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

Avi Wigderson David Xiao

November 21, 2005

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 [AW02], 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 analyses of the Ajtai-Komlos-Szemeredi sampler for real-valued functions [Gil93, Lez98, AKS87].

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 [AR94, LR04, LS04]: 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 [SW04]. 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. [AW02, HLSW04]), in which we efficiently attain the integrality gap when the fractional semi-definite relaxation cost is constant.

Keywords: sampler, random variables, Chernoff, Cayley graphs, expanders

Contents

1 Introduction **2**

1.1 Background . . . 2

1.2 Our results . . . 2

1.3 Applications . . . 3

1.4 Organization of the paper . . . 4

2 Preliminaries **4**

2.1 Expander graphs . . . 4

2.2 Perturbation Theory . . . 5

2.3 Probability theory of matrix-valued random variables . . . 7

3 Randomness-efficient sampling of matrix-valued functions **8**

3.1 Chernoff bound . . . 8

3.2 The 1-dimensional case . . . 9

3.3 Expander walks for matrix-valued functions . . . 11

3.4 A randomness-efficient sampler for matrix-valued functions . . . 15

4 Applications **17**

4.1 A Derandomization of the Alon-Roichman Theorem . . . 17

4.2 Improved Affine Homomorphism Testers . . . 18

4.3 Covers of Hypergraphs and Quantum Hypergraphs . . . 19

4.3.1 Hypergraphs . . . 19

4.3.2 Quantum Hypergraphs . . . 20

5 Acknowledgments **22**

A Simple proof of a weaker expander walk bound **25**

A.1 A weak bound . . . 25

A.2 Boosting the weak bound . . . 26

B Tightness **27**

B.1 Tightness in d . . . 27

B.2 Exponent of bound is linear in ε . . . 28

C Multiple functions **29**

1 Introduction

1.1 Background

The Chernoff bound [Che52] 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 [AKS87] 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 [Gil93], which is the first to consider sampling any bounded real function (see also [Kah95, Lez98]), 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 [GG81, Mar88, LPS88, RVW00] 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 a variety of areas such as deterministic amplification [CW89, IZ89], security amplification in cryptography [GIL+90], hardness of approximation [ALM+98, AFWZ95], extractor construction (e.g. see surveys [NT99, Gol97, Sha02]), construction of efficient error-correcting codes [Spi95, BH04], construction of ε-biased spaces [NN93] 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 [AW02] 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 [AW02] generalization of the Chernoff bound states that the average of k independent points deviates significantly in norm from the mean with probability bounded by $d2^{-\Omega(k)}$. Note the linear dependence on d.

Like the Chernoff bound, this generalization has quickly found applications. Many of them are in quantum information theory (and private quantum channels) [AW02, HLSW04], where such matrices arise naturally. A notably different one is to a new proof [LR04, LS04] of the Alon-Roichman theorem [AR94], 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" [AW02] in complete analogy to the way [AKS87, Gil93] derandomized Chernoff in the real (1-dimensional) case.

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

Let $G = (V, E)$ be an expander graph with spectral gap ε. Define Y_i $(0 \le i \le k)$ 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 $\|f(v)\| \le 1$ 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 $1 \ge \gamma > 0$ and every $k \ge \frac{4}{\gamma}$ we have*

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

The dependence on d is linear, just as in the independent case of [AW02], and we show (by the examples in Appendix B) that our dependence on d, ε is essentially optimal.

Note that for $\varepsilon = 1$ (i.e. a complete graph) this bound is just independent sampling and thus the Chernoff bound of [AW02] (we state this in Theorem 2.15). For $d = 1$ it is just the 1-dimensional AKS sampler of [Gil93, AKS87, Lez98, Kah95]. 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 [Gil93, Lez98]. We also give a simpler analysis in Appendix A, 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. Interestingly, this proof follows quite closely that of the bound on the second-largest eigenvalue in the *zig-zag product* [RVW00]. We believe that the connection between the two results may be deeper and deserves further investigation.

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). [AR94] 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 [SW04] 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 $\lambda_2(X(H; S)) < \beta$.*

[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.

This will immediately imply the following optimal solution to a problem of [SW04] (see also [GS02]), 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 [AW02] concerning quantum hypergraphs. The result is discussed in Section 4.3, where we present a randomness-efficient algorithm to find a small cover of a quantum hypergraph. Here we simply note that this is a generalization of the set cover problem, and from [AW02] it can be shown to have the same integrality gap. However in this noncommutative setting the standard deterministic greedy method for obtaining an integral solution to set cover does not extend in any obvious way, whereas a derandomization of our randomness-efficient algorithm provides an efficient alternative, at least when the fractional cover (found by semidefinite programming) has constant value. Other quantum information theoretic results in [AW02, HLSW04] pose interesting derandomization problems for which we believe our result may be useful.

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.

The following other results appear in the Appendix. In Appendix A we give a simple proof of a weak version of Theorem 1.1 based on ideas from the proof of the Zig-Zag theorem [RVW00]. In Appendix B we show that our bound is essentially optimal with regard to two parameters (up to constant factors). In Appendix C we give a generalization of Theorem 1.1 that allows for an ensemble of different functions $f_1, \ldots, f_k$ rather than a single function.

2 Preliminaries

2.1 Expander graphs

Given a connected undirected d-regular graph $G = (V, E)$ on n vertices, we define its normalized adjacency 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 poly log in the size of the graph. An explicit example is the following.

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

Theorem 2.1 ([LPS88, Mar88]). *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 multi-set with elements in H. Let $S = T \sqcup T^{-1}$ denote the multi-set 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:

1. If A has spectrum $(\lambda_i)_{1 \le i \le n}$ and B has spectrum $(\mu_j)_{1 \le j \le p}$ then $A \otimes B$ has spectrum $(\lambda_i \mu_j)_{1 \le i \le n, 1 \le j \le p}$.
2. If $(A \otimes B)(x \otimes v) = Ax \otimes Bv$ where A, B are matrices and x, v are vectors of the appropriate dimensions.
3. We have $\langle x \otimes y, u \otimes v \rangle = \langle x \otimes u \rangle \langle y \otimes v \rangle$ for x, u vectors of the same dimension and u, v vectors of the same dimension.

2.2 Perturbation Theory

The proof of Lemma 3.9, 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 [Bau84] (see also [Kat80]) 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 $(A_i)_{i \ge 0}$. 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 $A(t)$.

[Bau84] 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 [Bau84, Ch. 3.2], where the $\lambda^{(i)}(t)$ are continuous at $t = 0$ and furthermore $\lambda = \lambda^{(i)}(0)$ for all $1 \le i \le m$.

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 ([Bau84, 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 ([Bau84, 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 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 [Bau84, 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 ([Bau84, 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.

[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.

Theorem 2.8 ([Bau84, p. 156]). *If $A(t), P(t)$ are defined as above, then*

$$(A(t)-I)P(t) = \sum_{i=1}^{\infty} t^i Z^{(i)} \quad \text{where} \quad Z^{(i)} = -\sum_{k=1}^{i} \sum_{\substack{\mu_1+\ldots+\mu_k=i \\ \sigma_1+\ldots+\sigma_{k+1}=k-1 \\ \mu_j\geq 1, \sigma_j \geq 0}} S_0^{(\sigma_1)} A_{\mu_1} S_0^{(\sigma_2)} \ldots A_{\mu_k} S_0^{(\sigma_{k+1})} \tag{2.2}$$

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 [Bau84]. 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 [Bau84]. This also follows from the fact that $A(t)$ is diagonalizable.

