

# Derandomizing the AW matrix-valued Chernoff bound using pessimistic estimators and applications

Avi Wigderson

David Xiao

August 23, 2006

## Abstract

Ahlswede and Winter [AW02] introduced a Chernoff bound for matrix-valued random variables, which is a non-trivial generalization of the usual Chernoff bound for real-valued random variables. We present an efficient derandomization of their bound using the method of pessimistic estimators (see Raghavan [Rag88]). As a consequence, we derandomize a construction of Alon and Roichman [AR94] (see also [LR04, LS04]) to efficiently construct an expanding Cayley graph of logarithmic degree on any (possibly non-abelian) group. This also gives an optimal solution to the homomorphism testing problem of Shpilka and Wigderson [SW04]. We also apply these pessimistic estimators to the problem of solving semi-definite covering problems, thus giving a deterministic algorithm for the quantum hypergraph cover problem of [AW02].

The results above appear as theorems in the paper [WX05a] (see also [WX05b]), as consequences to the main theorem of that paper: a randomness efficient sampler for matrix valued functions via expander walks. However, we discovered an error in the proof of that main theorem (which we briefly describe in the appendix). That main theorem stating that the expander walk sampler is good for matrix-valued functions thus remains open. One purpose of the current paper is to show that the applications in that paper hold true despite our inability to prove the expander walk sampler theorem for matrix-valued functions.

## 1 Introduction

Chernoff bounds are extremely useful throughout theoretical computer science. Intuitively, they say that a random sample approximates the average, with probability of deviation that goes down exponentially with the number of samples. Typically we are concerned with real-valued random variables, but recently several applications have called for matrix-valued random variables. Such a bound was given by Ahlswede and Winter [AW02].

In particular, the matrix-valued bound seems useful in giving new proofs of probabilistic constructions of expander graphs [AR94] and also in the randomized rounding of semi-definite covering problems, with further applications in quantum information theory [AW02].

In this paper we use the method of pessimistic estimators, originally formulated in [Rag88]<sup>1</sup>, to derandomize the Chernoff bound of [AW02], and in the process derandomize the Alon-Roichman theorem and the randomized rounding of covering SDP's.

The results of this paper prove the claimed applications of our previous paper [WX05a], and in fact supersede them in simplicity and efficiency. We regret to inform the community that

---

<sup>1</sup>The simpler method of conditional probabilities was described earlier in the first edition of [Spe94].

we discovered a fatal mistake in the analysis of using an expander sampler in [WX05a], and it remains open whether the expander sampler achieves the deviation bound claimed there (or something asymptotically equivalent).

For details on the problem with the previous work, see [Appendix A](#).

Kale [Kal] independently reached results similar to the ones presented in this paper that imply the applications to constructing expanding Cayley graphs and semi-definite covering programs.

## 2 Matrix-valued random variables and Alhswede-Winter's Chernoff Bound

We will work with real symmetric  $d \times d$  matrices, which we will denote  $\mathcal{M}_d$ .<sup>2</sup> We let  $I_d$  denote the identity matrix in  $\mathcal{M}_d$ , and will write simply  $I$  when the dimension is clear. For any  $A \in \mathcal{M}_d$  we let  $\lambda_1(A) \geq \dots \geq \lambda_d(A)$  denote the eigenvalues of  $A$  in non-increasing order. Recall that every matrix  $A \in \mathcal{M}_d$  is diagonalizable in an orthonormal basis.

We will measure distance between matrices in the operator norm  $\|A\| = \max_i |\lambda_i(A)|$ . We will also frequently use the trace,  $\text{Tr}(A) = \sum_{i=1}^d \lambda_i(A)$ . It is well-known that for any orthonormal basis  $v_1, \dots, v_d \in \mathbb{R}^d$  we have that  $\text{Tr}(A) = \sum_{i=1}^d \langle v_i, Av_i \rangle$ , where  $\langle \cdot, \cdot \rangle$  denotes the usual inner product over  $\mathbb{R}^d$ .

We say that a matrix  $A \in \mathcal{M}_d$  is positive semi-definite (p.s.d.) if all its eigenvalues are non-negative. We will use the fact that  $A$  is p.s.d. iff for all  $v \in \mathbb{R}^d$ ,  $\langle v, Av \rangle \geq 0$ . We let  $A \geq 0$  denote that  $A$  is p.s.d. We use the ordering of symmetric matrices given by this definition, namely  $A \leq B$  iff  $B - A \geq 0$ . For two matrices  $A \leq B$ , we will let  $[A, B]$  denote the set of all matrices  $C$  such that  $A \leq C$  and  $C \leq B$ .

We will work with the matrix exponential, which is defined by

$$\exp(A) = \sum_{\ell=0}^{\infty} \frac{A^\ell}{\ell!}$$

Recall that the matrix exponential is convergent for all matrices. Furthermore, it is not hard to see for  $A \in \mathcal{M}_d$  that  $\exp(A)$  is diagonalizable in the same basis as  $A$ , and that  $\lambda_i(\exp(A)) = e^{\lambda_i(A)}$  for all  $1 \leq i \leq d$ . Also, for all  $A \in \mathcal{M}_d$ , it holds that  $\exp(A) \geq 0$ .

We will consider matrix-valued random variables of the following form. We let  $f : [n] \rightarrow [-I_d, I_d]$ , where  $[n] = \{1, \dots, n\}$ . Let  $X$  be a distribution (not necessarily uniform) over  $[n]$ , and consider the variable  $f(X)$ . This is a natural extension of bounded random variables over the reals, which may be thought of as functions  $f : [n] \rightarrow [-1, 1]$ . We will let the expectation of  $f(X)$  be the obvious thing:  $\mathbb{E}[f(X)] = \sum_{i=1}^n \Pr[X = i]f(i)$ . Note that because  $\text{Tr}$  is linear that  $\mathbb{E}$  and  $\text{Tr}$  commute:  $\mathbb{E}[\text{Tr}(f(X))] = \text{Tr}(\mathbb{E}[f(X)])$ . We let  $\text{supp}(X)$  denote the set of all values of  $X$  that occur with non-zero probability. When we say that something holds for a random variable  $X$  *always*, we mean that it holds for every element in  $\text{supp}(X)$ .

We will use the following useful facts several times:

**Fact 2.1.** *If  $A, B \in \mathcal{M}_d$  and  $B \geq 0$ , then  $\text{Tr}(AB) \leq \|A\|\text{Tr}(B)$ .*

---

<sup>2</sup>All our results extend to complex Hermitian matrices, or abstractly to self-adjoint operators over any vector space where the operations of addition, multiplication, trace, exponential, and norm are efficiently computable.

*Proof.* Let  $v_1, \dots, v_d$  be the orthonormal diagonal basis of  $A$ , with corresponding eigenvalues  $\lambda_i = \lambda_i(A)$ . Then we may write

$$\begin{aligned}\mathrm{Tr}(AB) &= \sum_{i=1}^d \langle v_i, ABv_i \rangle \\ &= \sum_{i=1}^d \lambda_i \langle v_i, Bv_i \rangle\end{aligned}$$

Since  $B \geq 0$  we know that  $\langle v_i, Bv_i \rangle \geq 0$ , so we get

$$\begin{aligned}&\leq \sum_{i=1}^d \max_i \lambda_i \langle v_i, Bv_i \rangle \\ &\leq \|A\| \mathrm{Tr}(B)\end{aligned}$$

■

**Theorem 2.2** (Golden-Thompson inequality, [Gol65, Tho65]). *For  $A, B \in \mathcal{M}_d$ , we have*

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

The proof of this is outside the scope of this paper.

Ahlsvede and Winter introduce a generalization of Markov's inequality for matrix-valued random variables.

