Concentration of Measure in Matrix-Valued Dependent Settings

Juan C. Martínez Mori, Ayush Sekhari
Cornell University

1 Introduction

Concentration of measure is a central topic in the design and analysis of randomized algorithms. In particular, concentration of measure allows one to argue that the observed behavior of a randomized algorithm is *essentially deterministic*. Thus, tight concentration bounds are desired for the efficient de-randomization of algorithms and probabilistic constructions.

Perhaps the most familiar example is Chernoff's bound (Chernoff et al., 1952), which states that the empirical mean of a sequence $X_1, \ldots, X_k$ of independent and identically distributed (IID) random variables deviates from its true mean with a probability that decays exponentially in the number of samples and the square of the deviation. That is,

$$\mathbb{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} X_i - \mathbb{E}[X]\right| \geq \epsilon\right] \leq 2 \cdot \exp\left(-\Omega\left(k\epsilon^2\right)\right). \tag{1}$$

An important generalization of this bound was achieved by Gilman (Gillman, 1998), who significantly relaxed the independence assumption to Markov dependence. Specifically, suppose that $v_1, \ldots, v_k$ is a k-length stationary random walk of a regular graph $G = (\mathcal{V}, E)$. Additionally, assume a scalar valued function $F : [\mathcal{V}] \mapsto \mathbb{C}$. Then, even though the random variables v_i are not independent, one has

$$\mathbb{P}\left[\left|\frac{1}{k}\sum_{i=1}^{k} F(v_i) - \mathbb{E}[F]\right| > \epsilon\right] \leq 2 \cdot \exp(-\Omega((1-\lambda)k\epsilon^2)), \tag{2}$$

where $1 - \lambda$ is the spectral gap of the transition matrix of the random walk induced on the graph G.

Another important generalization Chernoff's bound is that for matrix-valued IID random variables (Rudelson, 1999; Ahlswede and Winter, 2002; Tropp et al., 2015). In particular, if $X_1, \ldots, X_k$ are IID $d \times d$ complex Hermitian random matrices with $\|X_i\| \leq 1$, then we have

$$\mathbb{P}\left[\left\|\frac{1}{k}\sum_{i=1}^{k} X_i - \mathbb{E}[X]\right\| > \epsilon\right] \leq 2d \cdot e^{-\Omega(k\epsilon^2)}, \tag{3}$$

which is optimal. Note that this bound only differs from that for the scalar-valued case by a factor of d.

1.1 Motivation

Matrix-valued random variables are at the core of a variety of applications, including approximation algorithms for classical combinatorial problems such as the Traveling Salesman Problem (TSP) (Gharan et al., 2011) and the Asymmetric Traveling Salesman Problem (ATSP) (Asadpour et al., 2010), solving Laplacian linear equations (Kyng and Sachdeva, 2016), solving Covering Semidefinte Programs (SDPs) (**?**), and constructions of a class of expander graphs (**?**). Crucially, the efficient de-randomization of randomized results in these applications may leverage the use of matrix-valued random variables in dependent settings.

Concentration of measure in scalar-valued dependent settings is a mature area of research (Schmidt et al., 1995; Kahale, 1997; Gillman, 1998). On the other hand, results for the matrix-valued setting have only recently appeared (**?**Tropp, 2011; Tropp et al., 2011; Garg et al., 2018; Kyng and Song, 2018). In this project, we study two such recent results. First, we consider a matrix Chernoff Bound for strongly Rayleigh distributions, due to Kyng and Zhao (Kyng and Song, 2018). Strongly Rayleigh distributions are of special interest as they exhibit negative correlation, which intuitively suggests good concentration. Moreover, these distributions have found applications in algorithm design (Gharan et al., 2011; Asadpour et al., 2010). Their matrix Chernoff bound also implies results in spectral sparsification of graphs by sampling random spanning trees, which follow a strongly Rayleigh distribution. It is known that the union of two uniformly random spanning trees of the complete graph has constant vertex expansion with high probability (Goyal et al., 2009), leading us to the second result studied. In particular, we study a concentration result for matrix-valued expander random walks, due to Garg, Lee, Song, and Srivastava (Garg et al., 2018). Their result generalizes both Alhswede and Winter's (Ahlswede and Winter, 2002) bound for IID matrix-valued random variables and Gilman's (Gillman, 1998) bound for scalar-valued expander random walks. The main technique of (Garg et al., 2018) is a new multi-matrix extension of the Golden-Thompson inequality derived through complex analysis of sub-harmonic functions, thus confirming a conjecture by Wigderson and Xiao (Wigderson and Xiao, 2005).

The goal of the project is two fold. Firstly, we aim to become familiar with the techniques used in developing concentration inequalities for dependent and matrix-valued random variables. Secondly, we are interested in developing new applications of these results, and their connections to open problems.

2 Notation

We will consider graphs G with vertex set $\mathcal{V}$ and edge set E. We will adhere to $|\mathcal{V}| = n$. Occasionally we will consider a pseudo vertex set $\tilde{\mathcal{V}}$, which is a copy of $\mathcal{V}$. We will use $z_1, \ldots, z_k$ to denote a k-step random walk on an expander graph G. Similarly, we will denote a random walk with rejections by $z'_1, \ldots, z'_k$. Throughout this work, the norm of a matrix refers to its spectral norm.

3 A Matrix Chernoff Bound for Strongly Rayleigh Distributions (Kyng and Song, 2018)

3.1 Preliminaries

Let $x = (x_1, \cdots, x_n) \in \{0,1\}^n$ be a random vector with probability measure $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$. The generating polynomial f_μ of μ is defined as

$$f_\mu(x) = \sum_{S \subseteq [n]} \mu(S) \cdot \prod_{i \in S} x_i. \tag{4}$$

The probability measure μ is said to be k-homogeneous if f_μ is a homogeneous polynomial of degree k. Moreover, f_μ is said to be real if all of its coefficients are real, which is always true by definition of μ. Lastly, f_μ is said to be stable if $\mathrm{Im}(x_i) > 0$ for all $i \in [n]$ implies that $f_\mu(x) \neq 0$.

Definition 1 (Strongly Rayleigh Distribution)**.** *A probability measure μ is a strongly Rayleigh distribution if f_μ is real stable.*

Strongly Rayleigh distributions are of special interest because they are a class of discrete distributions that exhibit negative dependence. Crucially, strongly Rayleigh distributions are preserved under conditioning and marginalization. In particular, suppose x follows a k-homogeneous strongly Rayleigh distribution. We have the following facts.

Fact 1 (Conditioning Property)**.** Suppose we have $b = (b_1, \cdots, b_t) \in \{0,1\}^t$ with l non-zero entries and some $S \subseteq [n]$ such that $|S| = t$. Then, conditional on $x_S = b$, the distribution of $x_{[m] \setminus S}$ is $(k-l)$-homogeneous strongly Rayleigh.