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

[AW02] 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 [AW02] here without proof.

Lemma 2.11 (Markov's inequality [AW02]). *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*:

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

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 ([Gol65, Tho65]). *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 ([AW02]). *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 $t > 0$*

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

[AW02] 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, [AW02]). *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 3.6 but qualitatively the bound achieves the same effect. See the example in Theorem B.1 of the Appendix for a discussion of the factor of d in the bound.

3 Randomness-efficient sampling of matrix-valued functions

In Section 3.1 we prove the classical Chernoff bound to give a feel of the style of proof. In Section 3.2 we prove the 1-dimensional case as another warm-up to the main theorem. In Section 3.3 we prove Theorem 1.1 and finally in Section 3.4 we derive the randomness-efficient and derandomized samplers.

3.1 Chernoff bound

We recall the following one-sided classical Chernoff bound. Better constants in the bound may be obtained; we give this proof for its simplicity.

Theorem 3.1 ([Che52]). *Let Y_i for $1 \leq i \leq k$ be i.i.d. random variables taking value in $[-1, 1]$. Suppose $\mathbb{E}[Y_i] = 0$. Then for any $1/2 \geq \gamma > 0$ we have*

$$\Pr\left[\frac{1}{k}\sum_{i=1}^{k} Y_i \geq \gamma\right] < e^{-\gamma^2 k/2.4}$$

Proof. We apply the "Bernstein trick": multiply $\frac{1}{k}\sum_{i=1}^{k} Y_i \geq \gamma$ by a positive constant t (to be set later), take the exponential, and apply Markov's inequality to get

$$\begin{aligned}\Pr\left[\frac{1}{k}\sum_{i=1}^{k} Y_i \geq \gamma\right] &= \Pr[e^{t(\sum_{i=1}^{k} Y_i)} \geq e^{\gamma k t}] \\ &\leq e^{-\gamma k t}\mathbb{E}[e^{t(\sum_{i=1}^{k} Y_i)}]\end{aligned}$$

At this point we apply independence and some analysis to bound the expectation on the RHS. We note for reference that the proof of the expander walk bound diverges exactly at this point.

By the independence of the Y_i we have that this is equal to

$$e^{-\gamma k t}\mathbb{E}[e^{tY}]^k$$

where Y is distributed like the Y_i's. Expanding e^{tY} as its power series and noticing that its $O(t^2)$ terms are bounded by $0.6t^2$ for $t < 1/2$, we have that this is at most

$$\begin{aligned}e^{-\gamma k t}\mathbb{E}[1 + Yt + 0.6t^2]^k &\leq e^{-\gamma k t}(1 + 0.6t^2)^k \\ &< e^{-\gamma k t}e^{0.6kt^2} \\ &= e^{-\gamma k t + 0.6kt^2}\end{aligned}$$

It is easy to see that the exponent is most negative when $t = \gamma/1.2$, which confirms our assumption that $t < 1/2$. This gives that the entire probability is at most $e^{-\gamma^2 k/2.4}$. ∎

3.2 The 1-dimensional case

In this section we prove the 1-dimensional expander walk Chernoff bound of [AKS87, Gil93, Lez98, Kah95]. This will be instructive because the proof of both the 1- and d-dimensional cases consist of two steps. In the first step we reduce the problem of bounding the sampling error to the problem of bounding the largest eigenvalue of a perturbation matrix. The second step consists of bounding this eigenvalue. The first step is simpler in the 1-dimensional case so we choose to illustrate it as a warm-up to the d-dimensional case. We defer the proof of the second step to the proof of the main theorem, as it is essentially identical in both cases.

Suppose we are given an expander with spectral gap $\varepsilon > 0$. Define Y_i $(1 \leq i \leq k)$ 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 walk and $f(W) = \sum_{i=1}^{k} f(Y_i)$ be the value of the walk.

Theorem 3.2 ([AKS87, Gil93, Lez98]). *For any $f : V \to [-1, 1]$ with $\mathbb{E}[f] = 0$, for any $1/2 \geq \gamma > 0$ and any $k \geq \frac{4}{\gamma}$ we have*

$$\Pr[\tfrac{1}{k}f(W) > \gamma] \leq e^{-\gamma^2 \varepsilon k/60}$$

An identical bound can be proved for $\Pr[\frac{1}{k}f(W) < -\gamma]$ by replacing f with $-f$.

We use the proof of Theorem 3.1 as a model for our proof of Theorem 3.2.

Proof of Theorem 3.2. We begin as in the proof of the traditional Chernoff bound:

$$\begin{aligned}\Pr[\tfrac{1}{k}f(W) > \gamma] &= \Pr[e^{tf(W)} > e^{\gamma k t}] \\ &\leq e^{-\gamma k t}\mathbb{E}[e^{tf(W)}]\end{aligned}$$

Thus it remains to bound $\mathbb{E}[e^{tf(W)}]$. Unfortunately we can't use independence because $Y_1, \ldots, Y_k$ are not independent. However, let A be the normalized adjacency matrix of G, and $D_t = \mathrm{diag}(e^{tf(i)})$, then we can show the following:

Claim 3.3. *Let $u = [\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}]^T$ be the unit uniform column vector. Then $\mathbb{E}[e^{tf(W)}] = \langle u, (D_t A)^k u\rangle$.*

Proof of Claim 3.3. We can use the fact that $Y_1, \ldots, Y_k$ come from a walk. Let $w = (y_1, \ldots, y_k)$ represent a particular walk, p_w its probability, and $f(w) = \sum_{i=1}^{k} f(y_i)$ its value. Then we have that

$$\mathbb{E}[e^{tf(W)}] = \sum_w p_w e^{tf(w)}$$

One interprets the right-hand side as follows. We keep track of the value of the walk, which starts at 1. Each time we arrive at a vertex y_i we multiply this value by $e^{tf(y_i)}$. It is easy to see that the expectation of this corresponds exactly to $\langle u, (D_t A)^k u\rangle$ ∎

We have by Cauchy-Schwarz that

$$\begin{aligned}\mathbb{E}[e^{tf(W)}] &= \langle u, (D_t A)^k u\rangle \\ &\leq \|(D_t A)^k\|\end{aligned}$$

Thus we require a bound on $\|(D_t A)^k\|$. The following definition will be useful:

Definition 3.4. $A(t) = D_{t/2} A D_{t/2}$

Notice that

$$(D_t A)^k = D_{t/2}(A(t))^k D_{-t/2}$$

The following lemma will give us an appropriate bound:

Lemma 3.5. $\|A(t)\| \leq 1 + (7.5/\varepsilon)t^2$ *for all* $t \leq \varepsilon/15$.

This is in fact a consequence of the Main Lemma 3.9. Indeed Lemma 3.5 is just the 1-dimensional case of Lemma 3.9. There is no advantage in clarity to treat this 1-dimensional case separately here with our technique[7], so we will use it here and prove the more general Lemma 3.9 later in Section 3.3.

Applying Lemma 3.5 and the fact that $\|D_{t/2}\| \leq e^{t/2}$ and $\|D_{-t/2}\| \leq e^{t/2}$ we get that

$$\|(D_t A)^k\| \leq \|D_{t/2}(A(t))^k D_{-t/2}\| \leq e^t(1 + (7.5/\varepsilon)t^2)^k \leq e^{(7.5k/\varepsilon)t^2 + t}$$

Then we finish the probability calculation:

$$\Pr[\tfrac{1}{k}f(W) > \gamma] \leq e^{-\gamma k t + (7.5k/\varepsilon)t^2 + t}$$

[7] Alternatively one could apply the analyses of [Gil93, Lez98] at this point.