**Theorem 2.3** (Markov's inequality [AW02]). *For  $g : [n] \rightarrow \mathcal{M}_d$ ,  $g(x) \geq 0$  for all  $x \in [n]$ , and for any random variable  $X$  over  $[n]$ , we have*

$$\Pr[g(X) \not\leq \gamma I] \leq \frac{1}{\gamma} \mathrm{Tr}(\mathbb{E}[g(X)])$$

*Proof.*

$$\begin{aligned}\Pr[g(X) \not\leq \gamma I] &= \Pr[\|g(X)\| > \gamma] \\ &\leq \frac{1}{\gamma} \mathbb{E}[\|g(X)\|]\end{aligned}$$

Since  $g(X) \geq 0$  always, we have  $\|g(X)\| \leq \mathrm{Tr}(g(X))$  always, so we get:

$$\begin{aligned}&\leq \frac{1}{\gamma} \mathbb{E}[\mathrm{Tr}(g(X))] \\ &= \frac{1}{\gamma} \mathrm{Tr}(\mathbb{E}[g(X)])\end{aligned}$$

■

The following [Theorem 2.4](#) is the main theorem proving [\[AW02\]](#)'s Chernoff-type bound. We will use [Theorem 2.4](#), which holds for all distributions, to derive two corollaries ([Theorem 2.6](#) and [Theorem 2.8](#)), which hold for more specific kinds of distributions. In addition, the proof of [Theorem 2.4](#) will give us the pessimistic estimators corresponding to the two corollaries.

**Theorem 2.4** ([AW02]). *Suppose  $f : [n] \rightarrow [-I_d, I_d]$  and let  $X_1, \dots, X_k$  be arbitrary independent random variables distributed over  $[n]$ . Then for all  $\gamma \in \mathbb{R}$ :*

$$\Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\leq \gamma I\right] \leq de^{-t\gamma k} \prod_{j=1}^k \|\mathbb{E}[\exp(tf(X_j))]\|$$

*Proof.* The proof begins analogously to the real-valued case, generalizing the classical Bernstein trick. We first multiply by an optimization constant  $t > 0$  and exponentiate to obtain

$$\Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\leq \gamma I\right] = \Pr\left[\exp\left(t \sum_{j=1}^k f(X_j)\right) \not\leq e^{t\gamma k} I\right]$$

The equality holds because for  $A, B \in \mathcal{M}_d$  such that  $AB = BA$ , we have that  $A \leq B$  is equivalent to  $\exp(A) \leq \exp(B)$ . Then the following inequality is a direct consequence of [Theorem 2.3](#) since  $\exp(A) \geq 0$  for all  $A \in \mathcal{M}_d$ .

$$\Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\leq \gamma I\right] \leq e^{-t\gamma k} \text{Tr}(\mathbb{E}[\exp(t \sum_{j=1}^k f(X_j))]) \quad (2.1)$$

Then we apply [Fact 2.1](#) and the Golden-Thompson Inequality [Theorem 2.2](#) to bound the expression in a manageable form. This step will be expressed in the following lemma.

**Lemma 2.5.** *For any matrix  $A \in \mathcal{M}_d$ , any  $f : [n] \rightarrow \mathcal{M}_d$  and any random variable  $X$  over  $[n]$ , we have*

$$\text{Tr}(\mathbb{E}_X[\exp(A + f(X))]) \leq \|\mathbb{E}[\exp(f(X))]\| \cdot \text{Tr}(\exp(A))$$

To obtain [Theorem 2.4](#), we simply apply [Lemma 2.5](#) to [Inequality 2.1](#) repeatedly:

$$\Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\leq \gamma I\right] \leq e^{-t\gamma k} \text{Tr}(\mathbb{E}[\exp(t \sum_{j=1}^k f(X_j))])$$

By independence

$$= e^{-t\gamma k} \mathbb{E}_{X_1, \dots, X_{k-1}} \left[ \text{Tr}(\mathbb{E}_{X_k}[\exp(t \sum_{j=1}^{k-1} f(X_j) + tf(X_k))]) \right]$$

Apply [Lemma 2.5](#)

$$\leq e^{-t\gamma k} \mathbb{E}_{X_1, \dots, X_{k-1}} \left[ \|\mathbb{E}[\exp(tf(X_k))]\| \cdot \text{Tr}(\exp(t \sum_{j=1}^{k-1} f(X_j))) \right]$$

Pulling the expectation back inside

$$= e^{-t\gamma k} \|\mathbb{E}[\exp(tf(X_k))]\| \cdot \mathbb{E}_{X_1, \dots, X_{k-1}} \left[ \text{Tr}(\exp(t \sum_{j=1}^{k-1} f(X_j))) \right]$$

Repeating  $k$  times ...

$$\begin{aligned} &\leq e^{-t\gamma k} \prod_{j=1}^k \|\mathbb{E}[\exp(tf(X_j))]\| \text{Tr}(I) \\ &= de^{-t\gamma k} \prod_{j=1}^k \|\mathbb{E}[\exp(tf(X_j))]\| \end{aligned}$$

This completes the proof modulo [Lemma 2.5](#). ■

*Proof of [Lemma 2.5](#).*

Since trace and expectation commute:

$$\text{Tr}(\mathbb{E}[\exp(A + f(X))]) = \mathbb{E}[\text{Tr}(\exp(A + f(X)))]$$

Applying the Golden-Thompson inequality

$$\leq \mathbb{E}[\text{Tr}(\exp(f(X)) \exp(A))]$$

Commuting trace and expectation again:

$$\leq \text{Tr}(\mathbb{E}[\exp(f(X))] \exp(A))$$

By [Fact 2.1](#)

$$\leq \|\mathbb{E}[\exp(tf(X))]\| \cdot \text{Tr}(\exp(A))$$
■

Now we will draw two corollaries from this main theorem. These two corollaries are useful in different settings; the first guarantees that the probability of an additive deviation is small, while the second a multiplicative deviation.

**Theorem 2.6** ([\[AW02\]](#)). *Let  $f : [n] \rightarrow [-I_d, I_d]$ . Let  $X$  be distributed over  $[n]$  with  $\mathbb{E}_X[f(X)] = 0$ , and let  $X_1, \dots, X_k$  be i.i.d. copies of  $X$ . Then for all  $1 > \gamma > 0$ <sup>3</sup>:*

$$\Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\leq \gamma I\right] \leq de^{-\gamma^2 k/4}$$

Note that the other direction  $\frac{1}{k} \sum_{i=1}^k f(X_i) \not\geq -\gamma I$  holds with the same bound by considering  $-f$ .

*Proof.* We require only [Theorem 2.4](#) and a simple claim. Because all the  $X_i$  are i.i.d. [Theorem 2.4](#) gives us

$$\Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\leq \gamma I\right] \leq de^{-t\gamma k} \|\mathbb{E}[\exp(tf(X))]\|^k$$

We use the following claim to bound the RHS.

---

<sup>3</sup>For the sake of simplicity, no attempt was made to optimize the constant in the exponent of the bound.

**Claim 2.7.**  $\|\mathbb{E}[\exp(tf(X))]\| \leq 1 + t^2$  for  $t \leq 1/2$ .

*Proof.* This follows from the Taylor expansion of  $\exp$ :

$$\begin{aligned} \|\mathbb{E}[\exp(tf(X))]\| &= \|\mathbb{E}[I + tf(X) + \frac{(tf(X))^2}{2} + \dots]\| \\ &= \|I + t\mathbb{E}[f(X)] + \mathbb{E}[(tf(X))^2/2 + \dots]\| \end{aligned}$$

Since  $\mathbb{E}[f(X)] = 0$ , applying the triangle inequality, and using  $\|f(X)\| \leq 1$  always, we have

$$\leq 1 + \sum_{\ell=2}^{\infty} t^\ell / \ell!$$

Since  $t = \gamma/2 \leq 1/2$  this is bounded by

$$\leq 1 + t^2$$

■

We will choose  $t = \gamma/2 \leq 1/2$ , so we may apply [Claim 2.7](#) to [Theorem 2.4](#) to get

$$\Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\leq \gamma I\right] \leq de^{-t\gamma k} (1 + t^2)^k$$

Using  $1 + x \leq e^x$  for all  $x \in \mathbb{R}$

$$\leq de^{-t\gamma k + t^2 k}$$

Choosing  $t = \gamma/2$

$$\leq de^{-\gamma^2 k/4}$$

■

**Theorem 2.8** ([\[AW02\]](#)). *Let  $f : [n] \rightarrow [0, I_d]$ . Let  $X$  be distributed over  $[n]$ , with  $M = \mathbb{E}_X[f(X)] \geq \mu I$  for some  $\mu \in (0, 1)$ . Let  $X_1, \dots, X_k$  be i.i.d. copies of  $X$ . Then we have, for all  $\gamma \in [0, 1/2]$ ,*

$$\Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\geq (1 - \gamma)\mu I\right] \leq de^{\frac{-\gamma^2 \mu}{2} k}$$