Fact 2 (Stochastic Covering Property)**.** Suppose we have an index $i \in [m]$. Let $x' = x_{[m] \setminus \{i\}}$ be the marginal distribution on the entries $[m] \setminus \{i\}$ and let $x'' = x_{[m] \setminus \{i\}}$ be the distribution on the entries $[m] \setminus \{i\}$ conditioned on $x_i = 1$. Then, there exists a coupling between x' and x'' such that in every outcome of the coupling, the value of x'' can be turned into the value of x' by flipping at most one of its entries from 0 to 1.

The stochastic covering property is an indication that strongly Rayleigh distributions do not change too quickly under conditioning. Intuitively, these facts are essential in the martingale-based proof of the main results discussed in the next section.

Two corollaries of the main result deal with matrix concentration of randomly sampled spanning trees. Given a weighted undirected graph $G = (V, E, w)$, define its Laplacian matrix L_G as follows.

Definition 2 (Laplacian Matrix)**.** *The Laplacian matrix $L_G \in \mathbb{R}^{n \times n}$ of a graph G is defined as*

$$L_G = \sum_{e = \{u,v\} \in E} w_e b_e b_e^T,$$

where b_e is a signed $\{u,v\}$-incidence vector with 1 at u, -1 at v, and 0 otherwise.

For two symmetric matrices $A, B \in \mathbb{R}^{n \times n}$, we write $A \preceq B$ to indicate that $x^T A x \leq x^T B x$, for all $x \in \mathbb{R}^n$. Moreover, A and B are spectrally similar if $B/\alpha \preceq A \preceq \alpha \cdot B$. Naturally, the notion of matrix concentration of graphs considered herein is with respect to their Laplacian matrices.

Random spanning trees have been useful in the design and analysis of algorithms. Specifically, let $\mathcal{T}_G$ denote the set of all spanning trees of G and consider the following distribution.

Definition 3 (w-uniform Spanning Tree Distribution). *A w-uniform distribution on $\mathcal{T}_G$, denoted by $\mathcal{D}_G$, is a probability distribution such that*

$$\mathbb{P}_{X\sim\mathcal{D}_G}[X=T]\propto\prod_{e\in T}w_e.$$

Note that this can be sampled faster than matrix multiplication (Durfee et al., 2017).

Fact 3 (Edge Marginals). The probability $\mathbb{P}[e]$ that an edge $e=\{u,v\}\in E$ appears in a tree samples according to $\mathcal{D}_G$ is given by

$$\mathbb{P}[e]=w_e b_e^T L_G^{\dagger} b_e,$$

where $L_G^{\dagger}$ is the pseudo-inverse of L_G. This is referred to as the leverage score l_e of $e\in E$.

Fact 3 is of much importance as it provides the freedom of tuning w to produce the edge marginals desired for the specific applications of random spanning tree sampling. Lastly, in connection to matrix concentration, the following fact is essential for the corollaries presented.

Fact 4. If G is connected, $\mathcal{D}_G$ is $(n-1)$-homogeneous strongly Rayleigh.

3.2 Main Results

The main result relies on the following lemma, whose inductive proof relies on Fact 1.
Lemma 1 (Shrinking Marginals Lemma). Suppose $\xi=(\xi_1,\cdots,\xi_m)\in\{0,1\}^m$ is a random vector that follows a k-homogeneous strongly Rayleigh distribution. Then, for any set $S\subseteq[m]$ with $|S|\le k$ and all $j\in[m]\setminus S$ we have

$$\mathbb{P}[\xi_j=1|\xi_S=1_S]\le\mathbb{P}[\xi_j=1].$$

Theorem 1 (Restatement of Theorem **??**). *Let $\xi=(\xi_1,\cdots,\xi_m)\in\{0,1\}^m$ be a random vector of $\{0,1\}$ variables whose distribution is k-homogeneous strongly Rayleigh. Let $A_1,\cdots,A_m\in\mathbb{R}^{n\times n}$ be a collection of PSD matrices such that $\|\mathbb{E}[\sum_e\xi_eA_e]\|\le\mu$ and $\|A_e\|\le R$ for all $e\in[m]$. Then, for any $\epsilon>0$ we have*

$$\mathbb{P}\left[\left\|\sum_e\xi_eA_e-\mathbb{E}\left[\sum_e\xi_eA_e\right]\right\|\ge\epsilon\mu\right]\le n\cdot\exp\left(-\frac{\epsilon^2\mu}{R(\log k+\epsilon)}\Theta(1)\right).$$

Proof sketch. For all $e\in[m]$, associate a PSD matrix satisfying $\|Y_e\|\le R$. The random vector ξ may equivalently be represented as a sequence $\gamma=\gamma_1,\cdots,\gamma_k$, each representing the indices of the entries that take value one in ξ. Thus, we have $Z=\sum_{e\in[m]}\xi_eY_e=\sum_{i=1}^kY_{\gamma_i}$. Now, define the martingale sequence given by $M_0=\mathbb{E}[Z]$ and $M_i=\mathbb{E}_{\gamma\ge i+1}[Z|\gamma_1,\cdots,\gamma_i]$. Denote the difference sequence by $X_{i+1}=M_{i+1}-M_i$. Via a coupling argument that relies on Fact 2, together with Lemma 1, we may bound: *i)* $\lambda_{\max}(X_i)$ and *ii)* the predictable quadratic variation $W_i=\sum_{j=1}^i\mathbb{E}_{j-1}[X_j^2]$. Lastly, we may use these bounds to apply the Matrix Freedman Inequality (Tropp et al., 2011) and complete the proof. □

Theorem 2. *Given a graph $G = (V, E, w)$ with $|V| = n$, let T be a random spanning tree. Then, with probability at least $1 - 1/poly(n)$ we have*

$$L_T \preceq O(\log n) L_G.$$

Proof sketch. Sample $T = (V, E') \sim \mathcal{D}_G$ with an associated random vector ξ, which by Fact 4 is $(n-1)$-homogeneous strongly Rayleigh. Let the weight w' of the edges in E' be given by $w'_e = w_e/l_e$. To apply Theorem 1, let ξ_e be the eth entry of ξ and let $A_e = (L_G^\dagger)^{1/2} L'_e (L_G^\dagger)^{1/2}$, where L'_e is the Laplacian of the subgraph of T induced by $e \in E'$. Then, we may apply Theorem 1 with $R = 1$, $\mu = 1$, and $\epsilon = 100 \log n$. □

Theorem 3. *Given a connected graph $G = (V, E, w)$ with $|V| = n$ and a parameter $\epsilon > 0$, let $T_1, \cdots, T_t$ denote t independent inverse leverage score weighted random spanning trees. If we choose $t = O(\epsilon^{-2} \log^2 n)$, then with probability $1 - 1/poly(n)$ we have*

$$(1-\epsilon) L_G \preceq \frac{1}{t} \sum_{i=1}^{t} L_{T_i} \preceq (1+\epsilon) L_G.$$

Proof sketch. The proof is similar to that of Theorem 3, except the edges of the $t = O(\epsilon^{-2} \log^2 n)$ independent samples are seen as following a $t(n-1)$-homogeneous strongly Rayleigh distribution; this holds because the product of strongly Rayleigh distribution remains strongly Rayleigh. Then, we apply Theorem 1 with $R = 1/t$ and $\mu = 1$. □

4 A matrix Expander Chernoff Bound (Garg et al., 2018)

Generalizations to Matrix-Random Variables It is natural to wonder whether bounds like Equation 3 could be extended for matrix-valued dependent random variables observed through random walks on a graph (in the same spirit as (Gillman, 1998)). In this section, we will positively answer the question and briefly discuss tools and techniques to derive them.

