THEORY OF COMPUTING, Volume 4 (2008), pp. 53–77
http://theoryofcomputing.org

Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications

Avi Wigderson* David Xiao†

Received: November 28, 2007; published: May 15, 2008.

Abstract: Ahlswede and Winter [IEEE Trans. Inf. Th. 2002] 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 [JCSS 1988]). As a consequence, we derandomize an efficient construction by Alon and Roichman [RSA 1994] of an expanding Cayley graph of logarithmic degree on any (possibly non-abelian) group. This gives an optimal solution to the homomorphism testing problem of Shpilka and Wigderson [STOC 2004]. We also apply these pessimistic estimators to the problem of solving semidefinite covering problems, thus giving a deterministic algorithm for the quantum hypergraph cover problem of Ahslwede and Winter.

*Partially supported by NSF grant CCR-0324906
†Supported by an NDSEG Graduate Fellowship and a NSF Graduate Fellowship

ACM Classification: G.3, G.2.2, F.2.1, F.1.2

AMS Classification: 68W20, 68R10, 60F10, 20D60, 81P68, 15A18

Key words and phrases: Chernoff bounds, matrix-valued random variables, derandomization, pessimistic estimators

 DOI: 10.4086/toc.2008.v004a003

The results above appear as theorems in our paper "A randomness-efficient sampler for matrix-valued functions and applications" [FOCS 2005, ECCC 2005], as consequences of the main claim 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 claim 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 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 a 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 large-deviation bounds for matrix-valued random variables. Such a bound was given by Ahlswede and Winter [1] (see Theorem 2.6 and Theorem 2.8 for a precise statement of their bounds).

In particular, the matrix-valued bound seems useful in giving new proofs of probabilistic constructions of expander graphs [3] and also in the randomized rounding of semidefinite covering problems, with further applications in quantum information theory [1].

In this paper we use the method of pessimistic estimators, originally formulated in [24],[1] to derandomize the Chernoff bound of [1], 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 [33], and in fact supersede them in simplicity and efficiency. However, we discovered a fatal mistake in the analysis of using an expander sampler in [33], 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.

Arora and Kale [4] independently reached results similar to the ones presented in this paper that imply the applications to constructing expanding Cayley graphs and semidefinite covering programs.

The paper is organized as follows. In Section 2 we define the linear algebra notation we use and prove the Chernoff bounds of Ahlswede-Winter, given in Theorem 2.6 and Theorem 2.8. In Section 3 we review the method of pessimistic estimators and how it is used to derandomize algorithms. In Section 4 we construct pessimistic estimators for the Ahlswede-Winter Chernoff bounds. Finally we apply these estimators to derandomize the construction of Cayley expanders in Section 5 and to derandomize the rounding of integer covering SDP's in Section 6.

[1] The simpler method of conditional probabilities was described earlier in the first edition of [29]. Ideas similar to those of [24] also appeared in [7].

2 Matrix-valued random variables and the Chernoff bound of Alhswede and Winter

We will work with the set $\mathcal{M}_d$ of real symmetric $d \times d$ matrices.[2] 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 $\lambda_1(A) \geq \ldots \geq \lambda_d(A)$ denote the eigenvalues of A in non-increasing order. Recall that every matrix $A \in \mathcal{M}_d$ has an orthonormal eigenbasis.

We will measure distance between matrices using the operator norm

$$\|A\| = \max_v \frac{\|Av\|}{\|v\|} = \max_i |\lambda_i(A)|\,.$$

We will also frequently use the trace, $\mathrm{Tr}(A) = \sum_{i=1}^d \lambda_i(A)$. It is well-known that for any orthonormal basis $v_1, \ldots, v_d \in \mathbb{R}^d$ we have that $\mathrm{Tr}(A) = \sum_{i=1}^d \langle v_i, Av_i \rangle$, 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 semidefinite (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 symmetric 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 an eigenbasis of A is also an eigenbasis of $\exp(A)$ 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 $f : [n] \to [-I_d, I_d]$, where $[n] = \{1, \ldots, n\}$. Let X be a distribution (not necessarily uniform) over $[n]$, and consider the variable $f(X)$. This is a natural extension of bounded discrete random variables over the reals, which may be thought of as functions $f : [n] \to [-1, 1]$. 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 Tr is linear, $\mathbb{E}$ and Tr commute: $\mathbb{E}[\mathrm{Tr}(f(X))] = \mathrm{Tr}(\mathbb{E}[f(X)])$. We let $\mathrm{supp}(X)$ 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 $\mathrm{supp}(X)$.

We will use the following useful facts several times:

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

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

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

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

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

$$\mathrm{Tr}(AB) \leq \sum_{i=1}^{d} \max_j \lambda_j \langle v_i, Bv_i \rangle \leq \|A\| \cdot \mathrm{Tr}(B)\,.$$

$\square$

Theorem 2.2 (Golden-Thompson inequality, [12, 31]). *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.