*Proof.* We can assume without loss of generality that  $M = \mu I$ .<sup>4</sup> Because the direction of this bound is the opposite of what we proved in [Theorem 2.4](#), we will work with  $I - f$  to get:

$$\Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\geq (1 - \gamma)\mu I\right] = \Pr\left[\frac{1}{k} \sum_{i=1}^k (I - f(X_i)) \not\leq (1 - (1 - \gamma)\mu)I\right] \quad (2.2)$$

---

<sup>4</sup>If not, we could work with  $g(x) = \mu M^{-1/2} f(x) M^{-1/2}$  instead

Applying [Theorem 2.4](#)

$$\leq d e^{-t(1-(1-\gamma)\mu)^k} \|\mathbb{E}[\exp(t(I - f(X)))]\|^k \quad (2.3)$$

$$= d \|\mathbb{E}[\exp(-tf(X))e^{t(1-\gamma)\mu}]\|^k \quad (2.4)$$

This last quantity was analyzed in [\[AW02\]](#) using information-theoretic techniques, with the following conclusion which we state without proof:

**Claim 2.9** ([\[AW02\]](#)). For  $t = \log(\frac{1-(1-\gamma)\mu}{1-\mu} \frac{1}{(1-\gamma)})$ , we have

$$\|\mathbb{E}[\exp(-tf(X))e^{t(1-\gamma)\mu}]\| \leq e^{-\gamma^2\mu/2}$$

Applying this claim to [Inequality 2.4](#) gives us the theorem. ■

### 3 Method of pessimistic estimators

First we review the method of pessimistic estimators, due to Raghavan [\[Rag88\]](#). The setting is the following: we have a random variable  $X$  and we know that with some non-zero probability an event  $\sigma(X)$  occurs, i.e.  $\Pr[\sigma(X) = 1] > 0$ , where  $\sigma : \text{supp}(X) \rightarrow \{0, 1\}$ ,  $\sigma(x) = 1$  iff  $x$  is in the event. We wish to efficiently and deterministically find a particular  $x \in \text{supp}(X)$  such that  $\sigma(x) = 1$ .

Our application of pessimistic estimators is to derandomizing probabilistic algorithms. In particular, suppose we have a randomized algorithm that constructs an object, and with some non-zero probability that object satisfies some property. Thus, our event  $\sigma$  is the event that the object satisfies the property, and our goal is to deterministically and efficiently find the object. In this paper our two main applications are to deterministically and efficiently find a small generating set of a group that satisfies expansion, and to find an integer solution to a SDP covering problem that satisfies feasibility and some approximation guarantee. Both problems were previously known to have randomized algorithms, and we use our pessimistic estimators to derandomize these algorithms.

We will only be concerned with random variables with finite state space with a product structure, and we will sub-divide the variable into many parts. Thus we use the notation  $\vec{X}$  to denote a random variable where *w.l.o.g.*  $\text{supp}(\vec{X}) \subseteq [n]^k$  for some  $k, n \in \mathbb{N}$  (these will be chosen according to the application). Let  $\vec{X} = (X_1, \dots, X_k)$ , where each  $X_i \in [n]$ . To find a “good” setting of  $\vec{X}$ , we will iteratively find settings of  $X_1$ , then  $X_2$ , and so forth until we have a complete setting of  $\vec{X}$ .

By the definition of expectation

$$\Pr_{\vec{X}}[\sigma(\vec{X}) = 0] = \mathbb{E}_{X_1}[\Pr[\sigma(\vec{X}) = 0 \mid X_1]]$$

Now by averaging there must exist at least one setting  $x_1 \in [n]$  of  $X_1$  such that

$$\Pr[\sigma(\vec{X}) = 0 \mid X_1 = x_1] \leq \mathbb{E}_{X_1}[\Pr[\sigma(\vec{X}) = 0 \mid X_1]]$$

We set  $X_1 = x_1$ , and then repeat the same reasoning for  $X_2, \dots, X_k$ . Let us denote the resulting setting of  $\vec{X}$  by  $\vec{x}$ . Thus at the end we have  $\Pr[\sigma(\vec{x}) = 0] \leq \Pr[\sigma(\vec{X}) = 0]$ . But note that we

supposed that  $\Pr[\sigma(\vec{X}) = 0] < 1$ , and since  $\vec{x}$  is a *fixed* vector, it must be that  $\Pr[\sigma(\vec{x}) = 0] = 0$  and therefore  $\sigma(\vec{x}) = 1$ .

The difficulty with turning this into an algorithm is in calculating the probabilities, for each  $1 \leq i \leq k$  and,  $\forall x_1, \dots, x_i \in [n]$

$$\Pr_{X_{i+1}, \dots, X_k} [\sigma(\vec{X}) = 0 \mid X_1 = x_1, \dots, X_i = x_i]$$

since they may not be efficiently computable. Fortunately we may relax the requirements slightly by the following.<sup>5</sup>

**Definition 3.1.** Let  $\sigma : [n]^k \rightarrow \{0, 1\}$  be an event on a random variable  $\vec{X}$  distributed over  $[n]$  and suppose  $\Pr[\sigma(\vec{X}) = 1] > 0$ . We say that  $\phi_0, \dots, \phi_k$ ,  $\phi_i : [n]^i \rightarrow [0, 1]$  (here  $\phi_0$  is just a number in  $[0, 1]$ ), are *pessimistic estimators* for  $\sigma$  if the following hold.

1. For any  $i$  and any fixed  $x_1, \dots, x_i \in [n]$ , we have that

$$\Pr_{X_{i+1}, \dots, X_k} [\sigma(x_1, \dots, x_i, X_{i+1}, \dots, X_k) = 0] \leq \phi_i(x_1, \dots, x_i)$$

2. For any  $i$  and any fixed  $x_1, \dots, x_i \in [n]$ :

$$\mathbb{E}_{X_{i+1}} \phi_{i+1}(x_1, \dots, x_i, X_{i+1}) \leq \phi_i(x_1, \dots, x_i)$$

We will also want the pessimistic estimators to be *efficient*, namely each  $\phi_i$  is efficiently computable, and *useful*, which means  $\phi_0 < 1$ . This last condition is because  $\phi_0$  is a bound on the initial probability of failure, which we need to be strictly less than 1.

**Theorem 3.2** ([Rag88]). *If there exist efficient and useful pessimistic estimators  $(\phi_0, \dots, \phi_k)$  for an event  $\sigma$ , then one can efficiently compute a fixed  $\vec{x} \in [n]^k$  such that  $\sigma(\vec{x}) = 1$ .*

*Proof.* We pick  $x_1, \dots, x_k$  one by one. At step 0 we have  $\phi_0 < 1$  since the estimators are useful. At step  $i$ , we have  $x_1, \dots, x_i$  already fixed. Enumerate over  $x_{i+1} \in [n]$  and choose the value such that  $\phi_{i+1}(x_1, \dots, x_{i+1}) \leq \phi_i(x_1, \dots, x_i) < 1$ . We are guaranteed that

$$\mathbb{E}_{X_{i+1}} [\phi_{i+1}(x_1, \dots, x_i, X_{i+1})] \leq \phi_i(x_1, \dots, x_i)$$

by property 2 of [Definition 3.1](#), and so by averaging there must exist a fixed  $x_{i+1} \in [n]$  that is at most the expectation on the LHS of the above inequality. We can compute the value of the estimator efficiently by hypothesis.

Finally, we have after  $k$  steps that  $\phi_k(\vec{x}) < 1$  and by property 1 we have that  $\Pr[\sigma(\vec{x}) = 0] < 1$ , and therefore  $\sigma(\vec{x}) = 1$ .

The algorithm runs through  $k$  steps, and each step is efficient, so the overall algorithm is efficient. ■

---

<sup>5</sup>Our definition is stronger than the standard definition of pessimistic estimators, in that in the second condition usually all that is required is for all  $x_1, \dots, x_i \in [n]$ , there exists  $x_{i+1} \in [n]$  such that  $\phi_{i+1}(x_1, \dots, x_{i+1}) \leq \phi_i(x_1, \dots, x_i)$ . But our estimators satisfy the stronger definition and we will find it useful, especially when composing estimators (see [Lemma 3.3](#)).