Theorem 4. *Let $G = (\mathcal{V}, E)$ be a regular undirected graph whose transition matrix has second eigenvalue λ, and let $F : V \to \mathbb{C}^{d \times d}$ be a function such that:*

1. *For each $v \in V$, $F(v)$ is Hermitian and $\|F(v)\| \leq 1$.*
2. *$\sum_{v \in V} F(v) = 0$.*

Then, for a stationary random walk $v_1, \ldots, v_k$ with $\epsilon \in (0, 1)$ we have:

$$\mathbb{P}\left[\lambda_{\max}\left(\frac{1}{k}\sum_{j=1}^{k} F(v_j)\right) \geq \quad \epsilon\right] \leq d \cdot \exp\left(-\Omega\left(\epsilon^2 (1-\lambda) k\right)\right),$$
$$\mathbb{P}\left[\lambda_{\min}\left(\frac{1}{k}\sum_{j=1}^{k} F(v_j)\right) \leq -\epsilon\right] \leq d \cdot \exp\left(-\Omega\left(\epsilon^2 (1-\lambda) k\right)\right). \tag{5}$$

This theorem adds to the amazingly long list of pseudorandom properties of expander graphs, and has numerous applications, a few of which are discussed in section 4.3.

History Interestingly, an attempt to solve this problem was made by (Wigderson and Xiao, 2005) in 2005, but unfortunately a bug was discovered in the proof a year later and the gap could not be fixed. Following which, an alternate proof was provided for the applications studied in (Wigderson and Xiao, 2005) using "the method of pessimistic estimators" (discussed in more details in Section A.1). This is the first such attempt to positively answer the conjecture due to Wigderson and Xiao, about existence of a chernoff-type (as (5)) for the matrix-valued dependent setting.

4.1 Proof Sketch

Let us first recall how the usual scalar Chernoff bound is proved, in both the iid and expander random walk setting. The key observation is that if $X_1, \ldots, X_k$ are independent random variables (for eg in (1)), then the moment generating function of the sum is equal to the product of the moment generating functions:

$$\mathbb{E}\left[\exp\left(t\sum_{i=1}^{k} X_i\right)\right] = \prod_{i=1}^{k} \mathbb{E}[\exp(tX_i)].$$

This is no longer true in case where the X_i come from a random walk (for the case in (2)), but we still have the algebraic fact that

$$\exp\left(t\sum_{i=1}^{k} X_i\right) = \prod_{i=1}^{k} \exp(tX_i), \tag{6}$$

which allows one to decompose the sum as a product, allowing an independent analysis for each step of the random walk.

The analogue of the moment generating function in the matrix setting is

$$\mathbb{E}\left[\operatorname{tr}\left[\exp\left(t\sum_{i=1}^{k} X_i\right)\right]\right],$$

and the main difficulty is that (6) no longer holds if the matrices X_i do not commute. A substitute of this fact is given by the *Golden-Thompson inequality*, but with limited applications. In fact, the old proof due to (Wigderson and Xiao, 2005) was based on an erroneous extension of the Golden-Thompson inequality to more than two matrices (see Section A.2 for more details). The main ingredient of the new proof is a multi-matrix generalization of the Golden-Thomspon inequality (see Section A.3 for more details), as follows:

Theorem 5 (Bounded Multi-matrix Golden-Thompson inequality)**.** *Let $H_1, \ldots, H_k \in \mathbb{C}^{d\times d}$ be Hermitian matrices. Then*

$$\log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j\right)\right]\right) \le \frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \log\left(\operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\imath\phi}}{2}H_j\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\imath\phi}}{2}H_j\right)\right]\right)\mathrm{d}\mu(\phi)$$

where μ is some probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Proof Sketch of Theorem 4

Proof. Due to symmetry, it suffices to prove just one of the statements. Let $t > 0$ be a parameter to be chosen later. Then

$$\mathbb{P}\left[\lambda_{\max}\left(\sum_{j=1}^{k} f(v_j)\right) \geq k\epsilon\right] \leq \mathbb{P}\left[\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \geq \exp(tk\epsilon)\right]$$

$$\text{((using Markov's Inequality))}$$

$$\leq \frac{\mathbb{E}\left[\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right]\right]}{\exp(tk\epsilon)} \tag{7}$$

Using Theorem 5 and the concavity of log, we bound the RHS above as:

$$\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \leq \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\imath\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\imath\phi}}{2}tf(v_j)\right)\right]\mathrm{d}\mu(\phi)\right)^{\frac{4}{\pi}}$$

Further nothing using the Jensen's inequality:

$$\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \geq d^{4/\pi-1}\left(\operatorname{tr}\left[\exp\left(\frac{\pi}{4}t\sum_{j=1}^{k} f(v_j)\right)\right]\right)^{\frac{4}{\pi}}$$

we get,

$$\operatorname{tr}\left[\exp\left(\frac{\pi}{4}t\sum_{j=1}^{k} f(v_j)\right)\right] \leq d^{1-\pi/4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\imath\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\imath\phi}}{2}tf(v_j)\right)\right]\mathrm{d}\mu(\phi) \tag{8}$$

The matrix-expander chernoff bound follows by using (8) along with Lemma 2 and then appropriately trading off the parameter t. □

Lemma 2. Let G be a regular λ-expander on V, let f be a function $f : V \to \mathbb{C}^{d\times d}$ and $\sum_{v\in V} f(v) = 0$, let $v_1, \cdots, v_k$ be a stationary random walk on G, for any $t > 0$, $\gamma \geq 0$, $b > 0$, $t^2(\gamma^2 + b^2) \leq 1$, and $t\gamma \leq \frac{1-\lambda}{4\lambda}$ we have

$$\mathbb{E}\left[\operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{tf(v_j)(\gamma+\imath b)}{2}\right)\prod_{j=k}^{1}\exp\left(\frac{tf(v_j)(\gamma-\imath b)}{2}\right)\right]\right] \leq d\cdot\exp\left(kt^2(\gamma^2+b^2)\left(1+\frac{8}{1-\lambda}\right)\right).$$

Proof. The proof of this lemma is analogous to the main lemma in (Healy, 2008) for the scalar valued expander Chernoff bound, which proceeds by controlling how much does a uniform vector shears on repeated application of the transition matrix (followed by the cost matrix) defining the random walk (similar to Lemma 4). □

4.2 Extension: Random Walk with Rejection Sampling

The Chernoff type bounds for random walks on an expander (eg. (2) and Theorem 4) are already quite strong and have numerous applications for efficient de-randomization. However, the constructions of λ–expanding graphs have, unfortunately, been restricted to constant degree graphs in the literature. **Fact 5.** Stationary distributions for constant degree graphs is a uniform distribution over the vertices.