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

Theorem 2.3 (Markov's inequality [1]). *For any $\gamma > 0$, any function $g : [n] \to \mathcal{M}_d$ such that $g(x) \geq 0$ for all $x \in [n]$, and for any random variable X over $[n]$, we have*

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

Proof.

$$\Pr[g(X) \not\leq \gamma I] = \Pr[\|g(X)\| > \gamma] \leq \tfrac{1}{\gamma}\mathbb{E}[\|g(X)\|]\,.$$

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

$$\leq \tfrac{1}{\gamma}\mathbb{E}[\mathrm{Tr}(g(X))] = \tfrac{1}{\gamma}\mathrm{Tr}(\mathbb{E}[g(X)])\,.$$

$\square$

The following Theorem 2.4 is the main theorem proving [1]'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 ([1]). *Suppose $f : [n] \to [-I_d, I_d]$ and let $X_1, \ldots, X_k$ be arbitrary independent random variables distributed over $[n]$. Then for all $\gamma \in \mathbb{R}$:*

$$\Pr\left[\tfrac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] \leq d e^{-t\gamma k} \prod_{j=1}^{k} \left\|\mathbb{E}[\exp(t f(X_j))]\right\|\,.$$

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

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

The equality holds because for any $A \in \mathcal{M}_d, \alpha \in \mathbb{R}$, the statement $A \not\leq \alpha I$ is equivalent to saying some eigenvalue of A is larger than α, which is the same as saying that some eigenvalue of $\exp(A)$ is larger than e^{α}, which in turn is equivalent to $\exp(A) \not\leq e^{\alpha} I$. Then the following inequality is a direct consequence of Theorem 2.3 since $\exp(A) \geq 0$ for all $A \in \mathcal{M}_d$.

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

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 $f : [n] \to \mathcal{M}_d$ and any random variable X over $[n]$, we have*

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

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

$$\begin{aligned}
\Pr\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma I\right] &\leq e^{-t\gamma k}\mathrm{Tr}\left(\mathbb{E}\left[\exp\left(t\sum_{j=1}^{k} f(X_j)\right)\right]\right) \\
&= e^{-t\gamma k}\,\mathbb{E}_{X_1,\dots,X_{k-1}}\left[\mathrm{Tr}\left(\mathbb{E}_{X_k}\left[\exp\left(t\sum_{j=1}^{k-1} f(X_j) + tf(X_k)\right)\right]\right)\right] && \textit{(By independence)} \\
&\leq e^{-t\gamma k}\,\mathbb{E}_{X_1,\dots,X_{k-1}}\left[\big\|\mathbb{E}[\exp(tf(X_k))]\big\| \cdot \mathrm{Tr}\left(\exp\left(t\sum_{j=1}^{k-1} f(X_j)\right)\right)\right] && \textit{(Apply Lemma 2.5)} \\
&= e^{-t\gamma k}\big\|\mathbb{E}[\exp(tf(X_k))]\big\| \cdot \mathrm{Tr}\left(\mathbb{E}_{X_1,\dots,X_{k-1}}\left[\exp\left(t\sum_{j=1}^{k-1} f(X_j)\right)\right]\right) && \textit{(Put expectation back inside)} \\
&\leq e^{-t\gamma k}\prod_{j=1}^{k}\big\|\mathbb{E}[\exp(tf(X_j))]\big\| \cdot \mathrm{Tr}(I) && \textit{(Repeat k times \dots)} \\
&= de^{-t\gamma k}\prod_{j=1}^{k}\big\|\mathbb{E}[\exp(tf(X_j))]\big\|. && (2.2)
\end{aligned}$$

This completes the proof modulo Lemma 2.5. □

Proof of Lemma 2.5.

$$\begin{aligned}
\mathrm{Tr}(\mathbb{E}[\exp(A + f(X))]) &= \mathbb{E}[\mathrm{Tr}(\exp(A + f(X)))] && \textit{(Since trace and expectation commute)} \\
&\leq \mathbb{E}[\mathrm{Tr}(\exp(f(X))\exp(A))] && \textit{(Applying the Golden-Thompson inequality)} \\
&\leq \mathrm{Tr}(\mathbb{E}[\exp(f(X))]\exp(A)) && \textit{(Commuting trace and expectation again)} \\
&\leq \big\|\mathbb{E}[\exp(tf(X))]\big\| \cdot \mathrm{Tr}(\exp(A)). && \textit{(By Fact 2.1)}
\end{aligned}$$

□

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 that of a multiplicative deviation.

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

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

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

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

We use the following claim to bound the RHS.