We will find it useful to compose estimators, which is possible from the following lemma.

**Lemma 3.3.** *Suppose  $\sigma, \tau : [n]^k \rightarrow \{0, 1\}$  are events on  $\vec{X}$ , which is distributed over  $[n]^k$ . Suppose that  $(\phi_0, \dots, \phi_k), (\psi_0, \dots, \psi_k)$  are pessimistic estimators for  $\sigma, \tau$  respectively. Then  $(\phi_0 + \psi_0, \dots, \phi_k + \psi_k)$  are pessimistic estimators for the event  $\sigma \cap \tau$ .*

*Proof.* We need to verify the properties of [Definition 3.1](#).

1. This is verified by a union bound:

$$\begin{aligned} \Pr[(\sigma \cap \tau)(x_1, \dots, x_i, X_{i+1}, \dots, X_k) = 0] \\ \leq \Pr[\sigma(x_1, \dots, x_i, X_{i+1}, \dots, X_k) = 0] + \Pr[\tau(x_1, \dots, x_i, X_{i+1}, \dots, X_k) = 0] \\ \leq (\phi_i + \psi_i)(x_1, \dots, x_i) \end{aligned}$$

2. This is immediate from linearity of expectation. ■

## 4 Applying pessimistic estimators to the AW bound

The method of pessimistic estimators extends to the AW Chernoff bound. We will first describe pessimistic estimators for [Theorem 2.6](#) and then for [Theorem 2.8](#). They are essentially identical except for the difference in distributions in the two settings, and the proofs that the pessimistic estimators satisfy [Definition 3.1](#) rely mainly on [Lemma 2.5](#). In both cases, they allow us to efficiently and deterministically find settings  $x_1, \dots, x_k$  such that bad event bounded by [Theorem 2.6](#) (resp. [Theorem 2.8](#)) does not occur.

**Theorem 4.1.** *Let  $f : [n] \rightarrow [-I_d, I_d]$ . Let  $X$  be distributed over  $[n]$  with  $\mathbb{E}_X[f(X_i)] = 0$ , and let  $X_1, \dots, X_k$  be i.i.d. copies of  $X$ . Fix  $1 > \gamma > 0$ . Let  $t = \gamma/2$ . Suppose that  $\mathbb{E}[\exp(tf(X))]$  is efficiently computable.*

*Combining the notation of [Section 2](#) and [Section 3](#), we let  $\vec{X} = (X_1, \dots, X_k)$  with  $X_i \in [n]$  and we let  $\sigma : [n]^k \rightarrow \{0, 1\}$  be the event  $\sigma(\vec{X}) = 1$  if  $\frac{1}{k} \sum_{i=1}^k f(X_i) \leq \gamma I$  and  $\sigma(\vec{X}) = 0$  otherwise. Then the following  $(\phi_0, \dots, \phi_k), \phi_i : [n]^i \rightarrow [0, 1]$  are efficient pessimistic estimators for  $\sigma$ .*

$$\begin{aligned} \phi_0 &= de^{-t\gamma k} \|\mathbb{E}[\exp(tf(X))]\|^k \quad (\leq de^{-\gamma^2 k/4}) \\ \phi_i(x_1, \dots, x_i) &= de^{-t\gamma k} \text{Tr}(\exp(t \sum_{j=1}^i f(x_j))) \cdot \|\mathbb{E}[\exp(tf(X))]\|^{k-i} \end{aligned}$$

*Proof.* We verify the properties of [Definition 3.1](#).

1. From [Inequality 2.1](#):

$$\begin{aligned} \Pr[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\leq \gamma I] &\leq de^{-t\gamma k} \text{Tr}(\mathbb{E}[\exp(t \sum_{j=1}^k f(X_j))]) \\ &\leq de^{-t\gamma k} \text{Tr}(\mathbb{E}[\exp(t \sum_{j=1}^i f(X_j))]) \prod_{j=i+1}^k \|\mathbb{E}[\exp(tf(X_j))]\| \end{aligned}$$

The last inequality follows from applying [Lemma 2.5](#)  $k - i$  times. Notice that the above is true even if the  $X_i$  are not i.i.d. Therefore we can consider the case where  $X_1 = x_1, \dots, X_i = x_i$  (i.e. they are constant random variables), while the remaining  $X_{i+1}, \dots, X_k$  are distributed i.i.d. according to  $X$ . This gives

$$\begin{aligned} \Pr\left[\frac{1}{k} \sum_{i=1}^k f(X_i) \not\leq \gamma I \mid X_1 = x_1, \dots, X_i = x_i\right] &\leq de^{-t\gamma k} \text{Tr}(\exp(t \sum_{j=1}^i f(x_j))) \cdot \|\mathbb{E}[\exp(tf(X))]\|^{k-i} \\ &= \phi_i(x_1, \dots, x_i) \end{aligned}$$

2. We use the following derivation, where the inequality follows from [Lemma 2.5](#):

$$\begin{aligned} &\mathbb{E}_{X_{i+1}}[\phi_{i+1}(x_1, \dots, x_i, X_{i+1})] \\ &= de^{-t\gamma k} \text{Tr}(\mathbb{E}_{X_{i+1}}[\exp(t \sum_{j=1}^i f(x_j) + tf(X_{i+1}))]) \cdot \|\mathbb{E}(\exp(tf(X)))\|^{k-i-1} \\ &\leq de^{-t\gamma k} \text{Tr}(\exp(t \sum_{j=1}^i f(x_j))) \cdot \|\mathbb{E}(\exp(tf(X)))\|^{k-i} \\ &= \phi_i(x_1, \dots, x_i) \end{aligned}$$

To see that the  $\phi_i$  are efficiently computable, we will specify the input to the algorithm as a function  $f$  (which we assume is given as a list of matrices  $d \times d$   $f(1), \dots, f(n)$  and  $1^k$ ). Thus we desire the algorithm be computable in time  $\text{poly}(n, d, k)$ . We require multiplication, addition, trace, matrix exponential, and norm computations. The first three are obviously efficient; the last two are efficient because eigenvalues of a  $d \times d$  matrix can be computed (and hence it can be diagonalized thus making the exponential and norm computations trivial) in  $O(d^3)$  numerical operations [[GL89](#)]. On a machine with finite precision, we can truncate the estimators to a sufficiently fine resolution so that the truncated estimators behave essentially as the real-valued estimators do. ■

[Theorem 4.1](#) gives us pessimistic estimators  $(\phi_0, \dots, \phi_k)$  for  $\sigma$ , and the same proof gives efficient pessimistic estimators  $(\psi_0, \dots, \psi_k)$  for the event  $\tau(\vec{X}) = 1$  iff  $\frac{1}{k} \sum_{i=1}^k f(X_i) \geq -\gamma I$  by applying [Theorem 2.6](#) to  $-f$ . Combining these with the  $\phi_i$  gives us the following.

**Corollary 4.2.** *Let  $f : [n] \rightarrow [-I_d, I_d]$ . Let  $X$  be distributed over  $[n]$  with  $\mathbb{E}_X[f(X)] = 0$ , and let  $X_1, \dots, X_k$  be i.i.d. copies of  $X$ . Fix  $1 > \gamma > 0$  and fix  $t = \gamma/2$ . Suppose that  $\mathbb{E}[\exp(tf(X))]$  and  $\mathbb{E}[\exp(-tf(X))]$  are efficiently computable.*

*Let  $\eta : [n]^k \rightarrow \{0, 1\}$  be the event  $\eta(X) = 1$  if  $\|\frac{1}{k} \sum_{i=1}^k f(X_i)\| \leq \gamma$  and  $\eta(X) = 0$  otherwise. Then  $(\phi_0 + \psi_0, \dots, \phi_k + \psi_k)$  are efficient pessimistic estimators for  $\eta$ .*

*Proof.* Note that  $\eta = \sigma \cap \tau$ . Efficiency is clear. We can apply [Lemma 3.3](#) to get that  $(\phi_0 + \psi_0, \dots, \phi_k + \psi_k)$  is a pessimistic estimator for the event  $\eta = \sigma \cap \tau$ . ■

This allows us to derandomize [Theorem 2.6](#) efficiently. Notice that in general the only property

of  $X$  that we need is to be able to compute  $\mathbb{E}[\exp(tf(X))]$  and  $\mathbb{E}[\exp(-tf(X))]$ .<sup>6</sup> This is of course true when  $X$  is uniform, or when we know  $\Pr[X = i]$  for each  $i \in [n]$ . The actual distribution is irrelevant, since we exhaustively search through the entire space for the choice of each  $X_i$ .

