delta}$ is the block diagonal matrix $\mathrm{diag}(f(i))$. This claim is easily derived by direct calculation using the Taylor expansion of $\tilde{D}_{t/2}$. Since $\tilde{A}$ and $\tilde{\Delta}$ are Hermitian it follows that $\tilde{A}(t)$ is Hermitian for all t, so its eigenvalues are real and the largest eigenvalue $\tilde{\lambda}(t) = \|\tilde{A}(t)\|$. Furthermore Theorem 2.4 applies to $\tilde{A}(t)$ and its perturbed eigenvalue $\tilde{\lambda}(t)$, because $\tilde{A}(0) = \tilde{A}$ is Hermitian and one can calculate from Claim 3.5 that $\|\tilde{A}_i\| \leq 1$ for all i.

We want to find the largest eigenvalue of $\tilde{A}(t)$. It is easy to verify using Claim 3.5 that $\|\tilde{A}_i\| \leq 1/2^{i-1}$ for all $i \geq 1$. In addition $t \leq \varepsilon/15$, so we can apply Lemma 2.5, which tells us that all the eigenvalues of $\tilde{A}(t)$ in the range $[1 - \varepsilon/2, 1 + \varepsilon/2]$ split from 1. In particular, the trivial bound $\|\tilde{A}(t)\| \leq e^t$ tells us that $\|\tilde{A}(t)\| < 1 + \varepsilon/2$ for $t \leq \varepsilon/15$, and therefore the largest eigenvalue of $\tilde{A}(t)$ splits from 1.

By Theorem 2.4 there is an analytic projection-valued function $\tilde{P}(t)$ with matrix coefficients $\tilde{P}_i$ that projects onto the eigenspace of all the eigenvalues splitting from the eigenvalue 1 of $\tilde{A}$. Recall that $\tilde{P}(0) = \tilde{P}_0$ is the projection onto the space spanned by the eigenvectors of the eigenvalue 1 of $\tilde{A}$.

We noted earlier that the eigenvalue 1 of $\tilde{A}$ may split into d distinct eigenvalues upon perturbation by $\tilde{D}_t$ because it is of multiplicity d. Fortunately we are simply interested in the largest one that splits from 1, which is still in the space that $\tilde{P}(t)$ projects onto.

We thus have that $\|\tilde{A}(t)\| = \|\tilde{A}(t)\tilde{P}(t)\|$. We remark for comparison here that the techniques of Gillman and Lezaud [11, 21] fail at this point because the assumption that 1 is an eigenvalue of multiplicity 1 is essential to their analyses.

Continuing onwards, we wish to bound $\tilde{\lambda}(t) = \|\tilde{A}(t)\tilde{P}(t)\|$. For intuition, consider that $\tilde{P}(t)$ is a projection onto eigenspaces of $\tilde{A}(t)$, so we have that $\tilde{A}(t)\tilde{P}(t) = \tilde{P}(t)\tilde{A}(t)\tilde{P}(t)$. By calculating the power series expansion of $\tilde{P}(t)\tilde{A}(t)\tilde{P}(t)$ one can see that the linear term is 0, and

the rest are $O(t^2)$ for small enough t. This is why one expects that $\tilde{\lambda}(t) \leq 1 + O(t^2)$. However we use a different approach to actually prove the lemma.

Formalizing this intuition, we wish to bound

$$\begin{aligned}\tilde{\lambda}(t) &= \|\tilde{A}(t)\tilde{P}(t)\| = \|\tilde{P}(t) + (\tilde{A}(t) - I)\tilde{P}(t)\| \\ &\leq 1 + \|(\tilde{A}(t) - I)\tilde{P}(t)\|\end{aligned} \tag{3.2}$$

$(\tilde{A}(t) - I)\tilde{P}(t)$ is a power series, which is given by Theorem 2.8. We will show shortly that the constant and linear coefficients of this series are 0 and whose i'th coefficient for $i \geq 2$ has norm $\leq (\frac{5}{\varepsilon})^{i-1}$. Therefore the norm of the entire series is bounded as in the claim below:

Claim 3.6. *$\|(\tilde{A}(t)-I)\tilde{P}(t)\| \leq (7.5/\varepsilon)t^2$ for all $t \leq \varepsilon/15$.*

Since our choice of $t = \gamma\varepsilon/15$ in the proof of Theorem 3.1 satisfies $t \leq \varepsilon/15$, we can apply this claim to Equation 3.2 to finally get $\tilde{\lambda}(t) \leq 1 + (7.5/\varepsilon)t^2$. ∎

Thus it only remains to prove Claim 3.6.

Proof of Claim 3.6. We apply Theorem 2.8 to our perturbation $\tilde{A}(t) = \sum_{i=0}^{\infty} t^i \tilde{A}_i$. Equation 2.2 implies that

$$\|(\tilde{A}(t) - I)\tilde{P}(t)\| = \left\|\sum_{i=1}^{\infty} t^i \tilde{Z}^{(i)}\right\| \leq \sum_{i=1}^{\infty} t^i \|\tilde{Z}^{(i)}\| \tag{3.3}$$

where

$$\tilde{Z}^{(i)} = -\sum_{k=1}^{i} \sum_{\substack{\mu_1+\ldots+\mu_k=i \\ \sigma_1+\ldots+\sigma_{k+1}=k-1 \\ \mu_j \geq 1, \sigma_j \geq 0}} \tilde{S_0}^{(\sigma_1)} \tilde{A}_{\mu_1} \tilde{S_0}^{(\sigma_2)} \ldots \tilde{A}_{\mu_k} \tilde{S_0}^{(\sigma_{k+1})} \tag{3.4}$$

where $\tilde{S_0}^{(0)} = \tilde{P}_0$, $\tilde{S_0}^{(\sigma)} = -(-\tilde{S_0})^{\sigma}$ for $\sigma \geq 1$, and $\tilde{S_0}$ is the reduced resolvent of $\tilde{A}$ for the eigenvalue 1.

We see that

$$\tilde{Z}^{(1)} = \tilde{P}_0 \frac{1}{2}(\tilde{\Delta}\tilde{A} + \tilde{A}\tilde{\Delta})\tilde{P}_0 = \tilde{P}_0\tilde{\Delta}\tilde{P}_0$$

and we claim that this last expression is actually 0. For any $\tilde{x} \in \mathbb{C}^{dn}$, we have

$$\begin{aligned}\tilde{P}_0\tilde{\Delta}\tilde{P}_0\tilde{x} &= \tilde{P}_0\tilde{\Delta}(x \otimes u) \\ &= \tilde{P}_0\left(\sum_{i=1}^{n} f(i)x \otimes e_i\right) \\ &= \left(\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n} f(i)x\right) \otimes u \\ &= 0\end{aligned}$$

We use two facts in the above. First, $\tilde{P}_0$ is the projection onto the space $\{x \otimes u \mid x \in \mathbb{C}^d\}$. That is, if we decompose $\tilde{x} = \sum_{i=1}^{n} x_i \otimes e_i$ where the $x_i \in \mathbb{C}^d$, then $\tilde{P}_0\tilde{x} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \otimes u$. The other fact, used in the last line, is that $\sum f(i) = n\mathbb{E}[f] = 0$.

For $i \geq 2$ we use Lemma 2.7 and the fact that the spectral gap is the separation of 1 from the other eigenvalues to see that $\|\tilde{S_0}\| = \frac{1}{\varepsilon}$. Also, it is evident that $\|\tilde{P}_0\| = 1$ since it is a projection, and we have already remarked that $\|\tilde{A}_i\| \leq 1$. Thus each summand of Equation 3.4 has norm at most $(1/\varepsilon)^{i-1}$.