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

Proof. This follows from the Taylor expansion of exp:

$$\begin{aligned}\left\|\mathbb{E}[\exp(t f(X))\right\| &= \left\|\mathbb{E}[I + t f(X) + (t f(X))^2/2 + \ldots]\right\| \\ &= \left\|I + t\,\mathbb{E}[f(X)] + \mathbb{E}[(t f(X))^2/2 + \ldots]\right\|.\end{aligned}$$

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 gives

$$\leq 1 + t^2\,.$$

□

[3]For the sake of simplicity, no attempt was made to optimize the constant in the exponent of the bound in this analysis. To get a tighter bound, we can apply the analysis of [1] to get a bound of

$$d e^{-kD\left(\frac{1+\gamma}{2}\|\frac{1}{2}\right)}\,.$$

Here $D(p\|q) = p(\log p - \log q) + (1-p)(\log(1-p) - \log(1-q))$ is the relative entropy function, and using the approximation $D((1+\gamma)/2\|1/2) \geq \gamma^2/(2\ln 2)$, which can be shown by looking at the Taylor expansion of $D(\cdot\|\cdot)$, we have the improved bound of $d e^{-k\gamma^2/(2\ln 2)}$.

We will choose $t = \gamma/2 \le 1/2$, so we may apply Claim 2.7 to Theorem 2.4 to get

$$\begin{aligned}
\Pr\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\leq \gamma I\right] &\le de^{-t\gamma k}(1+t^2)^k \\
&\le de^{-t\gamma k + t^2 k} && \textit{(Using } 1+x \le e^x \textit{ for all } x \in \mathbb{R}\textit{)} \\
&\le de^{-\gamma^2 k/4}. && \textit{(Choosing } t = \gamma/2\textit{)}
\end{aligned}$$

□

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

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

Proof. We can assume without loss of generality that $M = \mu I$.[4] 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:

$$\Pr\left[\frac{1}{k}\sum_{i=1}^{k} f(X_i) \not\geq (1-\gamma)\mu I\right] = \Pr\left[\frac{1}{k}\sum_{i=1}^{k} (I - f(X_i)) \not\leq (1-(1-\gamma)\mu)I\right]. \tag{2.3}$$

Applying Theorem 2.4

$$\begin{aligned}
&\le de^{-t(1-(1-\gamma)\mu)k} \left\| \mathbb{E}[\exp(t(I - f(X)))] \right\|^k && (2.4)\\
&= d \left\| \mathbb{E}[\exp(-tf(X))e^{t(1-\gamma)\mu}] \right\|^k. && (2.5)
\end{aligned}$$

This last quantity was analyzed in the proof of Theorem 19 of [1], with the following conclusion which we state without proof:

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

$$\left\| \mathbb{E}[\exp(-tf(X))]e^{t(1-\gamma)\mu} \right\| \le e^{-\gamma^2\mu/(2\ln 2)}.$$

Applying this claim to Inequality (2.5) gives us the theorem. □

3 Method of pessimistic estimators

First we review the method of pessimistic estimators, due to Raghavan [24]. 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 $\sigma : \mathrm{supp}(X) \to \{0,1\}$, $\sigma(x) = 1$ iff x is in the event. We wish to efficiently and deterministically find a particular $x \in \mathrm{supp}(X)$ such that $\sigma(x) = 1$.

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

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 σ 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.* $\operatorname{supp}(\vec{X}) \subseteq [n]^k$ 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

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

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

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

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.[5]

Definition 3.1. Let $\sigma : [n]^k \to \{0,1\}$ be an event on a random variable $\vec{X}$ distributed over $[n]^k$ and suppose $\Pr[\sigma(\vec{X}) = 1] > 0$. We say that $\phi_0, \dots, \phi_k$, $\phi_i : [n]^i \to [0,1]$ (here ϕ_0 is just a number in $[0,1]$), are *pessimistic estimators* for σ if the following hold.

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

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

[5]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).

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

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

We will also want the pessimistic estimators to be *efficient*, namely each ϕ_i is efficiently computable, and *useful*, which means $\phi_0 < 1$. This last condition is because ϕ_0 is a bound on the initial probability of failure, which we need to be strictly less than 1.

Theorem 3.2 ([24]). *If there exist efficient and useful pessimistic estimators $(\phi_0,\dots,\phi_k)$ for an event σ, 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 $\phi_{i+1}(x_1,\dots,x_{i+1}) \le \phi_i(x_1,\dots,x_i) < 1$. We are guaranteed that

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

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 $\Pr[\sigma(\vec{x}) = 0] \le \phi_k(\vec{x}) < 1$, and therefore $\sigma(\vec{x}) = 1$.