**Theorem 4.3.** *Let  $f : [n] \rightarrow [-I_d, I_d]$  be such that there exists a distribution  $X$  over  $[n]$  such that  $\mathbb{E}[f(X)] = 0$ . Then for  $k = O(\frac{1}{\gamma^2} \log d)$ , we can efficiently and deterministically find  $\vec{x} \in [n]^k$  such that  $\|\frac{1}{k} \sum_{i=1}^k f(x_i)\| \leq \gamma$ .*

*Proof.* Use the efficient and useful pessimistic estimators of [Corollary 4.2](#). Pick  $k = O(\frac{1}{\gamma^2} \log d)$  such that  $\phi_0 + \psi_0 < 1$  and so the estimators are useful. We may then apply [Theorem 3.2](#) to get the result.  $\blacksquare$

We can construct pessimistic estimators for [Theorem 2.8](#) in the same way.

**Theorem 4.4.** *Let  $f : [n] \rightarrow [0, I_d]$ . Let  $X$  be distributed over  $[n]$ , with  $M = \mathbb{E}_X[f(X)] \geq \mu I$  for some  $\mu \in (0, 1)$ . Let  $X_1, \dots, X_k$  be i.i.d. copies of  $X$ . Fix  $t = \log(\frac{1-(1-\gamma)\mu}{1-\mu} \frac{1}{1-\gamma})$ .*

*Let  $\vec{X} = (X_1, \dots, X_k)$  with  $X_i \in [n]^k$  and we let  $\sigma : [n]^k \rightarrow \{0, 1\}$  be the property  $\sigma(\vec{X}) = 1$  if  $\frac{1}{k} \sum_{i=1}^k f(X_i) \geq (1-\gamma)\mu I$  and  $\sigma(\vec{X}) = 0$  otherwise. Then the following  $(\phi_0, \dots, \phi_k), \phi_i : [n]^i \rightarrow [0, 1]$  are efficient pessimistic estimators for  $f$ .*

$$\phi_0 = de^{tk(1-\gamma)\mu} \|\mathbb{E}[\exp(-tf(X))]\|^k \leq de^{-\gamma^2 \mu k/2}$$

$$\phi_i(x_1, \dots, x_i) = de^{tk(1-\gamma)\mu} \text{Tr}(\exp(-t \sum_{j=1}^i f(x_j))) \cdot \|\mathbb{E}[\exp(-tf(X))]\|^{k-i}$$

*Proof.* The proof follows exactly along the lines of [Theorem 4.1](#).  $\blacksquare$

**Theorem 4.5.** *Let  $f : [n] \rightarrow [0, I_d]$  be such that there exists a distribution  $X$  over  $[n]$  and a number  $\mu \in (0, 1)$  such that  $\mathbb{E}[f(X)] \geq \mu I$ . Then for  $k = O(\frac{1}{\gamma^2 \mu} \log d)$ , we can efficiently and deterministically find  $\vec{x} \in [n]^k$  such that  $\frac{1}{k} \sum_{i=1}^k f(x_i) \geq (1-\gamma)\mu I$ .*

*Proof.* Use the efficient pessimistic estimators of [Theorem 4.4](#), and notice for our choice of  $k$  that  $\phi_0 < 1$  so they are useful. Then apply [Theorem 3.2](#).  $\blacksquare$

## 5 $O(\log n)$ expanding generators for any group

Our main application is a complete derandomization of the Alon-Roichman [[AR94](#)] theorem, which states that a certain kind of expander graph may be constructed by random sampling (details below). Expander graphs have a central role in theoretical computer science, especially in but not limited to the study of derandomization. Indeed, it has found a large number of

---

<sup>6</sup>In fact this is only necessary because we want a two-sided guarantee, i.e.  $\frac{1}{k} \sum_{i=1}^k f(X_i) \leq \gamma I$  and  $\frac{1}{k} \sum_{i=1}^k f(X_i) \geq -\gamma I$ . It is not necessary if we only require a one-sided guarantee, such as in the setting of [Theorem 4.4](#), where we only want  $\frac{1}{k} \sum_{i=1}^k f(X_i) \geq (1-\gamma)\mu I$ . In this second setting, when picking  $X_i$  to minimize  $\phi_i$ , notice that the quantity  $\|\mathbb{E}[\exp(tf(X))]\|$  does not change with different choices of  $X_i$ , so the only part we need to compute is the trace part, which does depend on the choice of  $X_i$ . Thus it suffices to compute the choice of  $X_i$  that minimizes the trace part of the  $\phi_i$ .

applications in a variety of areas such as deterministic amplification [CW89, IZ89], security amplification in cryptography [GIL<sup>+</sup>90], hardness of approximation [ALM<sup>+</sup>98, AFWZ95], extractor construction (e.g. see surveys [NT99, Gol97, Sha02]), construction of efficient error-correcting codes [Spi95, BH04], construction of  $\varepsilon$ -biased spaces [NN93] and much more. See [HLW06] for a comprehensive survey.

We derandomize the proof of the Alon-Roichman theorem given by [LR04] (see also [LS04]) to give a deterministic and efficient construction of the expanding generating set. We show how it implies an optimal solution to a problem of Shpilka and Wigderson [SW04] (see also [GS02]), significantly improving their results.

## 5.1 Definitions

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.

It is well-known that the set of eigenvalues of  $A$  is of the form  $1 = \lambda_1(A) > \lambda_2(A) \geq \dots \geq \lambda_n(A)$ . Note the strict separation between  $\lambda_1(A)$  and  $\lambda_2(A)$ , which follows from connectivity. The eigenvalues of  $G$  are the eigenvalues of  $A$ . Note that 1 is an eigenvalue of multiplicity 1, and that it corresponds to the uniform vector. Alternatively, the eigenvalue 1 also corresponds to the uniform eigenspace, given by the orthogonal projection matrix  $J/n$ , where  $J$  is the all 1's matrix.

The Cayley graph  $\text{Cay}(H; S)$  on a group  $H$  with respect to the generating multi-set  $S$  is the graph whose vertex set is the elements of  $H$ , and  $(h, h')$  are connected by an edge if there exists  $s \in S$  such that  $h' = hs$  (allowing for multiple edges for multiple elements in  $S$ ). We require  $S$  to be symmetric, namely for each  $s \in S$ , we also have  $s^{-1} \in S$  (this is to make the graph undirected). Let  $\lambda(\text{Cay}(H; S))$  denote the second-largest eigenvalue (in absolute value) of the normalized adjacency matrix of the Cayley graph.

Our goal is to construct an algorithm that, for a fixed  $\gamma < 1$ , takes as input the multiplication table of a group  $H$  of size  $n$  and efficiently constructs a small generating set  $S$  such that  $\lambda(\text{Cay}(H; S)) < \gamma$ . This is given by the following theorem.

**Theorem 5.1.** *Fix  $\gamma < 1$ . Then there exists an algorithm running in time  $\text{poly}(n)$  that, given  $H$ , a group of size  $n$ , constructs a symmetric set  $S \subseteq H$  of size  $|S| = O(\log n)$  such that  $\lambda(\text{Cay}(H; S)) \leq \gamma$ .*

We prove this after proving the randomized algorithm.

## 5.2 A randomized algorithm

**Theorem 5.2** ([AR94, LR04, LS04]). *Let  $H$  be a group of size  $n$ . Identify  $H$  with  $[n]$ . Let  $X_1, \dots, X_k$  be chosen randomly in  $H$ , where  $k = O(\frac{\log n}{\gamma^2})$ . We let the multi-set  $S$  be  $(X_1, \dots, X_k)$ , and we have*

$$\Pr_{S \subseteq H} [\lambda(\text{Cay}(H; S \sqcup S^{-1})) > \gamma] < 1$$

where  $S \sqcup S^{-1}$  denotes the symmetric closure of  $S$ , namely each time  $s$  appears in  $S$ ,  $s$  and  $s^{-1}$  appear in  $S \sqcup S^{-1}$ .