Choose $t = \frac{\gamma\varepsilon}{15}$ and use the fact $k \geq \frac{4}{\gamma}$, which gives finally

$$\Pr[\tfrac{1}{k}f(W) > \gamma] \leq e^{-\gamma^2\varepsilon k/60} \tag{3.1}$$

■

3.3 Expander walks for matrix-valued functions

In this section we prove the main theorem following the same structure as that of Theorem 3.2.

In addition, we apply perturbation theory akin to that of [Gil93, Lez98] to prove Lemma 3.9, the d-dimensional analogue of Lemma 3.5. 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 tightness of this factor is discussed in Section B.1 of the Appendix.

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, [Gil93, Lez98] 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 [Gil93, Lez98] 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 3.6 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 3.6, changing only constants in the bound. Thus our Main Theorem 1.1 follows immediately from Theorem 3.6 and Remark 2.10.

Theorem 3.6. *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*

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

Proof of Theorem 3.6. 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 3.9, the generalization of Lemma 3.5, 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:

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

Applying Theorem 2.13 and the fact that trace and expectation commute, we can write that

this is at most

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

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 3.7. *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 3.7. The reasoning to this is similar to the reasoning for Claim 3.3. The expectation on the LHS is taken over all walks on G. Let $w = (y_1, \ldots, y_k)$ be a walk, y_i the i'th vertex visited of the walk, and p_w be the probability of w. Then

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

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

$$\Pr[\tfrac{1}{k}f(W) \not\leq \gamma I] \leq de^{-\gamma kt}\|(\tilde{D}_t\tilde{A})^k\| \tag{3.2}$$

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

Definition 3.8. $\tilde{A}(t) = \tilde{D}_{t/2}\tilde{A}\tilde{D}_{t/2}$

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 3.9 (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 3.6. 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

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

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 3.2 we have

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

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

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

■

Now we turn to the proof of the Main Lemma:

Proof of Lemma 3.9. $\tilde{A}(t) = \tilde{D}_{t/2}\tilde{A}\tilde{D}_{t/2}$ 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 3.10.

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

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 3.10 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 3.10 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 [Gil93, Lez98] 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

$$\tilde{\lambda}(t) = \|\tilde{A}(t)\tilde{P}(t)\| = \|\tilde{P}(t) + (\tilde{A}(t) - I)\tilde{P}(t)\| \leq 1 + \|(\tilde{A}(t) - I)\tilde{P}(t)\| \tag{3.3}$$

$(\tilde{A}(t) - I)\tilde{P}(t)$ 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 3.11. *$\|(\tilde{A}(t) - I)\tilde{P}(t)\| \leq (7.5/\varepsilon)t^2$ for all $t \leq \varepsilon/15$.*

Since our choice of $t = \gamma\varepsilon/15$ in the proof of Theorem 3.6 satisfies $t \leq \varepsilon/15$, we can apply this claim to Equation 3.3 to finally get $\tilde{\lambda}(t) \leq 1 + (7.5/\varepsilon)t^2$. ■

Thus it only remains to prove Claim 3.11.

Proof of Claim 3.11. 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

$$\|(\tilde{A}(t) - I)\tilde{P}(t)\| = \left\| \sum_{i=1}^{\infty} t^i \tilde{Z}^{(i)} \right\| \leq \sum_{i=1}^{\infty} t^i \|\tilde{Z}^{(i)}\| \tag{3.4}$$

where

$$\tilde{Z}^{(i)} = -\sum_{k=1}^{i} \sum_{\substack{\mu_1+\ldots+\mu_k=i \\ \sigma_1+\ldots+\sigma_{k+1}=k-1 \\ \mu_j \geq 1, \sigma_j \geq 0}} \tilde{S_0}^{(\sigma_1)} \tilde{A}_{\mu_1} \tilde{S_0}^{(\sigma_2)} \ldots \tilde{A}_{\mu_k} \tilde{S_0}^{(\sigma_{k+1})} \tag{3.5}$$

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 3.5 has norm at most $(1/\varepsilon)^{i-1}$.

Notice that the number of terms in the summation in Equation 3.5 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

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

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 3.4 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 $(7.5/\varepsilon)t^2$. ∎

3.4 A randomness-efficient sampler for matrix-valued functions

Here we use Theorem 3.6 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 3.6 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 3.6 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 3.12. *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*

$$\Pr_{w \xleftarrow{R} \{0,1\}^r} \left[\left\| \frac{1}{k} \sum_{i=1}^{k} f(\sigma(w)_i) \right\| \leq \gamma \right] \geq 1 - 2de^{-\gamma^2 k/70} \tag{3.6}$$

Proof of Proposition 3.12. 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. Fix 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 $p_G(y) = \Pr_{v \xleftarrow{R} V}[\varphi(v) = y]$.

Claim 3.13. $\|\mathbb{E}_{p_G}[f]\| \leq \gamma/40$

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 3.6, Claim 3.13, and Remark 2.10 that

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

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 3.13 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 3.12. ∎

Proof of Claim 3.13. The only thing remaining is to show that the $G = (V, E)$ we chose is large enough to satisfy

$$\begin{aligned} \|\mathbb{E}_{p_G}[f]\| &= \left\| \sum_{y \in [n]} (p_G(y)) f(y) \right\| = \left\| \sum_{y \in [n]} ((p_G(y) - p(y)) f(y) + \mathbb{E}_p[f(y)]) \right\| \\ &= \left\| \sum_{y \in [n]} (p_G(y) - p(y)) f(y) \right\| \\ &\leq \gamma/40 \end{aligned}$$

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 3.14. *Suppose we are in the setting of Proposition 3.12. Then there exists $k = O(\log d)$ and a $n \cdot \mathrm{poly}(d/\gamma)$ algorithm (in fact an* **NC** *algorithm) to find a sample $T = (\sigma_1, \ldots, \sigma_k)$ such that $\|\frac{1}{k}\sum_{i=1}^{k} f(\sigma_i)\| \leq \gamma$.*

Proof. Take the smallest integer $k > \frac{70}{\gamma^2}(\log d + \log 2)$, then we have that the RHS of Equation 3.6 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 $\|\frac{1}{k}\sum_{i=1}^{k} f(\sigma(w_0)_i)\| \leq \gamma$. Let $T = \sigma(w_0)$. ■

Remark 3.15. We note that the f in Proposition 3.12 and Corollary 3.14 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. In Section 4.3 we define quantum hypergraphs and show an efficient algorithm for the quantum hypergraph cover problem for quantum hypergraphs with constant size fractional covers.

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 [LR04, LS04] 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 [LR04, LS04].

Theorem 4.1 ([AR94, LS04, LR04]). *Fix $\beta < 1$ and $q < 1$.*[8] *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

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

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

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 3.14

Proof of Theorem 1.2. We wish to apply Corollary 3.14. We identify H with $[|H|]$ and let p be the uniform distribution over $[|H|]$. We apply Corollary 3.14 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 $\lambda_2(X(H;T \sqcup T^{-1})) \leq \beta$. ∎

4.2 Improved Affine Homomorphism Testers

Theorem 1.2 answers a question about the derandomization of homomorphism testers posed in [SW04]. 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: $d(f,g) = \Pr[f(x) \neq g(x)]$.