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

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

Lemma 3.3. *Suppose $\sigma,\tau : [n]^k \to \{0,1\}$ 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 σ,τ 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:

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

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 $f:[n]\to[-I_d,I_d]$. 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 $1>\gamma>0$. 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 $\sigma:[n]^k\to\{0,1\}$ be the event $\sigma(\vec{x})=1$ if $\frac{1}{k}\sum_{i=1}^k f(x_i)\le\gamma I$ and $\sigma(\vec{x})=0$ otherwise. Then the following $(\phi_0,\dots,\phi_k)$, $\phi_i:[n]^i\to[0,1]$ are efficient pessimistic estimators *for σ:*

$$\begin{aligned}\phi_0 =& de^{-t\gamma k}\left\|\mathbb{E}[\exp(tf(X))]\right\|^k \quad (\textit{which is at most } de^{-\gamma^2k/4}),\\ \phi_i(x_1,\dots,x_i) =& de^{-t\gamma k}\mathrm{Tr}\left(\exp\left(t\sum_{j=1}^i f(x_j)\right)\right)\cdot\left\|\mathbb{E}[\exp(tf(X))]\right\|^{k-i}.\end{aligned}$$

Proof. We verify the properties of Definition 3.1.

1. From Inequality (2.1):

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

   By fixing $X_j=x_j$ for all $j\le i$, we derive that

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

2. We use the following derivation, where the inequality follows from Lemma 2.5:

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

To see that the ϕ_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 $d \times d$ matrices $f(1), \ldots, f(n)$) and 1^k. Thus we desire the algorithm to be computable in time $\mathrm{poly}(n,d,k)$. 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 [16]. 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. $\square$

Theorem 4.1 gives us pessimistic estimators $(\phi_0, \ldots, \phi_k)$ for σ, and the same proof gives efficient pessimistic estimators $(\psi_0, \ldots, \psi_k)$ 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 ϕ_i gives us the following.

Corollary 4.2. *Let $f : [n] \to [-I_d, I_d]$. Let X be distributed over $[n]$ with $\mathbb{E}_X[f(X)] = 0$, and let $X_1, \ldots, X_k$ 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 $\eta : [n]^k \to \{0,1\}$ be the event $\eta(\vec{x}) = 1$ if $\|\frac{1}{k}\sum_{i=1}^k f(x_i)\| \leq \gamma$ and $\eta(\vec{x}) = 0$ otherwise. Then $(\phi_0 + \psi_0, \ldots, \phi_k + \psi_k)$ are efficient pessimistic estimators for η.

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

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 $\mathbb{E}[\exp(-tf(X))]$.[6] This is of course true when X is uniform, or when we can efficiently compute $\Pr[X = x]$ for each $x \in [n]$. The actual distribution is irrelevant, since we exhaustively search through the entire space for the choice of each X_i.

Theorem 4.3. *Let $f : [n] \to [-I_d, I_d]$ 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 pessimistic estimators of Corollary 4.2. Pick $k = O(\frac{1}{\gamma^2}\log d)$ such that $\phi_0 + \psi_0 < 1$ and so that the estimators are useful. We may then apply Theorem 3.2 to get the result. $\square$

We can construct pessimistic estimators for Theorem 2.8 in the same way.

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

$$t = \log\left(\frac{1-(1-\gamma)\mu}{1-\mu}\frac{1}{(1-\gamma)}\right).$$

[6] 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 ϕ_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 ϕ_i.

Let $\vec{X} = (X_1, \ldots, X_k)$ with $X_i \in [n]$ and we let $\sigma : [n]^k \to \{0,1\}$ be the event $\sigma(\vec{x}) = 1$ 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 $(\phi_0, \ldots, \phi_k), \phi_i : [n]^i \to [0,1]$ are efficient pessimistic estimators for σ:

$$\phi_0 = de^{tk(1-\gamma)\mu} \left\| \mathbb{E}[\exp(-tf(X))] \right\|^k \quad (\textit{which is at most } de^{-\gamma^2 \mu k/(2\ \ln 2)}),$$

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

Proof. The proof follows exactly along the lines of Theorem 4.1. □

Theorem 4.5. *Let $f : [n] \to [0, I_d]$ 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. □

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

Our main application is a complete derandomization of the Alon-Roichman [3] 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, they have found a large number of applications in a variety of areas such as deterministic amplification [9, 18], security amplification in cryptography [14], hardness of approximation [5, 2], extractor construction (e.g. see surveys [23, 13, 25]), construction of efficient error-correcting codes [30, 8], construction of ε-biased spaces [22] and much more. See [17] for a comprehensive survey.

We derandomize the proof of the Alon-Roichman theorem given by [20] (see also [21]) 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 [28] (see also [15]), significantly improving their results.

5.1 Definitions

Given an undirected d-regular graph $G = (V,E)$ on n vertices, we define its normalized adjacency matrix A by setting $A_{ij} = e_{ij}/d$ where e_{ij} is the number of edges between vertices i and j (we allow self-loops and multiple edges). The matrix A is real and symmetric.