To identify the notation in the following proof precisely with that used in [Section 4](#), we have that  $S$  corresponds to  $\vec{X}$ ,  $|S| = k$ , and it will become clear that in this setting  $n = d = |H|$ .

*Proof.* Consider the  $n \times n$  matrices  $P_h$  for  $h \in H$ , where each  $P_h$  is the  $n \times n$  permutation matrix of the action of  $h$  on itself by right multiplication. Consider now  $\frac{1}{2}(P_h + P_{h^{-1}})$ . It is not hard to see that the normalized adjacency matrix  $A$  of  $\text{Cay}(H; S \sqcup S^{-1})$  is given by

$$A = \frac{1}{k} \sum_{i=1}^k \frac{1}{2}(P_{X_i} + P_{X_i^{-1}})$$

We wish to bound  $\lambda(A)$ . We know that the largest eigenvalue is 1 and corresponds to  $J/n$  where  $J$  is the all 1 matrix. To remove this eigenvalue from  $A$ , we consider

$$(I - J/n)A = \frac{1}{k} \sum_{i=1}^k (I - J/n) \frac{1}{2}(P_{X_i} + P_{X_i^{-1}})$$

We let our matrix-valued function be  $f(h) = (I - J/n) \frac{1}{2}(P_h + P_{h^{-1}})$ , so that

$$\lambda(A) = \|(I - J/n)A\| = \left\| \frac{1}{k} \sum_{i=1}^k f(X_i) \right\|$$

It is straight-forward to verify that  $f(h) \in \mathcal{M}_n$ ,  $\|f(h)\| \leq 1$  and  $\mathbb{E}_{h \in H}[f(h)] = 0$ .

Thus we may apply [Theorem 2.6](#) to get that

$$\Pr[\lambda(A) > \gamma] = \Pr\left[\left\| \frac{1}{k} \sum_{i=1}^k f(X_i) \right\| > \gamma\right] \tag{5.1}$$

$$\leq 2ne^{-\gamma^2|S|/4} \tag{5.2}$$

so picking  $k = O\left(\frac{\log n}{\gamma^2}\right)$  suffices to make this probability less than 1. ■

### 5.3 Derandomizing

*Proof of [Theorem 5.1](#).* To derandomize and obtain [Theorem 5.1](#), we apply [Corollary 4.2](#) to obtain efficient pessimistic estimators for the event  $\sigma(S) = 1$  iff  $\left\| \frac{1}{k} \sum_{i=1}^k f(X_i) \right\| \leq \gamma$ . We fix  $k = O\left(\frac{1}{\gamma^2} \log n\right)$  large enough such that the probability of this event is non-zero (i.e. the estimators we got are useful). We then apply [Theorem 3.2](#) to greedily choose successive elements of  $H$  to be put in  $S$  in order to make an expander. ■

### 5.4 Derandomized Homomorphism Testing

[Theorem 5.1](#) answers a question about the derandomization of homomorphism testers posed by Shpilka and Wigderson [[SW04](#)]. In this section we will use [Theorem 5.1](#) to prove [Corollary 5.4](#).

An *affine homomorphism* between two groups  $H, H'$  is a map  $f : H \rightarrow H'$  such that  $f^{-1}(0)f$  is a homomorphism. An  $(\delta, \eta)$ -test for affine homomorphisms is a tester that accepts any affine homomorphism surely and rejects with probability  $1 - \delta$  any  $f : H \rightarrow H'$  which is  $\eta$  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  $T_{H \times S}$  using an expander  $\text{Cay}(H; S)$  where  $\lambda(\text{Cay}(H; S)) < \lambda$ : simply pick a random element  $x \stackrel{R}{\leftarrow} H$  and a random element of  $y \stackrel{R}{\leftarrow} S$  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 5.3** ([SW04]). *For all groups  $H, H'$  and  $S \subseteq H$  an expanding generating set such that  $\lambda(\text{Cay}(H; S)) < \lambda$ , we can construct a tester  $T_{H \times S}$  that surely accepts any affine homomorphism  $f : H \rightarrow H'$  and rejects with probability at least  $1 - \delta$  any  $f : H \rightarrow H'$  which is  $\frac{4\delta}{1-\lambda}$  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 5.1 allows us indeed to improve this dependency to the following.

**Corollary 5.4.** *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.*

*Proof of Corollary 5.4.* Theorem 5.3 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 5.1 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. ■

## 6 Covering SDP's

Linear programming (LP) was one of the first tools computer scientists used to approximate NP-hard problems. As a natural relaxation of integer programming (IP), linear programs give fractional solutions to an IP, which may then be rounded to give provably good solutions to the original IP.

More recently, a more general class of relaxations, *semi-definite programs* (SDP's), have been used by computer scientists (e.g. [GW95, ARV04]) to give better approximation guarantees to NP-hard problems. SDP's may be solved in polynomial time (using e.g. the ellipsoid method or interior-point methods, see [Sho77, Sho87, YN77, VB96]), and again the solution may be rounded to give a solution to the original IP.

In this section we will define a restricted class of integer SDP's and show that our pessimistic estimators will give a good approximation guarantee.

### 6.1 Definition

We define the notion of *integer covering SDP's*, which are generalizations of integer covering linear programs (see e.g. [KY05]). These programs take the following form: given  $c \in [0, 1]^n$

and  $f : [n] \rightarrow [0, I_d]$ ,<sup>7</sup> find  $y \in \mathbb{N}^n$  where

$$\begin{aligned} & \text{minimize } c^T y \\ & \text{with feasibility constraint } y_1 f(1) + \dots + y_n f(n) \geq I \end{aligned} \tag{6.1}$$

where the feasibility inequality is using the p.s.d. ordering. The vector  $c$  may be interpreted as a cost vector, and we wish to minimize the cost of a solution  $y \in \mathbb{N}^n$ . This is relaxed into a covering SDP by allowing  $y \in \mathbb{R}_+^n$  where  $\mathbb{R}_+$  denotes the non-negative reals, which we would then like to round  $y$  to a solution  $\hat{y} \in \mathbb{N}^n$  that is not too much more costly. We will let  $OPT$  denote the optimal value of the *relaxed* covering SDP.

Our main theorem is as follows:

**Theorem 6.1.** *Suppose we have a program as in Equation 6.1 and suppose we have a feasible relaxed solution vector  $y \in \mathbb{R}_+^n$ . Then we can find in time  $\text{poly}(n, d)$  a feasible integer solution  $\hat{y}$  such that*

$$c^T \hat{y} \leq O(\log d) \cdot c^T y$$

**Corollary 6.2.** *Given an integer covering SDP with optimum  $OPT$ , we can efficiently find an integer solution with cost at most  $O(\log d) \cdot OPT$ .*

This is done by using a randomized rounding algorithm given implicitly in AW, and then derandomizing using pessimistic estimators.

Also, note that this is a natural generalization of integer covering linear programs of the following form: for a cost vector  $c \in \mathbb{R}_+^n$ , a matrix  $A \in \mathbb{R}_+^{d \times n}$

$$\begin{aligned} & \text{minimize } c^T y \\ & \text{subject to feasibility constraints that for all } i \in [d]: (Ay)_i \geq 1 \end{aligned}$$

This may be viewed as the special case of integer covering SDP's where all the matrices are diagonal; each  $f(i)$  is just the diagonal matrix with a column of  $A$  along the diagonal. Integer covering LP's, in turn, are a generalization of the very familiar set cover problem, which are exactly the programs where the columns of  $A$  are either 0 or 1. In the language of set cover, the universe is  $[n]$  and the columns of  $A$  are the indicator vectors for the sets we may use to cover  $[n]$ .

Our approximation for integer covering SDP's will imply a new approximation algorithm for all these covering problems with a logarithmic approximation guarantee. Thus in a sense our algorithm gives optimal approximation factors (up to constants), since a logarithmic approximation factor is optimal (up to constant factors) assuming that  $\mathbf{P} \neq \mathbf{NP}$ , as shown by [Fei98]. This connection is discussed in more detail in Section 6.4.1.