Notice that the number of terms in the summation in Equation 3.4 is exactly

$$\sum_{k=1}^{i} \binom{i-1}{k-1}\binom{2k-1}{k}$$

It is clear that $\binom{2k-1}{k} = \frac{1}{2}\binom{2k}{k}$ and by Stirling's formula we have $\binom{2k}{k} \leq 4^k/\sqrt{k\pi}$. Thus the number of terms is at most

$$\begin{aligned}\frac{1}{2} + \frac{1}{2\sqrt{\pi}}\sum_{k=2}^{i}\frac{4^k}{\sqrt{k}}\binom{i-1}{k-1} &\leq \frac{1}{2} + \frac{2}{\sqrt{2\pi}}\sum_{k=1}^{i-1} 4^k\binom{i-1}{k} \\ &\leq \frac{1}{2} + \sqrt{\frac{2}{\pi}}(5^{i-1} - 1)\end{aligned}$$

We obtain the last inequality by recognizing a binomial expansion. Finally

$$\frac{1}{2} + \sqrt{\frac{2}{\pi}}(5^{i-1} - 1) \leq 5^{i-1}$$

for all $i \geq 2$. Therefore $\|\tilde{Z}^{(i)}\| \leq (\frac{5}{\varepsilon})^{i-1}$ for all $i \geq 2$.

Since $\tilde{Z}^{(1)} = 0$ and $\|\tilde{Z}^{(i)}\| \leq (\frac{5}{\varepsilon})^{i-1}$ for $i \geq 2$, we have that the RHS of Equation 3.3 is at most

$$\frac{5}{\varepsilon}t^2\sum_{i=0}^{\infty}(\tfrac{5t}{\varepsilon})^i$$

Thus for $t \leq \frac{\varepsilon}{15}$ it is clear that this is at most $(7.5/\varepsilon)t^2$. ∎

3.2. A randomness-efficient sampler for matrix-valued functions

Here we use Theorem 3.1 to derive a randomness-efficient sampler for matrix-valued functions over arbitrary distributions. We then derandomize this sampler to get deterministic samples in polynomial time.

Theorem 3.1 treats sampling a function $f : [n] \to [-I, I]$ uniformly, where $[n] = \{1, \ldots, n\}$. That is, let $x \xleftarrow{R} X$ denote *sampling x from X uniformly*, then Theorem 3.1 allows us to (approximately) sample $f(x)$ where $x \xleftarrow{R} [n]$ using little randomness. Here we generalize this so that the distribution on $[n]$ is not necessarily uniform. In the following, let $\mathbb{E}_p[f]$ denote the expectation of $f(Y)$ where Y is sampled from $[n]$ according to the probability distribution p.

Proposition 3.7. *Let $p : [n] \to [0,1]$ be a probability distribution on $[n]$. For any $1 \geq \gamma > 0$ and every $k \geq \frac{4}{\gamma}$, we can construct a $\mathrm{poly}(n)$-time computable sampler $\sigma : \{0,1\}^r \to [n]^k$ with $r = \log n + O(k) + O(\log \frac{1}{\gamma})$ such that for all functions $f : [n] \to [-I_d, I_d]$ with $\mathbb{E}_p[f] = 0$ we have*

$$\Pr_{w \stackrel{R}{\leftarrow} \{0,1\}^r} \left[\left\| \frac{1}{k} \sum_{i=1}^{k} f(\sigma(w)_i) \right\| \leq \gamma \right] \geq 1 - 2de^{-\gamma^2 k/70} \tag{3.5}$$

Proof of Proposition 3.7. Our strategy is to construct in time polynomial in n a constant-degree expander graph $G = (V, E)$ and a map $\varphi : V \to [n]$. Our sampler σ will map a walk on the expander of length k (which can clearly be encoded using $r = \log |V| + O(k)$ bits) to $[n]^k$, namely all the vertices it visits on the walk.