We assume G is connected. It is well known that the set of eigenvalues of A is of the form $1 = \lambda_1(A) > \lambda_2(A) \geq \ldots \geq \lambda_n(A)$. (The strict separation between $\lambda_1(A)$ and $\lambda_2(A)$ follows from connectivity.) The eigenvalues of G are the eigenvalues of A. Note that 1 is an eigenvalue of multiplicity 1, and with corresponding eigenvector $u = [1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$. The orthogonal projection to the subspace spanned by the eigenvector u is given by the matrix J/n, where J is the all 1's matrix.

The Cayley graph $\mathrm{Cay}(H;S)$ on a group H with respect to the generating multi-set $S \subset H$ is the graph whose vertex set is H, and where h and 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 $\lambda(\mathrm{Cay}(H;S))$ denote the second-largest eigenvalue (in absolute value) of the normalized adjacency matrix of the Cayley graph.

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

Theorem 5.1. *Fix $\gamma < 1$. Then there exists an algorithm running in time $\mathrm{poly}(n)$ that, given a group H of order n, constructs a symmetric set $S \subseteq H$ of size $|S| = O(\frac{1}{\gamma^2}\log n)$ such that $\lambda(\mathrm{Cay}(H;S)) \leq \gamma$.*

We prove this after presenting the randomized algorithm.

5.2 A randomized algorithm

Theorem 5.2 ([3, 20, 21]). *Fix $0 < \gamma < 1$, and let H be a group of order n. Identify H with $[n]$. Let $X_1,\dots,X_k$ be chosen randomly in H, where $k = O(\frac{1}{\gamma^2}\log n)$. We let the multi-set S be $(X_1,\dots,X_k)$, and we have*

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

where $S\sqcup S^{-1}$ denotes the symmetric closure of S, namely the number of occurrences of s and s^{-1} in $S\sqcup S^{-1}$ equals the number of occurrences of s in S.

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 that $n = d = |H|$.

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 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 $\mathrm{Cay}(H;S\sqcup S^{-1})$ 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. Since we want to analyze the second-largest eigenvalue, 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

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

It is straightforward 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

$$\Pr[\lambda(A) > \gamma] = \Pr\left[\left\|\frac{1}{k}\sum_{i=1}^{k} f(X_i)\right\| > \gamma\right] \le 2ne^{-\gamma^2|S|/4} \tag{5.1}$$

so picking $k = O(\frac{1}{\gamma^2}\log n)$ 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 $\|\frac{1}{k}\sum_{i=1}^{k} f(X_i)\| \le \gamma$. We fix $k = O(\frac{1}{\gamma^2}\log n)$ 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 affine homomorphism testers posed by Shpilka and Wigderson [28]. 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 $f : H \to H'$ 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: $d(f,g) = \Pr[f(x) \ne g(x)]$, where the probability is over x chosen uniformly from H.

Shpilka and Wigderson [28] showed how to construct a tester $T_{H\times S}$ efficiently using an expander $\mathrm{Cay}(H;S)$ where $\lambda(\mathrm{Cay}(H;S)) < \lambda$: simply pick a random element $x \in H$ and a random element $y \in S$ and check to see that $f(0)f(x)^{-1}f(xy) = f(y)$. It is clear this test accepts f surely if f is an affine homomorphism. [28] shows that if $12\delta < 1-\lambda$ then this rejects with probability $1-\delta$ any f that is $4\delta/(1-\lambda)$-far from being an affine homomorphism.

Theorem 5.3 ([28]). *For all groups H, H' and $S \subseteq H$ an expanding generating set such that $\lambda(\mathrm{Cay}(H;S)) < \lambda$, 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 $12\delta/(1-\lambda) < 1$. That is, $T_{H\times S}$ is a $(\delta, 4\delta/(1-\lambda))$-test for affine homomorphisms.*

In [28] the deterministic construction of S gave a set of size $|H|^{\varepsilon}$ for arbitrary $\varepsilon > 0$. The explicit construction given in [28] 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 $|H|^{O(1)}$ an affine 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. $\square$

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, *semidefinite programs* (SDP's), have been used by computer scientists (e.g. [11, 6]) 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 [26, 27, 35, 32]), 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. [19]). These programs take the following form: given $c \in [0,1]^n$ and $f : [n] \to [0, I_d]$,[7] find $y \in \mathbb{N}^n$ where

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

where the feasibility inequality uses 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 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 $\mathrm{poly}(n,d)$ 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 [1], and then derandomizing using pessimistic estimators.

[7] We restrict ourself to this scale for simplicity. Our results apply to any bounded function with a constant loss in efficiency.

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 $A \in \mathbb{R}_+^{d\times n}$

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

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 i'th 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 [10]. 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 $Q = \sum_{j=}^n y_j$. 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 $Q \leq n$.