## 6.2 A randomized rounding algorithm

First suppose we have a solution to the SDP given by a vector  $y \in \mathbb{R}_+^n$ , and let us define  $Q = \sum_{j=1}^n y_j$ . In the case where  $Q \geq n$ , we can get a trivial deterministic rounding scheme with approximation factor 2 by always rounding up, since this will increase the value of the program at most by an additive  $n$ . Thus in the following we consider only programs where  $Q \leq n$ .

---

<sup>7</sup>We restrict ourselves to this scale for simplicity. Our results apply to any bounded function with a constant loss in efficiency.

Suppose we have a program as in [Equation 6.1](#) and we have solved it efficiently to obtain a solution  $y$ , where  $c^T y = OPT$ . Let  $X$  be distributed according to the distribution over  $[n]$  given by normalizing  $y$ , i.e.

$$\Pr[X = i] = y_i/Q$$

Note that, because  $y$  is a feasible solution, we have  $\mathbb{E}_X[f(X)] \geq \frac{1}{Q}I$ . It was implicitly shown in [AW] that sampling  $k = Q \cdot O(\log d)$  elements from  $[n]$  according to the distribution  $X$  and taking  $f(X_i)$  ( $1 \leq i \leq k$ ) gives us a feasible solution with approximation factor  $O(\log d)$ . We state this formally:

**Theorem 6.3.** *Suppose we sample  $k = Q \cdot 8 \ln 2d$  times from  $[n]$  according to  $X$  in order to get  $X_1, \dots, X_k$ . Furthermore, for each  $1 \leq j \leq n$ , we define the random variables*

$$\hat{Y}_j = |\{i \mid X_i = j\}|$$

*the number of times that  $j$  is sampled, and let  $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)$ . Then, with non-zero probability, we have that*

$$f(X_1) + f(X_2) + \dots + f(X_k) \geq I \quad \text{and} \quad c^T \hat{Y} \leq c^T y \cdot 16 \ln 2d$$

*Proof.* We will use a union bound to show that the probability that either  $\sum_j f(X_j) \not\geq I$  or  $c^T \hat{Y} > c^T y \cdot 16 \ln 2d$  occurs is strictly less than 1.

All expectations below are over the  $X_i$  (since the  $\hat{Y}_j$  are totally determined by the  $X_i$ ).

$$\Pr\left[\sum_{j=1}^k f(X_j) \not\geq I\right] = \Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\geq \frac{1}{k}I\right] \quad (6.2)$$

We know from the fact that  $y$  is feasible that  $\mathbb{E}[f(X)] \geq \frac{1}{Q}I$ , and so for  $k > 2Q$  we get:

$$\Pr\left[\sum_{j=1}^k f(X_j) \not\geq I\right] \leq \Pr\left[\frac{1}{k} \sum_{j=1}^k f(X_j) \not\geq \frac{1}{2} \frac{1}{Q}I\right] \quad (6.3)$$

Invoking [Theorem 2.8](#)

$$\leq de^{\frac{-k}{(8Q)}} \quad (6.4)$$

Therefore if we take  $k = Q \cdot 8 \ln 2d$  with probability greater than  $\frac{1}{2}$  we have  $\sum_j f(X_j) \geq I$ .

For the second event it is easy to see that  $c^T \hat{Y} = \sum_{j=1}^k c_{X_j}$ . Furthermore, a simple calculation shows that for each  $j$ ,  $\mathbb{E}[c_{X_j}] = c^T y/Q$ . Thus, by Markov we have:

$$\Pr[c^T \hat{Y} > c^T y \cdot 16 \ln 2d] = \Pr\left[\sum_{j=1}^k c_{X_j} > c^T y \cdot 16 \ln 2d\right] \quad (6.5)$$

$$< \frac{\mathbb{E}\left[\sum_{j=1}^k c_{X_j}\right]}{c^T y \cdot 16 \ln 2d} \quad (6.6)$$

$$= \frac{k \cdot c^T y/Q}{c^T y \cdot 16 \ln 2d} \quad (6.7)$$

Expanding  $k = Q \cdot 8 \ln 2d$  shows that this last expression is at most  $1/2$ .

Thus each bad event happens with probability less than  $1/2$ , and so the probability that either bad event happens is strictly less than 1. ■

### 6.3 Derandomizing

Derandomizing is a simple proposition. Given a program, first solve it using a standard efficient technique ([Sho77, Sho87, YN77], for a survey see [VB96]), with solution  $y$  and  $Q = \sum_{j=1}^n y_j$ . Let  $k = Q \cdot 8 \ln 2d$ . In the proof of [Theorem 6.3](#) at [Inequality 6.2](#), we can apply [Theorem 4.4](#) to get pessimistic estimators  $\phi_i$  for the event  $\sum_{j=1}^k f(X_j) \geq I$ , which we call  $\sigma$ . We only need now a pessimistic estimator  $(\psi_0, \dots, \psi_k)$  for the event of the solution not being too costly, which we call  $\tau$ .

We define  $\psi_i : [n]^i \rightarrow [0, 1]$  as follows:

$$\psi_i(x_1, \dots, x_i) = \frac{\sum_{j=1}^i c_{x_j} + (k - i)\mathbb{E}[c_X]}{c^T y \cdot 16 \ln 2d}$$

It is clear that the  $\psi_i$  are efficiently computable. They satisfy the properties of [Definition 3.1](#). This is easy to see, since the  $\psi_i$  are exactly the expressions given by a Markov bound on the event  $\tau$ , and such expressions always satisfy [Definition 3.1](#). We write this out explicitly here for completeness.

1. Looking at the proof of [Theorem 6.3](#), and substituting into [Inequality 6.5](#) shows that:

$$\Pr \left[ \sum_{j=1}^k c_{X_j} > c^T y \cdot 16 \ln 2d \mid X_1 = x_1, \dots, X_i = x_i \right] \leq \frac{\sum_{j=1}^i c_{x_j} + (k - i)\mathbb{E}[c_X]}{c^T y \cdot 16 \ln 2d} = \psi(x_1, \dots, x_i)$$

2. For estimators based on Markov, we actually have equality for this property.

$$\begin{aligned} \mathbb{E}_{X_{i+1}}[\psi_{i+1}(x_1, \dots, x_i, X_{i+1})] &= \mathbb{E}_{X_{i+1}} \left[ \frac{\sum_{j=1}^i c_{x_j} + c_{X_{i+1}} + (k - i - 1)\mathbb{E}[c_X]}{c^T y \cdot 16 \ln 2d} \right] \\ &= \frac{\sum_{j=1}^i c_{x_j} + (k - i)\mathbb{E}[c_{X_j}]}{c^T y \cdot 16 \ln 2d} \\ &= \psi_i(x_1, \dots, x_i) \end{aligned}$$

**Theorem 6.4.** *Since  $\phi_0 + \psi_0 < 1$  because of the choice of  $k = Q \cdot 8 \ln 2d$ , we may invoke [Lemma 3.3](#) to get that  $(\phi_0 + \psi_0, \dots, \phi_k + \psi_k)$  are efficient and useful pessimistic estimators for the event in [Theorem 6.3](#).*

Finally we may prove [Theorem 6.1](#).

*Proof of [Theorem 6.1](#).* By [Theorem 6.4](#) we have pessimistic estimators for the event in [Theorem 6.3](#), and so we may apply [Theorem 3.2](#), which says we can efficiently and deterministically find a suitable integer vector  $\hat{y}$  that satisfies [Theorem 6.1](#). The algorithm runs in time  $\text{poly}(n, k, d)$ , but since  $k = Q \cdot 8 \ln 2d$  and we only consider  $Q \leq n$ , this is  $\text{poly}(n, d)$ . ■

## 6.4 Quantum Hypergraph Covers

In this section we define hypergraphs and quantum hypergraphs and discuss the cover problem for both. The hypergraph cover problem is just the classical set cover problem, and the quantum hypergraph cover problem is a non-commutative generalization arising in quantum information theory [AW02]. Our efficient and useful pessimistic estimators for the integer covering SDP problem immediately give an efficient deterministic algorithm to find a quantum hypergraph cover that is optimal up to logarithmic factors.

### 6.4.1 Hypergraphs

Here we will describe the hypergraph cover problem, which is just another name for the classical set cover. 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  $f(e)$ . 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 denoted  $c(\Gamma)$ .