[SW04] 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. [SW04] 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 ([SW04]). *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 [SW04] the deterministic construction of S gave a set of size $|H|^{\varepsilon}$ for arbitrary $\varepsilon > 0$. The explicit construction given in [SW04] 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 [GS02] 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.

4.3 Covers of Hypergraphs and Quantum Hypergraphs

In this section we define hypergraphs and quantum hypergraphs and discuss the cover problem for both. The quantum hypergraph cover problem is a generalization of set cover arising in quantum information theory [AW02]. We apply our randomness-efficient sampler to give an algorithm to find a quantum hypergraph cover that is optimal up to logarithmic factors. This algorithm can be derandomized to run in deterministic polynomial time for quantum hypergraphs of constant cover size.

4.3.1 Hypergraphs

A hypergraph is a pair (V, E) where $E \subseteq 2^V$, i.e. E is a collection of subsets of V. Say $|V| = d$. One often views an edge e as a vector in $\{0,1\}^d$, where the i'th entry is 1 if vertex i is in the edge and 0 otherwise.

It will actually be convenient for us to view $e \in E$ as $d \times d$ diagonal matrix with 1 or 0 at each diagonal entry to signify whether that vertex is in the edge. In this section we will denote the matrix associated with e as M_e. This representation will naturally generalize to quantum hypergraphs.

A *cover* of a hypergraph $\Gamma = (V, E)$ is a set of edges C such that $\bigcup_{e \in C} e = V$, i.e. each vertex is in at least one edge. Note that this definition of cover coincides exactly with the definition of set cover. The size of the smallest cover is called the *cover number* and dentoted $c(\Gamma)$.

Using the matrix representation of E, one sees that

$$\bigcup_{e \in C} e = V \quad \Leftrightarrow \quad \sum_{e \in C} M_e \geq I$$

where the second expression uses our usual ordering of matrices.

A *fractional cover* is a set of non-negative weights w over E such that $\sum_{e \in E} w(e) M_e \geq I$. Likewise, we say that the *fractional cover number*

$$\tilde{c}(\Gamma) = \min_{w} \left\{ \sum_{e \in E} w(e) \middle| \sum_{e \in E} w(e) M_e \geq I \right\}$$

Clearly fractional cover is a LP relaxation of the cover,[9] and it is well-known that the integrality gap is $\log |V|$ and can be achieved efficiently[10] This is stated formally in the following theorem.

Theorem 4.3 (see e.g. [CLRS01, p. 1035]). *One can find a cover of a hypergraph $\Gamma = (V, E)$ of size $\tilde{c}(\Gamma) \log |V|$ in polynomial time.*

Proof. This may be done by a greedy algorithm. Simply take the edge that covers the most vertices, remove the vertices covered from consideration, take the edge that covers the most of the remaining vertices, and so on. This gives the desired approximation. ∎

The following weaker alternative method will be more useful for our generalization. We can use an LP solver to efficiently find the fractional cover, which as we mentioned is a LP. This solution is a set of non-negative weights, which we may normalize to be a probability distribution. Then we may use the AKS-sampler of [AKS87, Gil93] to sample according to this distribution which will give us a cover with high probability. Derandomizing this sampling will give us a cover with logarithmic integrality gap in time $|V|^{O(\tilde{c}(\Gamma)^2)}$, which is polynomial if $\tilde{c}(\Gamma)$ is constant. This is the idea we will generalize in the next section.

4.3.2 Quantum Hypergraphs

[AW02] defines *quantum hypergraphs* as generalizations of hypergraphs. Recall that we represented an edge of a hypergraph as a $d \times d$ diagonal matrix with $1, 0$ along the diagonal. So a hypergraph is equivalent to $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a d-dimensional complex Hilbert space, identified say with $\mathbb{C}^d$, $\mathcal{E}$ is a set of projections on V of the following kind. The vertex set V is identified with an orthonormal basis of $\mathcal{V}$. Each edge $e \in \mathcal{E}$ is identified with a projection $M_e \in \mathcal{E}$ onto the space spanned by the basis vectors corresponding to the vertices $v \in e$.

We generalize this to non-commutative "edges" by allowing $\mathcal{E}$ to contain other operators, i.e. M_e can be any Hermitian operator (i.e. matrix) in $[0, I]$.

Definition 4.4. $\Gamma = (\mathcal{V}, \mathcal{E})$ is a quantum hypergraph where $\mathcal{V}$ is a d-dimensional Hilbert space and $\mathcal{E}$ is a finite set such that each $e \in \mathcal{E}$ is identified with a Hermitian operator $M_e \in [0, I_d]$.

One can extend the definition of a cover of a quantum hypergraph $\Gamma = (\mathcal{V}, \mathcal{E})$ to be a finite subset $C \subseteq \mathcal{E}$ such that $\sum_{e \in C} M_e \geq I$. The cover number $c(\Gamma)$ is the size of the smallest cover of Γ.

[9] This is a linear program, since the M_e are diagonal.

[10] One can show a lower bound for the integrality gap using the Hadamard matrix. Consider the hypergraph on d nodes with d edges, such that the i'th edge is associated with the diagonal matrix whose j'th diagonal entry is the element (i, j) of the Hadamard matrix. That is, each edge is associated with a matrix that has a row of the Hadamard matrix along the diagonal. The Hadamard matrix is discussed in more detail in Section B.1. It is easy to show that the LP solution has weight 1 but the integral solution has weight $\log d$.

Likewise, we define a fractional cover to be a non-negative combination w of $e \in \mathcal{E}$ such that $\sum_{e\in\mathcal{E}} w(e)M_e \geq I$, and the fractional cover number as

$$\tilde{c}(\Gamma) = \min_w \left\{ \sum_{e\in\mathcal{E}} w(e) \middle| \sum_{e\in\mathcal{E}} w(e)M_e \geq I \right\}$$

Note that this corresponds exactly with our previous definitions for hypergraphs. The problem of finding the fractional cover has equivalent forms that are natural and interesting, which are discussed at the end of this section.

It is important to note that the notion of "vertex" is lost because the matrices $M_e \in \mathcal{E}$ are not necessarily diagonal in a common basis. This is why the greedy algorithm of the previous section does not extend in any obvious way.

However, in the following theorem we show that the sampler of Proposition 3.12 allows us to efficiently find a cover from a fractional cover provided certain conditions are met. Theorem 24 of [AW02] showed that the integrality gap of the SDP relaxation of fractional cover of quantum hypergraphs is $\log d$, and our derandomization gives a poly-time deterministic way of finding a cover provided that $\tilde{c}(\Gamma)$ is constant. It is an interesting problem whether one can efficiently find such a cover when $\tilde{c}(\Gamma)$ is super-constant.

Theorem 4.5. *Let $\Gamma = (\mathcal{V}, \mathcal{E})$ be a quantum hypergraph with cover number $\tilde{c}(\Gamma)$, with $d = |\mathcal{V}|$. Then one can find a cover of Γ of size $k = \tilde{c}(\Gamma)^2 O(\log d)$ in time $d^{O(\tilde{c}(\Gamma)^2)}$ (which is polynomial in d if $\tilde{c}(\Gamma)$ is constant).*

Proof of Theorem 4.5. Notice that the fractional cover optimization problem for hypergraphs is not a linear program but a *semi-definite program* (SDP's of this form are discussed in [VB96]). SDP's are also solvable in polynomial time ([Sho77, Sho87, YN77], for a survey see [VB96]), and we will show here how to find a cover using a derandomized sampler from a fractional cover recovered by the SDP.