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 [1] 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 ([1]). *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 $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)$.[8] *Then, with non-zero probability, we have that*

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

Proof. We will use a union bound to show that the probability that either

$$\sum_j f(X_j) \not\geq I \quad \text{or} \quad c^T\hat{Y} > c^T y \cdot 16\ln 2d$$

[8] Notice that $\sum_{i=1}^k f(X_i) = \sum_{j=1}^n \hat{Y}_j f(j)$.

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).

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

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:

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

Invoking Theorem 2.8, we obtain

$$\Pr\left[\sum_{j=1}^{k} f(X_j) \not\geq I\right] \leq d e^{-k/(8Q)}.$$

Therefore if we take $k = Q \cdot 8 \ln 2d$ with probability greater than $1/2$ 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:

$$\begin{aligned}
\Pr[c^T \hat{Y} > c^T y \cdot 16 \ln 2d] &= \Pr\left[\sum_{j=1}^{k} c_{X_j} > c^T y \cdot 16 \ln 2d\right] \\
&< \frac{\mathbb{E}\left[\sum_{j=1}^{k} c_{X_j}\right]}{c^T y \cdot 16 \ln 2d} \\
&= \frac{k \cdot c^T y/Q}{c^T y \cdot 16 \ln 2d}.
\end{aligned} \tag{6.3}$$

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. $\square$

6.3 Derandomizing

Derandomizing is a simple proposition. Given a program, first solve it using a standard efficient technique ([26, 27, 35], for a survey see [32]), 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 ϕ_i for the event $\sum_{j=1}^{k} f(X_j) \geq I$, which we call σ. We only need now a pessimistic estimator $(\psi_0, \ldots, \psi_k)$ for the event of the solution not being too costly, which we call τ.

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

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

It is clear that the ψ_i are efficiently computable. They satisfy the properties of Definition 3.1. This is easy to see, since the ψ_i are exactly the expressions given by a Markov bound on the event τ, and such expressions always satisfy Definition 3.1. We write this out explicitly here for completeness.

1. By an application of Markov (this is the same as in Inequality (6.3)), we see:

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

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

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

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 $(\phi_0 + \psi_0, \ldots, \phi_k + \psi_k)$ 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 $\mathrm{poly}(n,k,d)$, but since $k = Q \cdot 8 \ln 2d$ and we only consider $Q \leq n$, this is $\mathrm{poly}(n,d)$. □

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 [1]. 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 is 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 an assignment $w : E \to \mathbb{R}_+$ of non-negative weights to the edges such that $\sum_{e \in E} w(e) f(e) \geq I$. The *fractional cover number* is defined as

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

We know that the hypergraph cover problem is hard to approximate up to a $\ln n$ factor [10]. 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

Ahlswede and Winter [1] define *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 a pair $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \mathbb{C}^d$ and each $e \in \mathcal{E}$ is identified with a diagonal matrix $f(e)$ whose diagonal entries are 0 or 1. 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 $[0, I]$.[9]

Definition 6.5. A *quantum hypergraph* is a pair $\Gamma = (\mathcal{V}, \mathcal{E})$ 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 $f(e) \in [0, I_d]$.

We define a *cover* of a quantum hypergraph 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 Γ.

Likewise, we define a fractional cover to be an assignment $w : E \to \mathbb{R}_+$ of non-negative weights to the edges such that $\sum_{e \in \mathcal{E}} w(e) f(e) \geq I$, and the fractional cover number as

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

Note that this corresponds exactly to 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.

[9] A complex matrix A is Hermitian (self-adjoint) if $A = A^*$ where * denotes the conjugate transpose. Here we use the fact that all our previous results for real symmetric matrices generalize to complex Hermitian matrices.

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 (extended to complex matrices), 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 a quantum hypergraph $\Gamma = (\mathcal{V}, \mathcal{E})$ with fractional cover number $\tilde{c}(\Gamma)$, with $\dim(\mathcal{V}) = d$ and $|\mathcal{E}| = n$. Then we can find an integer cover of Γ of size $k = \tilde{c}(\Gamma) \cdot O(\log d)$ in time* $\mathrm{poly}(n,d)$.

6.5 Other Applications

Our integer covering SDP (and its extension to complex matrices) also encompasses two other natural problems from quantum information theory. Given a function $f : [n] \to [0, I_d]$, one may want to find a probability distribution X over $[n]$ that achieves the optimum of either of the following quantities:

1. $\min_X \lambda_1(\mathbb{E}_X[f(X)]) = \min_X \|\mathbb{E}_X[f(X)]\|$,
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 Ahslwede-Winter's Chernoff-type bounds. We thank Satyen Kale for interesting discussions.

A Error in [33]

Due to the error, the main claim of [33] remains open.