Using the matrix representation of  $E$ , one sees that

$$\bigcup_{e \in C} e = V \quad \Leftrightarrow \quad \sum_{e \in C} f(e) \geq I$$

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  $\sum_{e \in E} w(e)f(e) \geq I$ . Likewise, we say that the *fractional cover number*

$$\tilde{c}(\Gamma) = \min_w \left\{ \sum_{e \in E} w(e) \mid \sum_{e \in E} w(e)f(e) \geq I \right\}$$

We know that the hypergraph cover problem is hard to approximate up to a  $\ln n$  factor [Fei98]. From the definitions, it is clear that this problem is a special case of our integer covering SDP's. In the next section we generalize to the non-commutative case.

### 6.4.2 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  $\mathcal{V} = \mathbb{C}^d$  and each  $e \in \mathcal{E}$  is identified with a diagonal matrix whose diagonal entries are either 0 or 1, which we will call  $f(e)$ . We generalize this to non-commutative ‘‘edges’’ by allowing  $\mathcal{E}$  to contain other operators, i.e.  $f(e)$  can be any Hermitian operator (i.e. matrix) in  $[0, I]$ .<sup>8</sup>

<sup>8</sup>Here we are using the fact that all our previous results for real symmetric matrices generalize to complex Hermitian matrices. A complex matrix  $A$  is Hermitian if  $A = A^*$  where  $*$  denotes the conjugate transpose.

**Definition 6.5.**  $\Gamma = (\mathcal{V}, \mathcal{E})$  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  $f(e) \in [0, I_d]$ .

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  $\sum_{e \in C} f(e) \geq I$ . The cover number  $c(\Gamma)$  is the size of the smallest cover of  $\Gamma$ .

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

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

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  $f(e) \in \mathcal{M}_d$  are not necessarily diagonal in a common basis. However, it is again clear from the definitions that a quantum hypergraph cover problem is just a special case of integer covering SDP’s, so we may use [Theorem 6.1](#) to give an efficient deterministic approximation. Thus the theorem below follows.

**Theorem 6.6.** *Suppose we are given  $\Gamma = (\mathcal{V}, \mathcal{E})$  a quantum hypergraph with fractional cover number  $\tilde{c}(\Gamma)$ , with  $|\mathcal{V}| = d$  and  $|\mathcal{E}| = n$ . Then we can find an integer cover of  $\Gamma$  of size  $k = \tilde{c}(\Gamma) \cdot O(\log d)$  in time  $\text{poly}(n, d)$ .*

## 6.5 Other Applications

Our integer covering SDP also encompasses two other natural problems from quantum information theory. Given a function  $f : [n] \rightarrow [0, I_d]$ , one may want to find a probability distribution  $X$  over  $[n]$  one may want to solve either of the following

1.  $\min_X \|\mathbb{E}_X[f(X)]\|$
2.  $\max_X \lambda_d(\mathbb{E}_X[f(X)])$

The former minimizes the norm of the expected value of the distribution, which is also its largest eigenvalue, while the latter may be viewed as maximizing the lowest energy state of a quantum system, which is also its smallest eigenvalue. The second can be formulated as a covering SDP by using the cost vector  $c = \mathbf{1}$  the all 1’s vector, and then normalizing the solution vector  $y$  to be a probability distribution. The first can be formulated as the second by considering the function  $I - f$ .

In both cases, our pessimistic estimators give an “integral solution” that is worse by at most  $O(\log d)$ . In this case, an integral solution is actually a distribution with sparse support; we sample from the solution distribution  $X$  to get a distribution  $\hat{X}$  with support of size  $O(\frac{1}{\gamma^2} \log d)$  such that the corresponding objective is worse by at most a factor of  $O(\log d)$ .

## 7 Acknowledgments

We thank Sanjeev Arora for suggesting a connection between pessimistic estimators and Ahlswede-Winter's Chernoff-type bounds. We thank Satyen Kale for interesting discussions.

## References

- [AW02] R. Ahlswede and A. Winter. Strong Converse for Identification via Quantum Channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.
- [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<sup>+</sup>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.
- [ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *Proc. 36th STOC*, pages 222–231. ACM, 2004.
- [BH04] Y. Bilu and S. Hoory. Hypergraph Codes. *European Journal of Combinatorics*, 25(3):339–354, 2004.
- [CW89] A. Cohen and A. Wigderson. Dispersers, Deterministic Amplification, and Weak Random Sources. In *Proc. 30th FOCS*, pages 14–19. IEEE, 1989.
- [Fei98] U. Feige. A Threshold of  $\ln n$  for Approximating Set Cover. *Journal of the ACM*, 45(4):634–652, 1998.
- [GW95] M. X. Goemans and D. P. Williams. Improved approximation algorithms for Max-Cut and Satisfiability Problems using Semidefinite Programming. *J. of the ACM*, 42:1115–1145, 1995.
- [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<sup>+</sup>90] 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.
- [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.

- [GL89] G. H. Golub and C. F. V. Loan. *Matrix Computations*. Johns Hopkins University Press, 1989.
- [HLW06] S. Hoory, N. Linial, and A. Wigderson. *Expander Graphs and their Applications*. Manuscript, 2006.
- [IZ89] R. Impagliazzo and D. Zuckerman. How to Recycle Random Bits. In *Proc. 30th FOCS*, pages 248–253. IEEE, 1989.
- [Kal] S. Kale. Unpublished manuscript.
- [KY05] S. G. Kolliopoulos and N. E. Young. Approximation algorithms for covering/packing integer programs. *J. Comput. Syst. Sci.*, 71(4):495–505, 2005.
- [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.
- [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.
- [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.
- [Rag88] P. Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *J. Comput. Syst. Sci.*, 37(2):130–143, 1988.
- [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.
- [Spe94] J. Spencer. *Ten Lectures on the Probabilistic Method, 2nd Edition*. SIAM, 1994.
- [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.

- [WX05a] A. Wigderson and D. Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *Proc. 46th FOCS*, 2005.
- [WX05b] A. Wigderson and D. Xiao. A randomness-efficient sampler for matrix-valued functions and applications. ECCC Report TR05-107, 2005.
- [YN77] D. B. Yudin and A. S. Nemirovski. Informational complexity and efficient methods for solving complex extremal problems. *Matekon*, 13:25–45, 1977.

## A Error in [WX05a]

The error in [WX05a] is in the application of the Golden-Thompson inequality (Theorem 2.2). The following derivation, which appears in the proof of Theorem 3.6 in the second column of page 401 of [WX05a]<sup>9</sup>, is incorrect:

$$\mathbb{E} \left[ \text{Tr} \left( \exp \left( t \sum_{i=1}^k f(Y_i) \right) \right) \right] \leq \mathbb{E} \left[ \text{Tr} \left( \prod_{i=1}^k \exp(t f(Y_i)) \right) \right]$$

where the  $Y_i$  are the steps in a random expander walk and the expectation is over all walks. This is incorrect because the Golden-Thompson inequality does not generalize to more than two terms, i.e. the following does not hold in general for real symmetric matrices  $A, B, C$ :

$$\text{Tr}(\exp(A + B + C)) \leq \text{Tr}(\exp(A) \exp(B) \exp(C))$$

and it is not hard to come up with counterexamples.

We have tried various techniques to bypass this problem, but we have not discovered any method to get parameters that are sufficient for our applications. In the notation of [WX05a], it would suffice to prove

$$\text{Tr}(\mathbb{E}[\exp(t \sum_{i=2}^k f(Y_i)) \exp(t f(Y_1))]) \leq \|\tilde{A}\tilde{D}_t\| \cdot \text{Tr}(\mathbb{E}[\exp(t \sum_{i=2}^k f(Y_i))])$$

or even only

$$\text{Tr}(\mathbb{E}[\exp(t \sum_{i=1}^k f(Y_i))]) \leq d \|\tilde{A}\tilde{D}_t\|^k$$

We know from the proof of the Theorem 2.4 that both the previous inequalities hold when the normalized adjacency matrix of the graph  $A = J/n$ , i.e. we sample from the complete graph, which corresponds to independent sampling. We do not know counter-examples for either of these inequalities for sampling according to an expander walk. Thus, as far as we know, the Theorem 3.6 of [WX05a] may be true as stated.

---

<sup>9</sup>Top of page 12 in the ECCC version [WX05b]