This restricts us to only handle deviations of X from $\frac{1}{n}\sum_{i=1}^{n} f(v_i)$ by running a random walk. However, in many applications, we are interested in controlling deviations of the rewards accumulated along a random walk w.r.t. $\frac{1}{n}\sum_{i=1}^{n} y_i f(v_i)$ where $y_i \in [0,1]$. This motivates us to look at analyzing a random walk (with rejection) on a constant degree graph, where at each vertex v_i, the reward is only collected with probability y_i, and is ignored otherwise (described by 1). Interestingly, we can derive a similar chernoff-type bound for a random walk with rejection sampling but with a multiplicate loss of $\frac{1}{\sigma}(\approx n)$ in the exponential tail.

Theorem 6. *Let $G(\mathcal{V}, E)$ be a constant degree λ-spectral expander with $|\mathcal{V}| = n$. Let us further assume that we are given constants $y_1, \ldots, y_n \in [0,1]$. Additionally, we are given a scalar valued function $f : \mathcal{V} \mapsto [-1,1]$. Then for a random walk with rejectoin, $z_1, \ldots, z_k$, as defined by Algorithm 1, the following Chernoff-type bound holds:*

$$Pr[|X - \rho k| \geq \epsilon k] \leq \frac{2}{\sigma} e^{-\frac{k(1-\lambda)\epsilon^2}{4}}$$

where $\sigma = \frac{1}{n}\min(y_1, y_2, \ldots, y_n, 1 - y_1, \ldots, 1 - y_n)$.

The proof for is provided in Section A.4.

Algorithm 1 Random walk with Rejection Sampling

Require: Given: $(y_1, y_2, \ldots, y_n) \in [0,1]$– Rejection probabilities at the vertices.
Require: A constant degree graph $G(\mathcal{V}, E)$ with spectral expansion λ.

$X \leftarrow 0$
$z_0 \sim \text{Uniform}[n]$
for $t = 1, \ldots k$ **do**
  $z_t \leftarrow \text{Random\_step}(z_{t-1})$
  Let $X_i \coloneqq \begin{cases} f(z_t) & w.p. \quad q[z_t] \\ 0 & w.p. \quad 1 - q[z_t] \end{cases}$
  $X += X_i$
end for
return X

4.3 Applications

4.3.1 $O(\log n)$ expanding generators for any group

One of the applications salvaged in (**?**) is the deterministic construction of a $O(\log n)$-sized expanding generator set for any group. Let H be a group and $S \subset H$ be a generating multi-set. The Cayley graph Cay$(H;S)$ on H and S is a graph whose vertex set is H and where (h,h') is in the edge set if $h' = hs$. Require S to be symmetric, that is $s \in S$ if, and only if, $s^{-1} \in S$, making the graph undirected. Let A be the normalized adjacency matrix of the graph and let $\lambda(\mathrm{Cay}(H;S))$ denote the second largest eigenvalue of the graph. Their main theorem is as follows.

Theorem 7. *Fix $\gamma < 1$. Then, there exists a algorithm running in time poly(n) that, given any group H of order n, constructs a symmetric set $S \subseteq H$ of size $|S| = O(\frac{1}{\gamma^2}\log n)$ such that $\lambda(Cay(H;S)) \leq \gamma$.*

Their proof relies on an existential result of such generating sets (**?**). In particular, for any $h \in H$, let the matrix-valued function be $f(h) = \frac{1}{2}(I - J/n)(P_h + P_{h^{-1}})$, where J is the all ones matrix and P_h is the permutation matrix representing the action of h on any element of the group. Then, one can show that for all $h \in H$, $\|f(h)\| \leq 1$ and $\mathbb{E}_{h \in H}[f(h)] = 0$, which is ready for the application of their pessimistic estimators de-randomization of (Ahlswede and Winter, 2002), which has probability of error < 1 when $k = O(\frac{1}{\gamma^2}\log n)$.

The same result may be obtained by considering an expander $G = (V,E)$ with a bijection between V and H together with the bounds in Theorem 4, bypassing the need for the pessimistic estimators. This may turn out to be more efficient, as even though one will need to take more samples when considering de-randomization by evaluating all random walks on the expander, performing each step is comparatively cheaper than sampling all of them.

4.3.2 Approximation Algorithms for Covering SDP's

More recently, a more general class of relaxations, semi-definite programs (SDPs), have been used by computer scientists (e.g. (Goemans and Williamson, 1995) and (Arora et al., 2009)) to give better approximation guarantees to NP-hard problems. SDPs may be solved in polynomial time (using e.g. the ellipsoid method or interior-point methods), and again the solution may be rounded to give an integral solution to the original problem. In this application, we study a restricted class of SDPs of the following form: given $c \in [0,1]^n$ and $f : [n] \mapsto [0, \mathbb{I}_d]$, find y where,

$$\begin{aligned} &\text{minimize } \langle c, y\rangle \\ &\qquad\text{with feasibility constraint} \sum_{i=1}^{n} y_i f(v_i) \geq \mathbb{I} \text{ and } y \in \mathbb{N}^n \end{aligned} \tag{9}$$

This is relaxed into a covering SDP by allowing $y \in \mathbb{R}^n_+$, and then an integral solution is found by rounding the proposed solution, as stated in the following theorem:

Theorem 8. *Suppose we have a program as in (9) and suppose we have a feasible solution vector $y^* \in \mathbb{R}^n_+$. Then, we can find in time poly(n,d) a feasible integral solution $\hat{y}$ such that*

$$\langle c, \hat{y}\rangle \leq O(\log d)\langle c, y^*\rangle$$

Proof. Let us denote the optimal value of the relaxed covering SDP by OPT. The classical rounding techniques follows as below:

1. Take $8\|y^*\|_1 \log(2d)$ iid samples from the distribution induced by y^* and treat them as a bag of samples $\hat{y}$ (i.e. $\hat{y}[i]$ is equal to the number of times the i^{th} coordinate was sampled). Then,

with probability strictly greater than 0, the proposed solution $\hat{y}$ is feasible and satisfies the constraint

$$\langle \hat{y}, c \rangle \leq O(\log n)\langle y^*, c \rangle$$

2. Derandomize the above probabilistic construction using the method of pessimistic estimator (discussed more in Section A.1) to get a deterministic feasible solution $\hat{y}$ such that $\langle \hat{y}, c \rangle \leq O(\log n)\langle y^*, c \rangle$.

We will be looking at an alternative (and more efficient) de-randomization strategy using expander walks with rejection sampling as follows:

1. Perform a $k = 8 \cdot \frac{\|y^*\|_1}{(1-\lambda)} \cdot \log(\frac{2d}{\sigma})$ length random walk on a λ-expander with rejection sampling with probabilities defined by $\frac{y_i}{\|y^*\|_1}$ at the i^{th} node. Treat the collected nodes as a bag of samples $\hat{y}$ (i.e. $\hat{y}[i]$ is equal to the number of times the i^{th} coordinate was seen and selected in the random walk). Then, with probability strictly greater than 0, the proposed solution $\hat{y}$ is feasible and satisfies the constraint

$$\langle \hat{y}, c \rangle \leq O(\log n)\langle y^*, c \rangle$$

2. Derandomize the above probabilistic construction by evaluating all possible $k-$length walks on the expander graph G to get a deterministic feasible solution $\hat{y}$ such that $\langle \hat{y}, c \rangle \leq O(\log n)\langle y^*, c \rangle$.