We begin by solving the following SDP efficiently

$$\begin{array}{rc} \min : & \sum_{e\in\mathcal{E}} w(e) \\ \text{constraints} : & \sum_{e\in\mathcal{E}} w(e)M_e \geq I \quad \forall e,\ w(e) \geq 0 \end{array}$$

Solving this SDP (e.g. using the interior point method) gives us the fractional cover number $\tilde{c}(\Gamma)$ to an arbitrary accuracy and, by normalizing w, a probability distribution p on the edges, i.e. $p(e) = \frac{w(e)}{\tilde{c}(\Gamma)}$.

Suppose we have p and $\tilde{c}(\Gamma)$. The definition of $\tilde{c}(\Gamma)$ above says exactly that $\mathbb{E}_p[M_e] \geq \frac{1}{\tilde{c}(\Gamma)} I$. Thus we apply[11] Proposition 3.12 to the the distribution p on $\mathcal{E}$ (which we identify with $[|\mathcal{E}|]$). Let $A = \mathbb{E}_p[M_e]$. Then, using the resulting sampler σ we can get a sample $T \subseteq \mathcal{E}$ of size k such that

$$\begin{aligned} \Pr\left[\sum_{e\in T} M_e \geq I\right] &= \Pr\left[\sum_{e\in T}(M_e - A) \geq I - kA\right] \\ &\geq \Pr\left[\tfrac{1}{k}\sum_{e\in T}(M_e - A) \geq \left(\frac{1}{k} - \frac{1}{\tilde{c}(\Gamma)}\right) I\right] \end{aligned}$$

[11] One may apply the derandomization of Corollary 3.14 directly but this is slightly delicate and we choose to state the full proof for clarity.

We will choose $k \geq 2\tilde{c}(\Gamma)$. Since $M_e - A$ is a random variable in $[-I, I]$ with mean 0, Proposition 3.12 tells us that

$$\Pr\left[\frac{1}{k}\sum_{e\in T}(M_e - A) \geq -\frac{1}{2\tilde{c}(\Gamma)}I\right] \geq \Pr\left[\left\|\frac{1}{k}\sum_{e\in T}(M_e - A)\right\| \leq \frac{1}{2\tilde{c}(\Gamma)}\right] \geq 1 - 2de^{-(\frac{1}{2\tilde{c}(\Gamma)})^2 k/70}$$

which we want to be non-zero, hence $k \geq 70(2\tilde{c}(\Gamma))^2(\log 2d))$ confirming our previous assumption. Enumerating over all walks from the sampler gives us a deterministic algorithm to find a cover in time $d^{O(\tilde{c}(\Gamma)^2)}$, which is polynomial in $|\Gamma|$, as long as $\tilde{c}(\Gamma) = O(1)$. ■

It is interesting to note that the fractional cover SDP here is exactly the same (up to shifting) as two other natural problems from quantum information theory. Given a set $M_1, \ldots, M_n$ of Hermitian matrices, one may want to find a probability distribution p over $[n]$ one may want to solve either of the following

1. $\min_p \|\mathbb{E}_p[M_i]\|$, where i is drawn according to p
2. $\max_p \lambda_{\min}(\mathbb{E}_p[M_i])$ where i is drawn according to p and $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue.

The former minimizes the norm of the expected value of the distribution, which is also its largest eigenvalue in absolute value, while the latter may be viewed as maximizing the lowest energy state of a quantum system, which is also its smallest eigenvalue. In both cases, our sampler from Proposition 3.12 gives an integral solution using p that is worse by at most $\log d$. It can be derandomized in polynomial time when the corresponding eigenvalue is constant.

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

[AW02] R. Ahlswede and A. Winter. Strong Converse for Identification via Quantum Channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[AKS87] 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.

[Ald87] D. Aldous. On the Markov Chain Simulation Method for Uniform Combinatorial Distributions and Simulated Annealing. *Probab. Engrg. Inform. Sci.*, 1:33–46, 1987.

[AFWZ95] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized Graph Products. *Computational Complexity*, 5(1):60–75, 1995.

[AR94] N. Alon and Y. Roichman. Random Cayley Graphs and Expanders. *RSA: Random Structures & Algorithms*, 5, 1994.

[ALM+98] 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.

[Bau84] H. Baumgärtel. *Analytic Perturbation Theory for Matrices and Operators*, volume 15 of *Operator Theory: Advances and Applications*. Birkhuser, 1984.

[BH04] Y. Bilu and S. Hoory. Hypergraph Codes. *European Journal of Combinatorics*, 25(3):339–354, 2004.

[Che52] 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.

[CW89] A. Cohen and A. Wigderson. Dispersers, Deterministic Amplification, and Weak Random Sources. In *Proc.* 30*th FOCS*, pages 14–19. IEEE, 1989.

[CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms.* MIT Press, 2001.

[GG81] O. Gabber and Z. Galil. Explicit Constructions of Linear-Sized Superconcentrators. *J. Comput. Syst. Sci.*, 22(3):407–420, June 1981.

[Gil93] D. Gillman. A Chernoff bound for random walks on expander graphs. In *IEEE Symposium on Foundations of Computer Science*, pages 680–691, 1993.

[Gol65] S. Golden. Lower Bounds for the Helmholtz Function. *Physical Review*, 137B(4):B1127–1128, 1965.

[Gol97] O. Goldreich. A Sample of Samplers - A Computational Perspective on Sampling (survey). *Electronic Colloquium on Computational Complexity (ECCC)*, 4(020), 1997.

[GIL+90] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security Preserving Amplification of Hardness. In *Proc.* 31*st FOCS*, pages 318–326. IEEE, 1990.

[GS02] 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.

[HLSW04] P. Hayden, D. Leung, P. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[IZ89] R. Impagliazzo and D. Zuckerman. How to Recycle Random Bits. In *Proc.* 30*th FOCS*, pages 248–253. IEEE, 1989.

[Kah95] N. Kahale. Eigenvalues and Expansion of Regular Graphs. *Journal of the ACM*, 42(5):1091–1106, Sept. 1995.

[Kat80] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, 1980.

[LR04] 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.

[Lez98] P. Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, 8(3):849–867, 1998.

[LS04] P.-S. Loh and L. J. Schulman. Improved Expansion of Random Cayley Graphs. *Discrete Mathematics and Theoretical Computer Science*, 6(2):523–528, 2004.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Mar88] 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.

[NN93] J. Naor and M. Naor. Small-Bias Probability Spaces: Efficient Constructions and Applications. *SIAM J. Comput.*, 22(4):838–856, Aug. 1993.

[NT99] N. Nisan and A. Ta-Shma. Extracting Randomness: A Survey and New Constructions. *J. Comput. Syst. Sci.*, 58(1):148–173, 1999.

[RVW00] 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.

[Sha02] 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.

[Sho77] N. Z. Shor. Cut-off method with space extension in convex programming problems. *Cybernetics*, 13:94–96, 1977.

[Sho87] N. Z. Shor. Quadratic optimization problems. *Soviet Journal of Circuits and Systems Sciences*, 25:1–11, 1987.

[SW04] A. Shpilka and A. Wigderson. Derandomizing homomorphism testing in general groups. In *Proc. 36th STOC*, pages 427–435. ACM, 2004.

[Spi95] D. Spielman. *Computationally Efficient Error-Correcting Codes and Holographic Proofs*. PhD thesis, M.I.T., 1995.

[Tho65] C. J. Thompson. Inequality with Applications in Statistical Mechanics. *Journal of Mathematical Physics*, 6(11):1812–1823, 1965.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite Programming. *SIAM Review*, 38:49–95, March 1996.