Conjecture A.1. Let $f : [N] \to [-I_d, I_d]$ be a matrix-valued function with expectation $\mathbb{E}[f] = 0$. Let G be an expander graph on $[N]$ with spectral gap ε. Let $Y_1, \ldots, Y_k$ be a random walk of length k on G (for sufficiently large k). Then it holds that

$$\Pr\left[\left\|\frac{1}{k}\sum_{i=1}^{k} f(Y_i)\right\| > \gamma\right] \le d e^{-\Omega(\varepsilon\gamma^2 k)}.$$

The above statement gives a relatively strong bound; in its weakest non-trivial form, the conjecture would be that

$$\Pr\left[\left\|\frac{1}{k}\sum_{i=1}^{k} f(Y_i)\right\| > \gamma\right] \leq \mathrm{poly}(d)e^{-\mathrm{poly}(\varepsilon,\gamma)k}.$$

The error in [33] 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 [33][10], is incorrect:

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

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:

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

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 [33], it would suffice to prove

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

or even only

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

We know from the proof of the Theorem 2.4 that both of the inequalities hold when the normalized adjacency matrix of the graph is $A = J/n$, i. e., we sample from the complete graph with self-loops, which corresponds to independent sampling. We do not know counterexamples to either of these inequalities for sampling according to an expander walk. Thus, as far as we know, Theorem 3.6 of [33] may be true as stated.

References

[1] * RUDOLF AHLSWEDE AND ANDREAS WINTER: Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002. [IEEE:10.1109/18.985947]. 1, 2.3, 2, 2.4, 2.6, 3, 2.8, 2, 2.9, 6.1, 6.2, 6.3, 6.4, 6.4.2

[10] Top of page 12 in the ECCC version [34]

[2] * NOGA ALON, URIEL FEIGE, AVI WIGDERSON, AND DAVID ZUCKERMAN: Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995. [Springer:r591795p150lj86q]. 5

[3] * NOGA ALON AND YUVAL ROICHMAN: Random Cayley graphs and expanders. *Random Structures & Algorithms*, 5, 1994. 1, 5, 5.2

[4] * SANJEEV ARORA AND SATYEN KALE: A combinatorial, primal-dual approach to semidefinite programs. In *Proc. 39th STOC*, pp. 227–236, New York, NY, USA, 2007. ACM. [STOC:10.1145/1250790.1250823]. 1

[5] * SANJEEV ARORA, CARSTEN LUND, RAJEEV MOTWANI, MADHU SUDAN, AND MARIO SZEGEDY: Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998. 5

[6] * SANJEEV ARORA, SATISH RAO, AND UMESH VAZIRANI: Expander flows, geometric embeddings, and graph partitionings. In *Proc. 36th STOC*, pp. 222–231. ACM Press, 2004. [STOC:10.1145/1007352.1007355]. 6

[7] * JÓZSEF BECK AND JOEL SPENCER: Integral approximation sequences. *Mathematical Programming*, 30(1):88–98, 1984. [Springer:d77488n21q0222p0]. 1

[8] * YONATAN BILU AND SHLOMO HOORY: On codes from hypergraphs. *European Journal of Combinatorics*, 25(3):339–354, 2004. [Elsevier:10.1016/j.ejc.2003.10.002]. 5

[9] * AVIAD COHEN AND AVI WIGDERSON: Dispersers, deterministic amplification, and weak random sources. In *Proc. 30th FOCS*, pp. 14–19. IEEE Computer Society Press, 1989. 5

[10] * URIEL FEIGE: A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998. [JACM:10.1145/285055.285059]. 6.1, 6.4.1

[11] * M. X. GOEMANS AND D. P. WILLIAMS: Improved approximation algorithms for max-cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. [JACM:10.1145/227683.227684]. 6

[12] * S. GOLDEN: Lower bounds for the Helmholtz function. *Physical Review*, 137B(4):B1127–1128, 1965. [10.1103PhysRev.137.B1127]. 2.2

[13] * ODED GOLDREICH: A sample of samplers - a computational perspective on sampling (survey). *Electronic Colloquium on Computational Complexity (ECCC)*, 4(020), 1997. [ECCC:TR97-020]. 5

[14] * ODED GOLDREICH, RUSSELL IMPAGLIAZZO, LEONID LEVIN, RAMARATHNAM VENKATESAN, AND DAVID ZUCKERMAN: Security preserving amplification of hardness. In *Proc. 31st FOCS*, pp. 318–326. IEEE Computer Society Press, 1990. 5

[15] * ODED GOLDREICH AND MADHU SUDAN: Locally testable codes and PCPs of almost-linear length. In *Proc. 43rd FOCS*, pp. 13–22, Los Alamitos-Washington-Brussels-Tokyo, 2002. IEEE Computer Society, IEEE Computer Society Press. [FOCS:10.1109/SFCS.2002.1181878]. 5