$\square$

The second approch uses strictly more samples as compared to the first approach but getting each one of them is cheaper, as one only needs $O(\log d)$ bits of randomness to walk on the graph whereas one would need entropy(y^*) pure random bits for sampling and derandomization. The second approach can be strictly more efficient than the first approach in the case of well behaving y^*.

Conclusions For this project, we explored various tools and techniques to derive strong concentration bounds in matrix-valued dependent settings and their applications for de-randomization and in theoretical computer science. We also extended the previous results for a new concentration inequality for random walks on expander graphs with rejection. We are particularly excited in exploring the applications of random walks on an expander as pseudo-random generators, and look forward to working on it in the future.

A Additional Results

A.1 Method of Pessimistic Estimators

The method of pessimistic estimators was first introduced by Raghavan (**?**). In essence, given a random variable X, the method consists of transforming an existential statement of the occurrence of some event $\sigma(X)$, i.e., $\mathbb{P}[\sigma(X) = 1] > 0$, into an efficient and deterministic procedure that finds some $x \in \operatorname{supp}(X)$ such that $\sigma(x) = 1$. This method may be applied, for instance, in the derandomization of probabilistic algorithms and constructions.

Wigderson and Xiao (**?**) take this approach to salvage the applications described in (Wigderson and Xiao, 2005), which was later retracted (Wigderson and Xiao, 2006) due to a fatal flaw in the proof of their main theorem. If $\vec{X} = (X_1, \dots, X_k)$ denotes a random variable, the main observation is that by averaging, for any i, there must exists some x_i such that

$$\mathbb{P}[\sigma(\vec{X}) = 0 | X_1 = x_1, \dots, X_i = x_i] \leq \mathbb{E}_{X_1,\dots,X_i}\left[\mathbb{P}[\sigma(\vec{X}) = 0 | X_1 = x_1, \dots, X_i = x_i]\right]. \tag{10}$$

By the existential statement, this chain of reasoning leads to $\mathbb{P}[\sigma(\vec{x})] \leq \mathbb{P}[\sigma(\vec{X}) = 0]$. Note that $\vec{x}$ is a fixed vector and that $\mathbb{P}[\sigma(\vec{X}) = 0] < 1$, meaning that $\sigma(\vec{x}) = 1$. The main difficulty in turning this observation into an algorithm lies in computing $\mathbb{P}[\sigma(\vec{X}) = 0 | X_1 = x_1, \dots, X_i = x_i]$. To avoid this, the method of pessimistic estimators produces estimators $\phi_0, \dots, \phi_k : [n]^i \to [0,1]$ such that for any i and any fixed $x_1, \dots, x_i \in [n]$ we have that

$$\mathbb{P}[\sigma(\vec{X}) = 0 | X_1 = x_1, \dots, X_i = x_i] \leq \phi_i(x_1, \dots, x_i), \tag{11}$$

together with

$$E_{X_{i+1}} \phi_{i+1}(x_1, \dots, x_i, X_{i+1}) \leq \phi_i(x_1, \dots, x_i). \tag{12}$$

Obviously, we require $\phi_0 < 1$ and $\phi_1, \dots, \phi_k$ to be efficiently computable.

Thus, the main contribution of (**?**) is the design of pessimistic estimators to de-randomize matrix-valued Chernoff bounds (Ahlswede and Winter, 2002). In turn, this allows them to de-randomize a construction of $O(\log n)$ expanders as well as an approximation algorithm for covering SDPs. However, as noted by Garg et al. (Garg et al., 2018), their method requires an efficient approach to compute matrix moment generating functions, which is problem dependent, and hence their result does not constitute a black box derandomization of matrix-valued Chernoff bounds.

A.2 Golden Thompson Inequality

Golden-Thompson inequality (due to (Golden, 1965) and (Thompson, 1965)) states that for any Hermitian A, B:

$$\operatorname{tr}[\exp(A + B)] \leq \operatorname{tr}[\exp(A)\exp(B)]. \tag{13}$$

The latter expression may further be bounded by $\|\exp(A)\| \operatorname{tr}[\exp(B)]$, and this is sufficient to prove (3) in the independent case as is done in (Ahlswede and Winter, 2002), where an inductive application of it yields

$$\mathbb{E}\left[\operatorname{tr}\left[\exp\left(t\sum_{i=1}^{k} X_i\right)\right]\right] \leq \operatorname{tr}[I] \cdot \prod_{i=1}^{k} \|\mathbb{E}[\exp(tX_i)]\|.$$

However, this approach is too crude to handle the Markov case, roughly because in the absence of inependence, passing to the norm makes it difficult to utilize the fact that the expectation of each X_i is zero.

The original proof of Wigderson-Xiao was based on the following plausible multi-matrix generalization of (13):

$$\operatorname{tr}\left[\exp\left(\sum_{i=1}^{k} A_i\right)\right] \le \operatorname{tr}\left[\prod_{i=1}^{k}\exp(A_i)\right],$$

which turns out to be false for $k > 2$, as demonstrated by the following counter-example. This lead to a fatal gap in their proof.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 2 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

A.3 Bounded Multi-Matrix Golden Thompson Inequality

Here we provide intuition and proof sketch for the new multi-matrix golden thompson inequality.
Theorem 9 (Theorem 5 repeated)**.** *Let $H_1, \ldots, H_k \in \mathbb{C}^{d\times d}$ be Hermitian matrices. Then*

$$\log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j\right)\right]\right) \le \frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \log\left(\operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\imath\phi}}{2}H_j\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\imath\phi}}{2}H_j\right)\right]\right) \mathrm{d}\mu(\phi)$$

where μ is some probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

The proof relies on Hirschman's interpolation theorem (see Lemma 3) in complex analysis, which intuitively states that the value of a holomorphic function at any point inside a domain is a weighted average of the values that the function takes on the boundary (see (Sutter et al., 2017) for a more detailed discussion).