Recall we can construct Ramanujan graphs efficiently from Theorem 2.1, so let us pick the degree such that the spectral gap is at least 0.95. Choose such a graph of size $\geq \frac{40n}{\gamma}$. Call this graph $G = (V, E)$.

We define the function $\varphi : V \to [n]$ such that for each value $y \in [n]$ we map any $\llbracket p(y) \cdot |V| \rrbracket$ vertices in G to y, where the brackets $\llbracket \cdot \rrbracket$ denote rounding either up or down, so that in the end all the vertices V are mapped to $[n]$. Thus G, φ give an altered distribution p_G, which is $p_G(y) = \Pr_{v \stackrel{R}{\leftarrow} V}[\varphi(v) = y]$.

Claim 3.8. $\|\mathbb{E}_{p_G}[f]\| \leq \gamma/40$

We first use this claim to prove the proposition. Let $f'(v) = \frac{40}{40+\gamma}(f(v) - \mathbb{E}_{p_G}[f])$, then clearly $f' : V \to [-I, I]$ and $\mathbb{E}_{p_G}[f'] = 0$. Take a random walk of length k on G and let this sequence be called W. Then we have by Theorem 3.1, Claim 3.8, and Remark 2.10 that

$$\Pr[\tfrac{1}{k} f \circ \varphi(W) \in [-\gamma I, \gamma I]] \geq$$
$$\Pr\left[\tfrac{1}{k} f' \circ \varphi(W) \in \left[-\frac{39\gamma}{41} I, \frac{39\gamma}{41} I\right]\right] \geq 1 - 2de^{-\gamma^2 k/70}$$

where the inequality on the first line is obtained by adding $-\mathbb{E}_{p_G}[f]$ and scaling by $\frac{40}{40+\gamma}$ to both sides of the event and then applying Claim 3.8 and the fact that $\gamma \leq 1$.

We can encode each walk by $r = \log |V| + O(k)$ bits, which by our choice of $|V|$ is exactly $r = \log n + O(k) + O(\frac{1}{\gamma})$. Thus σ is the map that for any walk $w = (v_1, \ldots, v_k)$ outputs $(\varphi(v_1), \ldots, \varphi(v_k))$. We can plug σ into the above calculations to derive the bounds of Proposition 3.7. ∎

Proof of Claim 3.8. The only thing remaining is to show that the $G = (V, E)$ we chose is large enough to satisfy

$$\gamma/40 \geq \|\mathbb{E}_{p_G}[f]\| = \left\| \sum_{y \in [n]} (p_G(y) - p(y)) f(y) \right\|$$

where we use the fact that $\mathbb{E}_p[f(y)] = 0$. Note that since $\|f(y)\| \leq 1$ for all y, it suffices to show that

$$\sum_{y \in [n]} |p_G(y) - p(y)| \leq \gamma/40$$

Since $p_G(y) = \llbracket p(y)|V| \rrbracket / |V|$, this is

$$\sum_{y \in [n]} \left| \frac{\llbracket p(y)|V| \rrbracket - p(y)|V|}{|V|} \right|$$

The numerator is at most 1, so after summing we get $\|\mathbb{E}_{p_G}[f] - \mathbb{E}_p[f]\| \leq n/|V|$, and thus it suffices to take $|V| \geq \frac{40n}{\gamma}$. ∎

An easy corollary of the proposition states that for short enough walks we can completely derandomize the procedure.