[YN77] D. B. Yudin and A. S. Nemirovski. Informational complexity and efficient methos for solving complex extremal problems. *Matekon*, 13:25–45, 1977.

[Zuc05] D. Zuckerman. Linear Degree Extractors and the Inapproximability of Max Clique and Chromatic Number. ECCC Report TR05-100, 2005.

A Simple proof of a weaker expander walk bound

Here we demonstrate a simple proof of a weak version of Theorem 3.6 using ideas inspired by the proof of the Zig-Zag Theorem [RVW00] and without reference to perturbation theory. In particular, we can use the decomposition of the eigenvectors of $\tilde{D}_t\tilde{A}$ into uniform and orthogonal-to-uniform parts. For constant γ, ε this gives, as with our stronger Theorem 3.6, an exponential bound of $2^{-\Omega(k)}$.

Theorem A.1. *Suppose we are in the setting of Theorem 3.6 with the additional constraints that $0 < \gamma < 1/2$ and $k > \frac{2}{\varepsilon}\log\frac{4}{\gamma\varepsilon}$. Then we have*

$$\Pr[\tfrac{1}{k}f(W) \not\leq \gamma I] \leq d\left(\frac{2\log\frac{4}{\gamma\varepsilon}}{\varepsilon}\right)e^{-\frac{\gamma^2\varepsilon^3 k}{12(2\log\frac{4}{\gamma\varepsilon})^3}}$$

The proof of this statement will come in two steps. First we prove an even weaker bound with further restrictions on ε. Then we "boost" it to get the bound above.

A.1 A weak bound

Here we present a proof of a weak version of Theorem 3.6 using ideas from the proof of the Zig-Zag product [RVW00]. The idea is to decompose vectors into uniform and orthogonal-to-uniform components, and then maximize over all combinations of such decompositions.

Proposition A.2. *Suppose we are in the setting of Theorem 3.6 with the additional constraints that $2\lambda_2 < \gamma < 1/2$. Then we have:*

$$\Pr[\tfrac{1}{k}f(W) \not\leq \gamma I] \leq de^{-(\gamma-2\lambda_2)^2 k/6}$$

An identical bound for $\Pr[\frac{1}{k}f(W) \not\geq -\gamma I]$ follows similarly.

Proof. To prove this, we follow the proof of Theorem 3.6 until Equation 3.2. As before we wish to bound $\|(\tilde{D}_t\tilde{A})\|$, except here we will achieve the following weaker bound.

Lemma A.3. *For $t < \min\{\frac{1}{2}, \frac{1}{2}(\frac{1}{\lambda_2^2} - 1)\}$, we have $\|(\tilde{D}_t\tilde{A})\| \leq e^{2\lambda_2 t + \frac{3}{2}t^2}$*

We will see that the hypotheses of Lemma A.3 will be satisfied. To see how this lemma applies, we plug into Equation 3.2 and get

$$\Pr[\tfrac{1}{k}f(W) \not\leq \gamma I] \leq de^{-(\gamma-2\lambda_2)kt + \frac{3}{2}kt^2}$$

Minimizing over t gives us $t = (\gamma - 2\lambda_2)/3$ and plugging in gives us

$$\Pr[\tfrac{1}{k}f(W) \not\leq \gamma I] \leq de^{-(\gamma-2\lambda_2)^2 k/6}$$

A simple calculation shows that for all $2\lambda_2 < \gamma < 1/2$ the two conditions on t in the hypotheses of Lemma A.3 are satisfied. ∎

Proof of Lemma A.3. Our strategy follows from the proof of the Zig-Zag theorem [RVW00]. We look at the eigenvector $\tilde{x}$ of $\tilde{D}_t\tilde{A}$ with maximum eigenvalue and we decompose it into components in the $x \otimes u, \tilde{w}$ directions. Here $x \in \mathbb{R}^d$ and $u \in \mathbb{R}^n$ and $x \otimes u$ is a unit vector in the eigenspace (of dimension d) of the eigenvalue 1 of $\tilde{A}$ (i.e. $\tilde{A}(x \otimes u) = x \otimes u$). Also $\tilde{w} \in \mathbb{R}^{dn}$ is a unit vector orthogonal to this eigenspace, i.e. $\langle x \otimes u, \tilde{w}\rangle = 0$. So we have, $\tilde{x} = a(x \otimes u) + b\tilde{w}$ where $a^2 + b^2 = 1$.

We will find it easier to compute $\|\tilde{D}_t\tilde{A}\|^2$. Thus we may write the following, where Re denotes the real part of a complex number:

$$\begin{aligned}\|\tilde{D}_t\tilde{A}\|^2 &= \langle \tilde{D}_t\tilde{A}(a(x\otimes u)+b\tilde{w}), \tilde{D}_t\tilde{A}(a(x\otimes u)+b\tilde{w})\rangle \\ &= a^2\langle \tilde{D}_t(x\otimes u), \tilde{D}_t(x\otimes u)\rangle + 2ab\mathrm{Re}\langle \tilde{D}_t(x\otimes u), \tilde{D}_t\tilde{A}\tilde{w}\rangle + b^2\langle \tilde{D}_t\tilde{A}\tilde{w}, \tilde{D}_t\tilde{A}\tilde{w}\rangle\end{aligned}$$

It is easy to check that $\tilde{D}_t$ is Hermitian and $\tilde{D}_t\tilde{D}_t = \tilde{D}_{2t}$, which we apply to get

$$\leq a^2\langle x\otimes u, \tilde{D}_{2t}(x\otimes u)\rangle + 2ab\mathrm{Re}\langle x\otimes u, \tilde{D}_{2t}\tilde{A}\tilde{w}\rangle + b^2\langle \tilde{D}_t\tilde{A}\tilde{w}, \tilde{D}_t\tilde{A}\tilde{w}\rangle$$

Using the power series expansion $\tilde{D}_t = \exp(t\tilde{\Delta})$ gives us

$$a^2\sum_{i=0}^{\infty}\frac{1}{i!}\langle x\otimes u, (2\tilde{\Delta}t)^i(x\otimes u)\rangle + 2ab\sum_{i=0}^{\infty}\frac{1}{i!}\mathrm{Re}\langle x\otimes u, (2\tilde{\Delta}t)^i\tilde{A}\tilde{w}\rangle + b^2\langle \tilde{D}_t\tilde{A}\tilde{w}, \tilde{D}_t\tilde{A}\tilde{w}\rangle$$

We use the facts that $\langle x\otimes u, \tilde{\Delta}(x\otimes u)\rangle = 0$ because $x\otimes u$ is uniform across "clouds", $\langle x\otimes u, \tilde{A}\tilde{w}\rangle = 0$, and that $\|\tilde{D}_t\| \leq e^t$. Applying these facts and Cauchy-Schwarz we get

$$a^2(e^{2t}-2t) + 2ab\lambda_2(e^{2t}-1) + \lambda_2^2e^{2t}b^2$$

Maximizing this quantity over $a^2+b^2=1$ gives us that the expression is at most

$$\begin{aligned}&\leq \frac{e^{2t}-2t+\lambda_2^2e^{2t}}{2} + \frac{1}{2}\sqrt{(2\lambda_2(e^{2t}-1))^2 + (e^{2t}-2t-\lambda_2^2e^{2t})^2} \\ &\leq \frac{e^{2t}-2t+\lambda_2^2e^{2t}}{2} + \frac{1}{2}|2\lambda_2(e^{2t}-1)| + |e^{2t}-2t-\lambda_2^2e^{2t}|\end{aligned}$$