Proof. Define $G(z) \coloneqq \prod_{j=1}^{k}\exp\left(\frac{z}{2}H_j\right)$. We thus have,

$$\begin{aligned}
\log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j\right)\right]\right) &= \log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j/2 + \sum_{j=k}^{1} H_j/2\right)\right]\right) \\
&\qquad \text{(by the Lie-Trotter formula)} \\
&= \log\left(\operatorname{tr}\left[\lim_{\theta\to 0^+}(G(\theta)G(\theta)^*)^{1/\theta}\right]\right) \\
&\qquad \text{(since the } H_j \text{ are Hermitian and log is continuous away from 0)} \\
&= \lim_{\theta\to 0^+}\frac{2}{\theta}\log\|G(\theta)\|_{2/\theta} \\
&\qquad \text{(by Lemma 3 with } p_0 = \infty \text{ and } p_1 = 2)
\end{aligned}$$

$$\leq \lim_{\theta\to 0^+} 2\left(\frac{4}{\pi} + O(\theta)\right) \int_{-\pi/2}^{\pi/2} \log \|G(e^{\mathbf{i}\phi})\|_2 \mathrm{d}\mu_\theta(\phi)$$
$$= \lim_{\theta\to 0^+} \left(\frac{4}{\pi} + O(\theta)\right) \int_{-\pi/2}^{\pi/2} \log\left(\mathrm{tr}\left[G(e^{\mathbf{i}\phi})G(e^{\mathbf{i}\phi})^*\right]\right) \mathrm{d}\mu_\theta(\phi),$$

When $\theta \to 0^+$, $\mu_\theta(\phi)$ converges to some probability distribution μ, thus completing the proof. □

We state without proving an auxillary lemma used in the proof of Theorem 9. The proof of the above lemma requires heavy complex analysis and is out of the scope of this report. We recommend the reader to look at the proof of Theorem 3.3 (Garg et al., 2018) for more details.

Lemma 3 (Riesz-Thorin-type inequality, Also see Theorem 3.1 in (Sutter et al., 2017)). Let $S = \{z \in \mathbb{C} : |z| \leq 1 \text{ and } \mathrm{Re}(z) \geq 0\}$ and let G be a holomorphic map from S to square matrices. Further, let $p_0 \geq p_1 \in [1, \infty]$, for $\theta \in (0, 1)$, define p_θ by

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}.$$

If $z \to \|G(z)\|_{p_{\mathrm{Re}(z)}}$ is uniformly bounded on S and $\|G(\mathbf{i}t)\|_{p_0} \leq 1$, then for any $0 \leq \theta \leq 1/4$,

$$\log \|G(\theta)\|_{p_\theta} \leq \left(\frac{4\theta}{\pi} + O(\theta^2)\right) \int_{-\pi/2}^{\pi/2} \log \|G(e^{\mathbf{i}\phi})\|_{p_1} \mathrm{d}\mu_\theta(\phi)$$

where μ_θ is some probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ depending only on θ.

A.4 Proof of Theorem 6

Our proof is similar to the analysis of Theorem-1 from (Healy, 2008) with some changes to handle rejection sampling.

We first define some additional notation for analysis

Let A_G be the normalized adjacency matrix for the graph G. Define pseudo-vertices $\tilde{\mathcal{V}}$ such that $f(\tilde{v}_i) = 0$ for all $i \in [n]$. We can thus analyze our random walk by looking at an equivalent random walk defined on the graph induced on $\mathcal{V} \cup \tilde{\mathcal{V}}$. In the extension of the graph G with pseudo-vertices, number $\tilde{v}_i$ as $n + i$ respectively, and define the transition matrix P as follows:

$$P[i,j] = \begin{cases} y_1 \cdot A_G[i,j] & \text{if } i \in \mathcal{V} \text{ and } j \in \mathcal{V} \\ (1-y_1) \cdot A_G[i,j-n] & \text{if } i \in \mathcal{V} \text{ and } j \in \tilde{\mathcal{V}} \\ y_1 \cdot A_G[i-n,j] & \text{if } i \in \tilde{\mathcal{V}} \text{ and } j \in \mathcal{V} \\ (1-y_1) \cdot A_G[i-n,j-n] & \text{if } i \in \tilde{\mathcal{V}} \text{ and } j \in \tilde{\mathcal{V}} \end{cases}$$

Additionally, also define a rewards matrix E:

$$E[i,j] = \begin{cases} e^{rf(v_i)} & \text{if } i = j \leq n \\ 1 & \text{if } i = j > n \\ 0 & \text{otherwise} \end{cases}$$

Under the new notation and setup, we observe that:

1. The stationary distribution of the random walk induced by the transition matrix P is given by

$$\pi = \frac{1}{n}[y_1, \dots, y_n, 1 - y_1, \dots, 1 - y_n]$$

2. Let X be the random variable denoting the total reward accumulated so far (as in Algorithm 1). Aditionally, define b_t as follows:

$$b_t \coloneqq \underbrace{EPEP \dots EP}_{\text{t times}} \pi$$

and let $(b)^{\|} = \langle \pi, b \rangle \pi$ and $b^{\perp} = b - b^{\|}$. We thus have, after k-steps of the random walk:

$$\mathbb{E}\left[e^{rX}\right] = \mathbf{1}^T b_k$$

which can be controlled as

$$\begin{aligned}
\mathbb{E}\left[e^{rX}\right] &= \mathbf{1}^T b_k \\
&= \sum_{i=1}^{n} b_k[i] + \sum_{i=1}^{n} b_k[n+i] \\
&= \frac{1}{\sigma}\left(\sum_{i=1}^{n} \sigma b_k[i] + \sum_{i=1}^{n} \sigma b_k[n+i]\right) \\
&= \frac{1}{\sigma}\pi^T b_k \\
&= \frac{1}{\sigma}\frac{\left\|b_k^{\|}\right\|}{\|\pi\|}
\end{aligned} \tag{14}$$

where $\sigma = \frac{1}{n}\min(y_1, \dots, y_n, 1 - y_1, \dots, 1 - y_n)$.

3. Additionally, $\left\|Pb^{\|}\right\| = \left\|b^{\|}\right\|$ and $\|Pb^{\perp}\| \leq \lambda\|b^{\perp}\|$.

4. Observe that $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot f(v_i)$.

Proof. (Proof of Theorem 6) Let us first start by looking at $\Pr[X \geq \rho k + \epsilon k]$. Using the Markov's inequality, we have:

$$\Pr[X \geq \rho k + \epsilon k] \leq \frac{\mathbb{E}\left[e^{rX}\right]}{e^{r\rho k + r\epsilon k}} \tag{15}$$

Recalling that $\left\|b_0^{\|}\right\| = \|\pi\|$, and using 6 we have:

$$\left\|b_j^{\|}\right\| \leq \|\pi\| \prod_{i=1}^{j} \exp\left\{(e^r - 1)\rho + \frac{\lambda \cdot (e^r - 1)^2}{2(1-\lambda)}\right\}$$

and thus, using this with (14) we have:

$$\mathbb{E}\left[e^{rX}\right] = \frac{1}{\sigma}\frac{\left\|b_k^{\|}\right\|}{\|\pi\|}$$

$$
\begin{aligned}
&\le \frac{1}{\sigma}\prod_{i=1}^{k}\exp\left\{(e^r-1)\rho+\frac{\lambda\cdot(e^r-1)^2}{2(1-\lambda)}\right\}\\
&\le \frac{1}{\sigma}\exp\left\{(e^r-1)\rho k+\frac{\lambda\cdot(e^r-1)^2 k}{2(1-\lambda)}\right\}
\end{aligned}
$$

To simplify this expression, we shall assume that $r \le \frac{1}{2}$ (and thus, $e^r - 1 \le r + 2r^2/3 \le 4r/3$) and we note that $\rho \le 1$. Thus,

$$
\mathbb{E}\left[e^{rX}\right] \le \frac{1}{\sigma}\exp\left\{\left(r+r^2\right)\rho k+\frac{\lambda\cdot(4r/3)^2 k}{2(1-\lambda)}\right\} \tag{16}
$$