Corollary 3.9. *Suppose we are in the setting of Proposition 3.7. Then there is a $k = O(\log d)$ and $n \cdot \mathrm{poly}(d/\gamma)$ algorithm (in fact an $\mathbf{NC}$ algorithm) to find a sample $T = (\sigma_1, \ldots, \sigma_k)$ such that $\| \sum_{i=1}^{k} f(\sigma_i)\| \leq k\gamma$.*

Proof. Take the smallest integer $k > \frac{70}{\gamma^2}(\log d + \log 2)$, then we have that the RHS of Equation 3.5 is positive. Thus since $r = \log n + O(\log d) + O(\log \frac{1}{\gamma})$, by enumerating over all $w \in \{0,1\}^r$ in time $2^r = n(d/\gamma)^{O(1)}$ we can deterministically find w_0 such that $\| \sum_{i=1}^{k} f(\sigma(w_0)_i)\| \leq k\gamma$. Let $T = \sigma(w_0)$. ∎

Remark 3.10. We note that the f in Proposition 3.7 and Corollary 3.9 is not identical to the one in Theorem 1.1. This is unimportant as we may apply these results to any bounded function f by shifting and scaling f; this only changes the resulting bounds by constant factors.

4. Applications

In Section 4.1 we apply Theorem 1.1 to prove Theorem 1.2 and in Section 4.2 we apply this to affine homomorphism testing to get Corollary 1.3.

4.1. A Derandomization of the Alon-Roichman Theorem

In this section we prove Theorem 1.2, which gives a deterministic polynomial time algorithm for the Alon-Roichman theorem. We first give a simple version of the

proof of the Alon-Roichman Theorem due to [20, 22] that does not use representation theory. We note that better constants in the final size of S may be achieved using the proof based on representation theory given in [20, 22].

Theorem 4.1 ([4, 22, 20]). *Fix $\beta < 1$ and $q < 1$.[7] For an arbitrary group H, by picking a random generating multi-set T of size $O(\frac{1}{\beta^2}\log|H|)$ and taking its symmetric closure multi-set $S = T \sqcup T^{-1}$ we have that the second-largest eigenvalue of the Cayley graph $\lambda_2(X(H;S))$ satisfies*

$$\Pr[\lambda_2(X(H;S)) \leq \beta] > q$$

Proof. Pick a generating multi-set set T uniformly at random from H and take its symmetric closure $S = T \sqcup T^{-1}$ (i.e. if a is in T i times and in T^{-1} j times then a is in S $i+j$ times). Define the homomorphism R such that for each $h \in H$, $R(h)$ is the $|H| \times |H|$ (real-valued) permutation matrix associated with the action of h on H. Define

$$f(h) = \frac{1}{2}((R(h) - J/n) + (R(h^{-1}) - J/n))$$

where J is the matrix with 1 in all entries. It is easy to observe that $f(h)$ is symmetric (and thus Hermitian), and $\mathbb{E}[f] = 0$. If we let P be the projection onto the space orthogonal to u the uniform vector, then a calculation shows that $PR(h) = R(h) - J/n$. Thus $f(h) = \frac{1}{2}(PR(h) + PR(h^{-1}))$, and looking at $\|f(h)\|$ it is also clear that $-I \leq f(h) \leq I$.

Finally, a simple calculation shows that $\frac{1}{|T|}\sum_{h\in T} f(h) = PA$ where P is the projection mentioned above and A is the adjacency matrix of $X(G;S)$. Therefore we have $\lambda_2(X(H;S)) = \|\frac{1}{|T|}\sum_{h\in T} f(h)\|$. So we wish to bound

$$\Pr[\lambda_2(X(H;S)) \leq \beta] = \Pr\left[\left\|\frac{1}{|T|}\sum_{h\in T} f(h)\right\| \leq \beta\right] \tag{4.1}$$

We can apply Theorem 2.15 to get that the RHS is $\geq 1 - 2|H|e^{-\beta^2 k/(2\ln 2)}$. Thus choosing the smallest integer $|T| > \frac{2\ln 2}{\beta^2}(\log|H| + \log\frac{2}{1-q})$ shows that the RHS is $> q$. ∎

Our derandomization, Theorem 1.2, follows easily from Theorem 4.1 and Corollary 3.9

Proof of Theorem 1.2. We wish to apply Corollary 3.9. We identify H with $[|H|]$ and let p be the uniform distribution over $[|H|]$. We apply Corollary 3.9 to get a sample T of size $O(\frac{1}{\beta^2}\log|H|)$ in time $|H|2^{O(|T|)} = |H|^{O(1)}$ such that $\|\frac{1}{|T|}\sum_{h\in T} f(h)\| \leq \beta$ and hence $\lambda_2(X(H;T\sqcup T^{-1})) \leq \beta$. ∎

[7] Here we may take $q = 1 - 1/\mathrm{poly}(n)$, but constant suffices for our purposes.