In the following we will use the fact that $1+\alpha < e^\alpha$ for all $\alpha \in \mathbb{R}$. Since $t < \frac{1}{2}(\frac{1}{\lambda_2^2}-1)$, we have $(1-\lambda_2^2)e^{2t} - 2t > (1-\lambda_2^2)(1+2t) - 2t > 0$, and thus we get that the above is

$$< e^{2t} - 2t + \lambda_2(e^{2t}-1)$$

From $t < 1/2$ and simple calculations it follows that we have $e^{2t} - 2t < 1 + 3t^2$ and $e^{2t} - 1 < 4t$. Thus our bound becomes

$$< 1 + 4\lambda_2 t + 3t^2 < e^{4\lambda_2 t + 3t^2}$$

This bound is on $\|\tilde{D}_t\tilde{A}\|^2$, and taking the square root implies the lemma. ∎

A.2 Boosting the weak bound

First we point out the simple but useful fact that for any graph G we can consider its ℓ'th power G^ℓ. This is the graph on the same vertex set, where each vertex is connected to all vertices exactly ℓ steps away in G (with appropriate multiplicities). If the normalized adjacency matrix of G is A, then the normalized adjacency matrix of G^ℓ is the A^ℓ, and each eigenvalue λ of G corresponds to an eigenvalue of G^ℓ of λ^ℓ.

Proof of Theorem A.1. The results of the previous section can be strengthened to remove the restriction that $\gamma > 2\lambda_2$ when k is sufficiently large. When taking our walk of length k on the expander G, consider dividing this walk into ℓ samples (ℓ will be fixed later). Namely, we consider steps $1, 1+\ell, 1+2\ell, \ldots$ as one sample, $2, 2+\ell, 2+2\ell, \ldots$ as another, and so on. Each of these samples can be viewed as walking on the graph G^ℓ. Now the probability that the average of all these samples deviates by γ is bounded by the probability that at least one of them deviates by γ/ℓ. This idea was mentioned by [Gil93] who credited it to [Ald87].

Applying Proposition A.2 and the union bound, this is at most

$$\ell d e^{-(\gamma/\ell - 2\lambda_2^\ell)^2 k/(6\ell)} \tag{A.1}$$

where we assume for convenience that k is a multiple of ℓ.

Claim A.4. *If $\ell = \frac{2}{\varepsilon}\log\frac{4}{\gamma\varepsilon}$ then $\gamma/\ell - 2\lambda_2^\ell \geq \gamma/2\ell$.*

Proof. We use some simple facts about the exponential and the fact that ℓ is large. The claim is equivalent to saying

$$\ell(1-\varepsilon)^\ell \leq \gamma/4$$

The derivation is given below:

$$\begin{aligned}\ell(1-\varepsilon)^\ell &\leq \left(\frac{2\log\frac{4}{\gamma\varepsilon}}{\varepsilon}\right)(1-\varepsilon)^{\frac{2\log\frac{4}{\gamma\varepsilon}}{\varepsilon}} \leq \left(\frac{2\log\frac{4}{\gamma\varepsilon}}{\varepsilon}\right)e^{-2\log\frac{4}{\gamma\varepsilon}} \\ &\leq \tfrac{1}{\varepsilon}e^{-\log\frac{4}{\gamma\varepsilon}} \leq \gamma/4\end{aligned}$$

Here we used the simple fact that $\alpha e^{-\alpha} < e^{-\alpha/2}$ for all real α. ∎

Thus substituting this setting of ℓ and $\gamma/\ell - 2\lambda_2^\ell \geq \frac{\gamma}{2\ell}$ into Equation A.1 gives us the theorem. ∎

B Tightness

Here we show that two parameters (d and ε) in our bounds Theorem 1.1 (and Theorem 3.6) are essentially tight asymptotically.

B.1 Tightness in d

Note that the bound in Theorem 1.1 has the form $d2^{-\Omega(\gamma^2\varepsilon k)} = 2^{-\Omega(\gamma^2\varepsilon k)+O(\log d)}$. Here we show that the relationship to d cannot be improved beyond $2^{-\Omega(\gamma^2\varepsilon k)+\Theta(\log d)}$.

Consider the following simple example based on the Hadamard matrix. Suppose we are working in dimension $d = 2^{k+1}$ for some k to be fixed later. Each $z \in \{0,1\}^k$ may be viewed as a subset of $\{0,1\}^k$, namely all $i \in \{0,1\}^k$ such that $\langle i, z\rangle = 1$, where the inner product is taken viewing $\{0,1\}^k$ as $GF(2)^k$. With this idea, for each z construct the diagonal matrix $\mathrm{diag}(\langle i, z\rangle)$. Let D be a random variable that is uniform over these diagonal matrices. Note that $\mathbb{E}[D] = \frac{1}{2}I$.

It is easy to show that one needs at least $\log d = k+1$ such subsets to cover $\{0,1\}^k$. So the probability that $\bigcup_{i=1}^k Z_i \neq \{0,1\}^k$, where the $Z_i \overset{R}{\leftarrow} \{0,1\}^k$ are viewed as subsets, is 1 for

$k = \log d - 1$. To rephrase this in our matrix terminology, we have that the probability of the event above is exactly

$$\Pr\left[\sum_{i=1}^{k} D_i \not\geq I\right] = 1 \text{ (for } k = \log d - 1)$$

We put this into the form of Theorem 1.1:

$$\begin{aligned} 1 &= \Pr\left[\sum_{i=1}^{k} D_i \not\geq I\right] = \Pr\left[\sum_{i=1}^{k} D_i - \frac{k}{2} I \not\geq -k(\tfrac{1}{2} - \tfrac{1}{k})I\right] \\ &= \Pr\left[\sum_{i=1}^{k} D_i - \frac{k}{2} I \not\geq -\frac{k}{4} I\right] \end{aligned}$$

The last line holds if we take $k > 4$. Now suppose that the dependence on d in Theorem 1.1 could be improved, say so that the RHS of the above is at most $2^{-\Omega(\varepsilon k)+f(d)}$ for some function $f(d)$. By our reasoning above, for $k = \log d - 1$ it must be that

$$2^{-\Omega(\varepsilon(\log d - 1))+f(d)} \geq 1$$

Solving this inequality shows that $f(d) = \Omega(\log d)$ and hence the best bound we can hope for is $2^{-\Omega(k)+\Theta(\log d)}$.

B.2 Exponent of bound is linear in ε

The bound of Theorem 3.6 is of the form $d2^{-\Omega(\gamma^2 \varepsilon k)}$. Here we show that in general one could not hope for a sublinear dependence on ε in the exponent. We will work with the 1-dimensional case $d = 1$.

Consider a boolean hypercube $B_n = (\{0,1\}^n, E)$ where $x, y \in \{0,1\}^n$ are connected iff they differ in exactly one coordinate. It is easy to check that its second-largest eigenvalue is $1 - \frac{2}{n}$ and its spectral gap is $\varepsilon = \frac{2}{n}$. Consider a dimension cut of $\{0,1\}^n$: choose a coordinate i and put all vertices v with $v_i = 1$ in S and put the rest in $T = \{0,1\}^n \setminus S$. This defines a function $f : \{0,1\}^n \to \{-1, 1\}$, where $f(v) = 1$ for all $v \in S$ and $f(v) = -1$ for $v \notin S$. Note that $\mathbb{E}[f] = 0$ (here the image of f is one-dimensional).