[16] * G. H. GOLUB AND C. F. VAN LOAN: *Matrix Computations*. Johns Hopkins University Press, 1989. 4

[17] * SHLOMO HOORY, NATHAN LINIAL, AND AVI WIGDERSON: Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006. 5

[18] * RUSSELL IMPAGLIAZZO AND DAVID ZUCKERMAN: How to recycle random bits. In *Proc. 30th FOCS*, pp. 248–253. IEEE, 1989. 5

[19] * STAVROS G. KOLLIOPOULOS AND NEAL E. YOUNG: Approximation algorithms for covering/packing integer programs. *J. Computer and System Sciences*, 71(4):495–505, 2005. [JCSS:10.1016/j.jcss.2005.05.002]. 6.1

[20] * ZEPH LANDAU AND ALEXANDER RUSSELL: Random Cayley graphs are expanders: a simplified proof of the Alon-Roichman theorem. *The Electronic Journal of Combinatorics*, 11(1), 2004. 5, 5.2

[21] * PO-SHEN LOH AND LEONARD J. SCHULMAN: Improved expansion of random Cayley graphs. *Discrete Mathematics and Theoretical Computer Science*, 6(2):523–528, 2004. 5, 5.2

[22] * JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, August 1993. [SICOMP:10.1137/0222053]. 5

[23] * NOAM NISAN AND AMNON TA-SHMA: Extracting randomness: A survey and new constructions. *J. Computer and System Sciences*, 58(1):148–173, 1999. [JCSS:10.1006/jcss.1997.1546]. 5

[24] * PRABHAKAR RAGHAVAN: Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *J. Computer and System Sciences*, 37(2):130–143, 1988. [JCSS:10.1016/0022-0000(88)90003-7]. 1, 3, 3.2

[25] * RONEN SHALTIEL: Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 2002. 5

[26] * N.Z. SHOR: Cut-off method with space extension in convex programming problems. *Cybernetics and Systems Analysis*, 13:94–96, 1977. [Springer:w88055332t2p4215]. 6, 6.3

[27] * N.Z. SHOR: Quadratic optimization problems. *Soviet Journal of Circuits and Systems Sciences*, 25:1–11, 1987. 6, 6.3

[28] * AMIR SHPILKA AND AVI WIGDERSON: Derandomizing homomorphism testing in general groups. In *Proc. 36th STOC*, pp. 427–435. ACM Press, 2004. [STOC:10.1145/1007352.1007421]. 5, 5.4, 5.3, 5.4

[29] * JOEL SPENCER: *Ten Lectures on the Probabilistic Method, 2nd Edition*. SIAM, 1994. 1

[30] * DANIEL SPIELMAN: *Computationally Efficient Error-Correcting Codes and Holographic Proofs*. PhD thesis, M.I.T., 1995. 5

[31] * C. J. THOMPSON: Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1823, 1965. [doi:10.1063/1.1704727]. 2.2

[32] * L. VANDENBERGHE AND S. BOYD: Semidefinite programming. *SIAM Review*, 38:49–95, March 1996. [SIREV:10.1137/1038003]. 6, 6.3

[33] * AVI WIGDERSON AND DAVID XIAO: A randomness-efficient sampler for matrix-valued functions and applications. In *Proc. 46th FOCS*, pp. 397–406, 2005. [FOCS:10.1109/SFCS.2005.8]. 1, A, A

[34] * AVI WIGDERSON AND DAVID XIAO: A randomness-efficient sampler for matrix-valued functions and applications. ECCC Report TR05-107, 2005. [ECCC:TR05-107]. 10

[35] * D. B. YUDIN AND A. S. NEMIROVSKI: Informational complexity and efficient methos for solving complex extremal problems. *Matekon*, 13:25–45, 1977. 6, 6.3

AUTHORS

Avi Wigderson [About the author]
Professor
School of Mathematics
Institute for Advanced Study
avi@ias.edu
http://www.math.ias.edu/~avi

David Xiao [About the author]
Student
Department of Computer Science
Princeton University
dxiao@cs.princeton.edu
http://www.cs.princeton.edu/~dxiao

ABOUT THE AUTHORS

AVI WIGDERSON was born in Haifa, Israel in 1956, and received his Ph. D. in 1983 at Princeton University under Dick Lipton. He enjoys and is fascinated with studying the power and limits of efficient computation, and the remarkable impacts of this field on understanding our world. Avi's other major source of fascination and joy are his three kids, Eyal, Einat, and Yuval.

DAVID XIAO hopes to receive his Ph. D. soon from Princeton University under the supervision of Boaz Barak and Avi Wigderson. He is interested in complexity theory and cryptography, and in his spare time he enjoys wandering through city streets and rocking out to French house music.