4.2. Improved Affine Homomorphism Testers

Theorem 1.2 answers a question about the derandomization of homomorphism testers posed in [29]. In this section we will use Theorem 1.2 to prove Corollary 1.3.

Recall that an *affine homomorphism* between two groups H, H' is a map $f : H \to H'$ such that $f^{-1}(0)f$ is a homomorphism. An (δ,η)-test for affine homomorphisms is a tester that accepts any affine homomorphism surely and rejects with probability $1-\delta$ any $f : H \to H'$ which is η far from being an affine homomorphism. Here distance is measured by the normalized Hamming distance: $d(f,g) = \Pr[f(x) \neq g(x)]$.

[29] showed how to efficiently construct a tester $T_{H\times S}$ where $\lambda_2(X(H;S)) < \lambda$: simply pick a random element $x \overset{R}{\leftarrow} H$ and a random element of $y \overset{R}{\leftarrow} S$ and check to see that $f(0)f(x)^{-1}f(xy) = f(y)$. It is clear this accepts f surely if f is an affine homomorphism. [29] shows that if $12\delta < 1-\lambda$ then this rejects with probability $1-\delta$ any f that is $\frac{4\delta}{1-\lambda}$-far from being an affine homomorphism.

Theorem 4.2 ([29]). *For all groups H, H' and $S \subseteq H$ an expanding generating set such that $\lambda_2(X(H;S)) < \lambda$, we can construct a tester $T_{H\times S}$ that surely accepts any affine homomorphism $f : H \to H'$ and rejects with probability at least $1-\delta$ any $f : H \to H'$ which is $4\delta/(1-\lambda)$ far from being an affine homomorphism, given that $\frac{12\delta}{1-\lambda} < 1$. That is, $T_{H\times S}$ is a $(\delta, \frac{4\delta}{1-\lambda})$-test for affine homomorphisms.*

In [29] the deterministic construction of S gave a set of size $|H|^{\varepsilon}$. The explicit construction given in [29] requires that $T_{H\times S}$ use $(1+\varepsilon)\log|H|$ random bits and asks whether it is possible to improve this dependency on randomness. Theorem 1.2 allows us indeed to improve this dependency to the following.

Recall Corollary 1.3:

Corollary 1.3 (Restated). *Given an arbitrary group H, one can construct in time $|H|^{O(1)}$ a* homomorphism tester *for functions on H which uses only $\log|H| + \log\log|H| + O(1)$ random bits.*

This follows easily from Theorem 1.2:

Proof of Corollary 1.3. Theorem 4.2 says we can construct a homomorphism tester that only uses randomness to pick an element of H and an element of an expanding generating set of H. Theorem 1.2 implies this only requires $\log|H| + \log\log|H| + O(1)$ random bits since we can deterministically construct an expanding generating set of size $\log|H|$ in polynomial time. ∎

Note that Corollary 1.3 is essentially optimal for "Cayley testers" of the above form, i.e. testers that pick one element at random and a second from an expanding generating

set. This is because the tester requires that S be an expanding generating set of H and there are groups (for example, $\mathbb{Z}_2^n$) for which $\Omega(\log |H|)$ generators are necessary for the Cayley graph to expand. However, note that [15] prove the existence of testers for homomorphisms $H \to H'$ where $|H'| = O(1)$ that use only $\log |H| + O(1)$ bits of randomness. Finding explicit such constructions remains an interesting open problem.