The probability that the value of a two-step walk W deviates above is

$$\Pr[\tfrac{1}{2} f(W) \geq I] < e^{-\varepsilon/30}$$

by Theorem 3.6 (or Theorem 3.2). But this event is exactly the event that picking a random edge of the graph stays in one of the sets in the cut defined by f, and a direct calculation of this probability shows that it is in fact exactly $1 - \frac{\varepsilon}{2}$.

Now suppose for the sake of contradiction that in general one could get a bound like Theorem 3.6 for the two-step walk with $g(\varepsilon) = o(\varepsilon)$:

$$\Pr[\tfrac{1}{2} f(W) \geq I] < e^{-\frac{g(\varepsilon)}{30}}$$

Then by a power expansion one would have that

$$1 - \frac{\varepsilon}{2} = \Pr[\tfrac{1}{2} f(W) \geq I] < 1 - \tfrac{1}{60} g(\varepsilon) + (\tfrac{1}{7200} g(\varepsilon))^2$$

It is clear that for any $g(\varepsilon) = o(\varepsilon)$ that for ε close enough to 0 this is a contradiction, and since we take ε as close to 0 as we like using a large enough hypercube, it follows that $g(\varepsilon) = \Omega(\varepsilon)$ is the best we can hope for.

C Multiple functions

In this section we will generalize our Main Theorem 1.1 to allow for an ensemble of different functions $f_i : V \to [-I, I]$ (for $1 \leq i \leq k$). This is used by Zuckerman [Zuc05] to construct an extractor $\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that, for any constants $\delta, \varepsilon > 0$, gives parameters $d = \log n + O(1)$ and $m = \Omega(n)$ and that given any δn-source gives an output ε-close to uniform.

As before we will state our theorem in the setting of Theorem 3.6 for convenience of notation. The analogous result for the setting of Theorem 1.1 is obtained by shifting and scaling the functions f_i.

Theorem C.1. *Suppose we are given an expander graph $G = (V, E)$ whose spectrum is non-negative*[12] *and with spectral gap ε.*

For any $0 < \gamma \leq I$ and $k > 6/\gamma$ and any ensemble of functions $f_i : V \to [-I, I]$, define $W = (Y_1, \ldots, Y_k)$ to be a random walk on G of length k and let $f(W) = \sum_{i=1}^{k} f_i(Y_i)$ be the value of the walk. Then it holds that

$$\Pr[\tfrac{1}{k} f(W) \not\leq \gamma I] \leq d e^{-\gamma^2 \varepsilon k/60} \quad \text{and} \quad \Pr[\tfrac{1}{k} f(W) \not\geq -\gamma I] \leq d e^{-\gamma^2 \varepsilon k/60}$$

Proof. As before we only prove one of the bounds. It is straight-forward and entirely analogous to reasoning of proof of Theorem 3.6 to see that the bound reduces to

$$\Pr[\tfrac{1}{k} f(W) \not\leq \gamma I] \leq d e^{-\gamma k t} \left\| \prod_{i=1}^{k} \tilde{D}_t^{(i)} \tilde{A} \right\|$$

where $\tilde{D}_t^{(i)}$ is the diagonal block matrix with the j'th diagonal block being $\exp(t f_i(j))$ and $\tilde{A}$ is $A \otimes I_d$. This is the analogue of Equation 3.2.

At this point we write

$$\left\| \prod_{i=1}^{k} \tilde{D}_t^{(i)} \tilde{A} \right\| \leq \|\tilde{D}_t^{(1)}\| \cdot \|\sqrt{\tilde{A}}\| \cdot \left\| \prod_{i=2}^{k} (\sqrt{\tilde{A}} \tilde{D}_t^{(i)} \sqrt{\tilde{A}}) \right\| \cdot \|\sqrt{\tilde{A}}\|$$

$$e^t \prod_{i=2}^{k} \|\sqrt{\tilde{A}} \tilde{D}_t^{(i)} \sqrt{\tilde{A}}\|$$

Thus we need a bound on $\|\sqrt{\tilde{A}} \tilde{D}_t^{(i)} \sqrt{\tilde{A}}\|$.

Claim C.2. *For all $1 \leq i \leq k$ and any $t \leq \varepsilon/15$ we have that $\|\sqrt{\tilde{A}} \tilde{D}_t^{(i)} \sqrt{\tilde{A}}\| \leq 1 + (7.5/\varepsilon) t^2$.*

Proof sketch of claim. We will only sketch the ideas since the details are almost identical to the proof of the Main Lemma 3.9. Define our perturbation $\tilde{A}(t) = \sqrt{\tilde{A}} \tilde{D}_t^{(i)} \sqrt{\tilde{A}}$. It is clear

[12] This is so that the square root of the adjacency matrix is Hermitian. One can obtain such a graph from any expander either by squaring the expander, or by adding self-loops.

that $\tilde{A}(0) = \tilde{A}$. Note that since A has a non-negative spectrum, $\sqrt{\tilde{A}}$ is Hermitian and so is $\sqrt{\tilde{A}}\tilde{D}_t^{(i)}\sqrt{\tilde{A}}$ and so it suffices to find the largest eigenvalue of $\sqrt{\tilde{A}}\tilde{D}_t^{(i)}\sqrt{\tilde{A}}$. Also, the j'th term $\tilde{A}_j$ in the power series expansion of $\sqrt{\tilde{A}}\tilde{D}_t^{(i)}\sqrt{\tilde{A}}$ is given by $\frac{1}{j!}\sqrt{\tilde{A}}(\tilde{\Delta}^{(i)})^j\sqrt{\tilde{A}}$. Therefore it also holds that $\|\tilde{A}_j\| \leq 1/2^{j-1}$.

Therefore there exists a projection-valued power series $\tilde{P}(t) = \sum_{j=1}^{\infty} t^j \tilde{P}_j$ that is convergent for $t \leq \varepsilon/15$ that projects onto the eigenspace of eigenvalues splitting from the largest eigenvalue of $\tilde{A}(t)$. Thus $\|\tilde{A}(t)\| = \|\tilde{A}(t)\tilde{P}(t)\|$ and we have

$$\|\tilde{A}(t)\| \leq 1 + \|(\tilde{A}(t) - I)\tilde{P}(t)\| \leq \sum_{j=1}^{\infty} t^j \|\tilde{Z}^{(j)}\|$$

by Theorem 2.8. It is easy to verify that $\tilde{Z}^{(1)} = 0$ by direct computation, and it also holds that $\|\tilde{Z}^{(j)}\| \leq (5/\varepsilon)^{j-1}$ for $j \geq 2$ by the same reasoning as in the proof of Claim 3.11, i.e. $\|\tilde{A}\| \leq 1$ and $\|\tilde{\Delta}^{(i)}\| \leq 1$.

Thus it follows since $t \leq \varepsilon/15$ that

$$\|\sqrt{\tilde{A}}\tilde{D}_t^{(i)}\sqrt{\tilde{A}}\| \leq \|\tilde{A}(t)\| \leq 1 + (7.5/\varepsilon)t^2$$

■

We apply this to the bound and get that

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

Choosing $t = \gamma\varepsilon/15$ and applying the fact that $k > 6/\gamma$ gives us that

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

■

Remark C.3. One can do away with the assumption that G has a non-negative spectrum by using the simple proof of Theorem A.1, where we do not need to take square roots of the adjacency matrix in order to bound $\|\tilde{D}_t^{(i)}\tilde{A}\|$. Of course this leads to an inferior bound. The proof is straight-forward and left to the reader.