Using (16) with (15), we get:

$$
\begin{aligned}
\Pr[X \ge \rho k + \epsilon k] &\le \frac{1}{\sigma}\exp\left\{\left(r+r^2\right)\rho k+\frac{\lambda\cdot(4r/3)^2 k}{2(1-\lambda)}-r\rho k-r\epsilon k\right\}\\
&\le \frac{1}{\sigma}\exp\left\{\left(r^2\right)\rho k+\frac{\lambda\cdot(r)^2 k}{(1-\lambda)}-r\epsilon k\right\}\\
&\le \frac{1}{\sigma}\exp\left\{k\left(r^2\rho+\frac{\lambda r^2}{(1-\lambda)}-r\epsilon\right)\right\}\\
&\le \frac{1}{\sigma}\exp\left\{-\frac{k\epsilon^2(1-\lambda)}{2(\lambda+\rho-\rho\lambda)}\right\}\\
&\le \frac{1}{\sigma}\exp\left\{-\frac{k(1-\lambda)\epsilon^2}{4}\right\}
\end{aligned}
$$

Using the same argument for the starting function to be $1 - f$, we get:

$$
\Pr[|X-\rho k| \ge \epsilon k] \le \frac{2}{\sigma}e^{-\frac{k(1-\lambda)\epsilon^2}{4}}
$$

□

Auxillary Lemma's
The following lemma bounds the length of $b_{i+1}^{\|}$ and $b_{i+1}^{\perp}$ relative to $b_i^{\|}$ and $b_i^{\perp}$.

Lemma 4. (Stretching Bounds) Let the matrices P and E be as defined above. Also, let us define the mean $\rho = \frac{1}{n}\sum_{i=1}^{n} y_i f(v_i)$ and $0 \le r \le \frac{\log(1/\lambda)}{2}$. Then for any $b \in \mathbb{R}^{2n}$:

1. $\left\|\left(EPb^{\|}\right)^{\|}\right\| \le (1+(e^r-1)\rho)\cdot\left\|b^{\|}\right\|$
2. $\left\|\left(EPb^{\|}\right)^{\perp}\right\| \le \frac{e^r-1}{2}\cdot\left\|b^{\|}\right\|$
3. $\left\|\left(EPb^{\perp}\right)^{\|}\right\| \le \frac{e^r-1}{2}\cdot\lambda\cdot\left\|b^{\perp}\right\|$
4. $\left\|\left(EPb^{\perp}\right)^{\perp}\right\| \le \sqrt{\lambda}\left\|b^{\perp}\right\|$

Proof. Recall that $b_0 = \pi$ and $b_{i+1} = (EP)\ b_i$.

1. $\left(EPb^{\parallel}\right)^{\parallel} = (Eb^{\parallel})^{\parallel} = \langle \pi, E\pi \rangle b^{\parallel} \le \langle \mathbf{1}, E\pi \rangle b^{\parallel} = \mathbb{E}_{v\sim\pi}\left[e^{rf(v)}\right] b^{\parallel}$, and using the fact that $e^{rx} \le (1 + (e^r - 1)x)$ for all $x \in [0,1]$, we have:

$$\left\|(EPb^{\parallel})^{\parallel}\right\| = \mathbb{E}_{v\sim\pi}\left[e^{rf(v)}\right]\left\|b^{\parallel}\right\| = \mathbb{E}_v[1 + (e^r - 1)f(v)]\left\|b^{\parallel}\right\| = (1 + (e^r - 1)\rho)\left\|b^{\parallel}\right\|$$

2. Recalling that $(b^{\parallel})^{\perp} = 0$ for all b, we note that for any $\alpha \in \mathbb{R}$,

$$(EPb^{\parallel})^{\perp} = (Eb^{\parallel})^{\perp} = ((E - \alpha \cdot \mathbb{I})b^{\parallel})^{\perp}$$

   Thus, choosing $\alpha = \frac{e^r+1}{2}$ so that $E - \alpha \cdot I$ is diagonal with entries bounded by $\frac{e^r-1}{2}$ in absolute value (since $e^r - \alpha = \frac{e^r-1}{2}$ and $e^0 - \alpha = -\frac{e^r-1}{2}$). Then,

$$\left\|(EPb^{\parallel})^{\perp}\right\| = \left\|(E - \alpha \cdot I)Pb^{\parallel})^{\perp}\right\| \le \frac{e^r - 1}{2}\left\|Pb^{\parallel}\right\|$$

3. The proof is similar to part-2 with the extra property that $\|Pb^{\perp}\| \le \lambda\|b^{\perp}\|$.
4. $\|(EPb^{\perp})^{\perp}\| \le \|EPb^{\perp}\| \le e^r \cdot \|Pb^{\perp}\| \le e^r\lambda\|b^{\perp}\|$, and since we assumed that $r \le \frac{\log(1/\lambda)}{2}$, this is at-most $\sqrt{\lambda} \cdot \|b^{\perp}\|$.

□

We now show that $b_i^{\perp}$ remains short relative to previous $b_j^{\parallel}$.

Lemma 5. $\left\|b_i^{\perp}\right\| \le \frac{e^r-1}{1-\lambda} \cdot \max_{j<i}\left\{\left\|b_j^{\parallel}\right\|\right\}$ for $1 \le i \le k$.

Proof. By the triangle inequality, $\left\|b_i^{\perp}\right\| = \|(EPb_{i-1})^{\perp}\| \le \left\|(EPb_{i-1}^{\perp})^{\perp}\right\| + \left\|(EPb_{i-1}^{\parallel})^{\perp}\right\|$.

Thus, by items 2 and 4 from Lemma 4, we have $\left\|b_i^{\perp}\right\| \le \frac{e^r-1}{2} \cdot \left\|b_{i-1}^{\parallel}\right\| + \sqrt{\lambda} \cdot \left\|b_{i-1}^{\perp}\right\|$.

Recursively applying this bound, and noting that $\|b_o^{\perp}\| = 0$, we have:

$$\left\|b_i^{\perp}\right\| \le \frac{e^r - 1}{2} \cdot \sum_{j=0}^{i-1}(\sqrt{\lambda})^j\left\|b_{i-j+1}^{\parallel}\right\| \le \frac{e^r - 1}{2(1 - \sqrt{\lambda})} \cdot \max_{j<i}\left\{\left\|b_j^{\parallel}\right\|\right\}$$

The lemma follows by noting that $\frac{1}{1-\sqrt{\lambda}} = \frac{1+\sqrt{\lambda}}{1-\lambda} \le \frac{2}{1-\lambda}$ since $\lambda \in [0,1]$. □

We now use Lemma 5 to bound $\left\|b_i^{\parallel}\right\|$ inductively.