5. Acknowledgments

We would like to thank Boaz Barak for many helpful comments about this research. We would also like to thank Salil Vadhan for his careful reading and comments on an early draft of this paper. Finally, we thank Amir Shpilka, Sanjeev Arora, and Scott Aaronson for their suggestions and input.

References

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[2] M. Ajtai, J. Komlos, and E. Szemeredi. Deterministic simulation in logspace. In ACM, editor, *Proceedings of the nineteenth annual ACM Symposium on Theory of Computing, New York City, May 25–27, 1987*, pages 132–140, New York, NY, USA, 1987. ACM Press. ACM order no. 508870.

[3] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

[4] N. Alon and Y. Roichman. Random cayley graphs and expanders. *RSA: Random Structures & Algorithms*, 5, 1994.

[5] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, May 1998.

[6] H. Baumgrtel. *Analytic Perturbation Theory for Matrices and Operators*, volume 15 of *Operator Theory: Advances and Applications*. Birkhuser, 1984.

[7] Y. Bilu and S. Hoory. Hypergraph codes. *European Journal of Combinatorics*, 25(3):339–354, 2004.

[8] H. Chernoff. A measure of aymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493 – 507, 1952.

[9] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. In *Proc. 30th FOCS*, pages 14–19. IEEE, 1989.

[10] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *J. Comput. Syst. Sci.*, 22(3):407–420, June 1981.

[11] D. Gillman. A chernoff bound for random walks on expander graphs. In *IEEE Symposium on Foundations of Computer Science*, pages 680–691, 1993.

[12] S. Golden. Lower bounds for the helmholtz function. *Physical Review*, 137B(4):B1127–1128, 1965.

[13] O. Goldreich. A sample of samplers - a computational perspective on sampling (survey). *Electronic Colloquium on Computational Complexity (ECCC)*, 4(020), 1997.

[14] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. In *Proc. 31st FOCS*, pages 318–326. IEEE, 1990.

[15] O. Goldreich and M. Sudan. Locally testable codes and PCPs of almost-linear length. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, FOCS'2002 (Vancouver, BC, Canada, November 16-19, 2002)*, pages 13–22, Los Alamitos-Washington-Brussels-Tokyo, 2002. IEEE Computer Society, IEEE Computer Society Press.

[16] P. Hayden, D. Leung, P. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[17] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *Proc. 30th FOCS*, pages 248–253. IEEE, 1989.

[18] N. Kahale. Eigenvalues and expansion of regular graphs. *Journal of the ACM*, 42(5):1091–1106, Sept. 1995.

[19] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, 1980.

[20] Z. Landau and A. Russell. Random cayley graphs are expanders: a simplified proof of the alon-roichman theorem. *The Electronic Journal of Combinatorics*, 11(2), 2004.

[21] P. Lezaud. Chernoff-type bound for finite markov chains. *Annals of Applied Probability*, 8(3):849–867, 1998.

[22] P.-S. Loh and L. J. Schulman. Improved expansion of random cayley graphs. *Discrete Mathematics and Theoretical Computer Science*, 6(2):523–528, 2004.

[23] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[24] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.

[25] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, Aug. 1993.

[26] N. Nisan and A. Ta-Shma. Extracting randomness: A survey and new constructions. *J. Comput. Syst. Sci.*, 58(1):148–173, 1999.

[27] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proc. 32th STOC*, pages 3–13. ACM, 2000.

[28] R. Shaltiel. Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 2002. Available from `http://www.wisodm.weizmann.ac.il/~ronens`.

[29] A. Shpilka and A. Wigderson. Derandomizing homomorphism testing in general groups. In *Proc. 36th STOC*, pages 427–435. ACM, 2004.

[30] D. Spielman. *Computationally Effi cient Error-Correcting Codes and Holographic Proofs*. PhD thesis, M.I.T., 1995.

[31] C. J. Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1823, 1965.