Lemma 6. $\left\|b_i^{\parallel}\right\| \le \exp\left\{(e^r - 1)\rho + \frac{\lambda\cdot(e^r-1)^2}{2(1-\lambda)}\right\} \cdot \max_{j<i}\left\{\left\|b_j^{\parallel}\right\|\right\}$, for $1 \le i \le k$

Proof. By the triangle inequality,

$$\left\|b_i^{\parallel}\right\| = \left\|(EPb_{i=1})^{\parallel}\right\| \le \left\|(EPb_{i-1}^{\parallel})^{\parallel}\right\| + \left\|(EPb_{i-1}^{\perp})^{\parallel}\right\|$$

Thus, by items 1 and 3 of Lemma 4, we have $\left\| b_i^{\parallel} \right\| \leq (1 + (e^r - 1)\rho) \cdot \left\| b_{i-1}^{\parallel} \right\| + \frac{e^r - 1}{2} \cdot \lambda \cdot \left\| b_{i-1}^{\perp} \right\|$, and so by Lemma 5, we have,

$$
\begin{aligned}
\left\| b_i^{\parallel} \right\| &\leq (1 + (e^r - 1)\mu) \cdot \left\| b_{i-1}^{\parallel} \right\| + \frac{\lambda \cdot (e^r - 1)^2}{2(1 - \lambda)} \cdot \max_{j < i-1} \left\{ \left\| b_j^{\parallel} \right\| \right\} \\
&\leq \left(1 + (e^r - 1)\rho + \frac{\lambda \cdot (e^r - 1)^2}{2(1 - \lambda)} \right) \cdot \max_{j < i} \left\{ \left\| b_j^{\parallel} \right\| \right\}.
\end{aligned}
$$

Finally using the fact that $1 + x \leq e^x$ for all $x > 0$, we conclude that this is at most,

$$
\exp\left\{ (e^r - 1)\rho + \frac{\lambda \cdot (e^r - 1)^2}{2(1 - \lambda)} \right\} \cdot \max_{j < i} \left\{ \left\| b_j^{\parallel} \right\| \right\}
$$

□

A.5 Conjectures

Matrix-Valued Settings - Random walk with Rejection Since the proof of the matrix-expander chernoff bound in Garg et al. (2018) follows the similar analysis as the proof of Theorem 1 in (Healy, 2008) and thus consequently to our proof of Theorem 6, we believe the following theorem to be true:

Theorem 10. *Let $G(\mathcal{V}, E)$ be a constant degree λ-spectral expander with $|\mathcal{V}| = n$. Let us further assume that we are given scalars $y_1, \ldots, y_n \in [0, 1]$. Additionally, we are given a matrix valued function $F : \mathcal{V} \mapsto [-\mathbb{I}_d, \mathbb{I}_d] \subseteq \mathbb{R}^{d \times d}$. Then for a random walk with rejection, $z_1, \ldots, z_k$, as defined by Algorithm 1, the following Chernoff-type bound holds:*

$$
Pr\left[\lambda_{\max}\left(\frac{1}{k} \sum_{j=1}^{k} F(v_j) - \rho \right) \geq \epsilon \right] \leq \frac{d}{\sigma} e^{-\Omega\left(\epsilon^2 (1-\lambda) k\right)}
$$

$$
Pr\left[\lambda_{\min}\left(\frac{1}{k} \sum_{j=1}^{k} F(v_j) - \rho \right) \leq -\epsilon \right] \leq \frac{d}{\sigma} e^{-\Omega\left(\epsilon^2 (1-\lambda) k\right)}
$$

where $\rho = \frac{1}{n} \sum_{j=1}^{n} y_i F(v_i)$ and $\sigma = \frac{1}{n} \min(y_1, y_2, \ldots, y_n, 1 - y_1, \ldots, 1 - y_n)$.

Multiplicative Chernoff Bound We also suspect that the following multiplicative versions of the chernoff-type bound will follow from similar analysis using the new bound on $\mathbb{E}\left[e^{tX}\right]$.

Theorem 11. *Let $G(\mathcal{V}, E)$ be a constant degree λ-spectral expander with $|\mathcal{V}| = n$. Let us further assume that we are given scalars $y_1, \ldots, y_n \in [0, 1]$. Additionally, we are given a matrix valued function $F : \mathcal{V} \mapsto [-\mathbb{I}_d, \mathbb{I}_d] \subseteq \mathbb{R}^{d \times d}$. Then for a random walk with rejection, $z_1, \ldots, z_k$, as defined by Algorithm 1, the following multiplicative chernoff-type bound holds for all $\gamma \in [0, \frac{1}{2}]$:*

$$Pr\left[\frac{1}{k}\sum_{j=1}^{k}F(v_j)\geq\rho(1+\gamma)\right]\leq\frac{d}{\sigma}e^{-\Omega\left(\gamma^2\rho(1-\lambda)k\right)}$$
$$Pr\left[\frac{1}{k}\sum_{j=1}^{k}F(v_j)\leq\rho(1-\gamma)\right]\leq\frac{d}{\sigma}e^{-\Omega\left(\gamma^2\rho(1-\lambda)k\right)}$$

where $\rho=\frac{1}{n}\sum_{j=1}^{n}y_iF(v_i)$ *and* $\sigma=\frac{1}{n}\min(y_1,y_2,\ldots,y_n,1-y_1,\ldots,1-y_n)$.

References

Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

Arash Asadpour, Michel X Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An o (log n/log log n)-approximation algorithm for the asymmetric traveling salesman problem. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 379–389. SIAM, 2010.

Herman Chernoff et al. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

David Durfee, Rasmus Kyng, John Peebles, Anup B Rao, and Sushant Sachdeva. Sampling random spanning trees faster than matrix multiplication. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 730–742. ACM, 2017.

Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1102–1114. ACM, 2018.

Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 550–559. IEEE, 2011.

David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6): 1115–1145, 1995.

Sidney Golden. Lower bounds for the helmholtz function. *Physical Review*, 137(4B):B1127, 1965.

Navin Goyal, Luis Rademacher, and Santosh Vempala. Expanders via random spanning trees. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 576–585. SIAM, 2009.

Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17(1): 3–37, 2008.

Nabil Kahale. Large deviation bounds for markov chains. *Combinatorics, Probability and Computing*, 6(4):465–474, 1997.

Rasmus Kyng and Sushant Sachdeva. Approximate gaussian elimination for laplacians-fast, sparse, and simple. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 573–582. IEEE, 2016.

Rasmus Kyng and Zhao Song. A matrix chernoff bound for strongly rayleigh distributions and spectral sparsifiers from a few random spanning trees. *arXiv preprint arXiv:1810.08345*, 2018.

Mark Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1): 60–72, 1999.

Jeanette P Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, 1995.

David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352(1):37–58, 2017.

Colin J Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813, 1965.

Joel Tropp et al. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Joel A Tropp. User-friendly tail bounds for matrix martingales. Technical report, CALIFORNIA INST OF TECH PASADENA, 2011.

Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

Avi Wigderson and David Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *null*, pages 397–406. IEEE, 2005.

Avi Wigderson and David Xiao. Retraction of wigderson-xiao: A randomness-efficient sampler for matrix valued functions, eccc tr05-107. 2006.[eccc: Tr05-107]. 3. In *Department of Computer Science Haifa University*. Citeseer, 2006.