Matrix Expander Chernoff Bounds: Derandomization and Applications in Spectral Graph Theory and Quantum Information

Chris Junchi Li[◇]

Department of Electrical Engineering and Computer Sciences[◇]
University of California, Berkeley

October 8, 2024

Abstract

We present new extensions to the classical Chernoff bound in the context of matrix-valued functions, focusing on expander graphs and random walks. Our primary contribution is proving the conjectured *Matrix Expander Chernoff Bound*, which generalizes both scalar Chernoff bounds for random walks and matrix Chernoff bounds for independent variables. This result provides strong concentration guarantees for sums of Hermitian matrix-valued functions sampled along a stationary random walk on an expander graph. Our proof leverages advanced linear algebraic techniques and an improved multi-matrix generalization of the Golden-Thompson inequality. The theoretical results offer significant improvements in randomness efficiency for samplers and derandomization in computational settings such as spectral graph theory and quantum information.

Keywords: Matrix Chernoff bound, Expander graphs, Random walks, Spectral graph theory, Derandomization

1 Introduction

The Chernoff bound is a foundational result in probabilistic combinatorics, providing exponentially small bounds on the tail distributions of sums of independent random variables. This tool has become essential for applications in randomized algorithms, complexity theory, and data science. In recent decades, efforts have been made to generalize the Chernoff bound to more complex settings, such as dependent random variables and matrix-valued functions. Notably, the *Matrix Chernoff Bound*, developed for independent Hermitian random matrices, has become particularly valuable for quantum information theory and numerical linear algebra.

In this paper, we prove a new bound that merges two of these generalizations: the concentration of matrix-valued functions under random walks on expander graphs, leading to a *Matrix Expander Chernoff Bound*. This result, conjectured by Wigderson and Xiao, addresses matrix-valued random variables where the usual independence assumption is replaced by Markov dependence along an expander graph. The challenge lies in handling matrix operations where traditional scalar concentration tools fail, particularly due to non-commutativity and dependence in random walks.

Our proof relies on several key techniques: (1) a new multi-matrix Golden-Thompson inequality, (2) martingale approximation methods, and (3) spectral properties of expander graphs. The results achieved here provide tighter bounds for matrix-valued functions and offer practical improvements for derandomization tasks in both quantum computing and classical spectral graph theory.

Brief Contributions Our contributions can be summarized as follows:

- We prove the Matrix Expander Chernoff Bound, extending the classical scalar and matrix Chernoff bounds to random walks on expander graphs.
- We develop a new multi-matrix Golden-Thompson inequality that is critical in proving the bound in the matrix setting.
- We provide randomness-efficient sampling techniques for matrix-valued functions using strongly explicit expander graphs.

Backgrounds The Chernoff Bound [Che52] is one of the most widely used probabilistic results in computer science. It states that a sum of independent bounded random variables exhibits subgaussian concentration around its mean. In particular, when the random variables are i.i.d. samples from a fixed distribution, it implies that the empirical mean of k samples is ϵ-close to the true mean with exponentially small deviation probability proportional to $e^{-\Omega(k\epsilon^2)}$.

An important generalization of this bound was achieved by Gillman [Gil98] (with refinements later by [Lez98, Kah97, LP04, WX05, Hea08, Wag08, CLLM12, RR17]), who significantly relaxed the independence assumption to Markov dependence. In particular, suppose G is a regular graph with vertex set $V = [n]$, $X : V \to \mathbb{C}$ is a bounded function, and $v_1, \dots, v_k$ is a stationary random walk[1] of length k on G. Then, even though the random variables $X(v_i)$ are in not independent (except when G is the complete graph with self loops), it is shown that:

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

where $1 - \lambda$ is the spectral gap of the transition matrix of the random walk. The gain here is that sampling a stationary random walk of length k on a constant degree graph with constant spectral gap requires $\log(n) + O(k)$ random bits, which is much less than the $k\log(n)$ bits required to produce k independent samples. Since such graphs can be explicitly constructed, this leads to a generic "derandomization" of the Chernoff bound, which has had several important applications (see [WX05] for a detailed discussion). In particular, it leads to the following randomness efficient sampler for scalar-valued functions ([Gil98]) using known strongly explicit constructions of expander graphs [RVW00, LPS88]:

Theorem 1 ([Gil98])**.** *For any $\epsilon > 0$ and $k \geq 1$, there is a* $\mathrm{poly}(r)$*-time computable sampler $\sigma : \{0,1\}^r \to [n]^k$, where $r = \log(n) + O(k)$ s.t. for all functions $f : [n] \to [-1,1]$ satisfying $\mathbb{E}f = 0$, we have that*

$$\mathbb{P}_{w \in_R \{0,1\}^r}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(\sigma(w)_i)\right| \geq \epsilon\right] \leq 2\exp\left(-\Omega\left(-\epsilon^2 k\right)\right)$$

In many applications of interest k is about $\log(n)$, and going from $O(\log^2(n))$ to $O(\log(n))$ random bits leads to a complete derandomization by cycling over all seeds $w \in \{0,1\}^r$.

A different generalization of the Chernoff bound appeared in the works of Rudelson [Rud99], Ahlswede-Winter [AW02], and Tropp [Tro12], who showed that a similar concentration phenomenon is true for *matrix-valued* random variables. In particular, if $X_1, \dots, X_k$ are independent $d \times d$ complex Hermitian random matrices with $\|X_i\| \leq 1$, then the following is true:

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

[1] That is the first vertex v_1 is chosen uniformly at random – which is the stationary distribution of the graph G.

The only difference between this and the usual Chernoff bound is the factor of d in front of the deviation probability; to see that it is necessary, notice that the diagonal case simply corresponds to a direct sum of d arbitrarily correlated instances of the scalar Chernoff bound, so by the union bound the probability should be d times as large in the worst case. This so called "Matrix Chernoff Bound" has seen several applications as well, notably in quantum information theory, numerical linear algebra, and spectral graph theory; the reader may consult e.g. the book [Tro15] for many examples.

We present two different extensions of the above results in this paper.

1.1 A Matrix Expander Chernoff Bound

It is natural to wonder whether there is a common generalization of (1) and (2), i.e., a "Matrix Expander Chernoff Bound". Such a result was conjectured by Wigderson and Xiao in [WX] — in fact, [WX05] contained a proof of it, but the authors later discovered a gap in the proof. In this paper, we prove the Wigderson and Xiao conjecture, namely:

Theorem 2. *Let $G = (V, E)$ 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:*

(i) For each $v \in V$, $f(v)$ is Hermitian and $\|f(v)\| \leq 1$.

(ii) $\sum_{v\in V} f(v) = 0$.

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

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

$$\mathbb{P}\left[\lambda_{\min}\left(\frac{1}{k}\sum_{j=1}^{k} f(v_j)\right) \leq -\epsilon\right] \leq d \cdot \exp\left(-\Omega\left(\epsilon^2(1-\lambda)k\right)\right)$$

This theorem adds to the amazingly long list of pseudorandom properties of expander graphs. By applying the theorem with a strongly explicit bounded degree expander, one obtains the following randomness-efficient sampler for matrix-calued functions conjectured in [WX].

Theorem 3. *For any $\epsilon > 0$, $k \geq 1$ and $d \geq 1$, there is a $\mathrm{poly}(r)$-time computable sampler $\sigma : \{0,1\}^r \to [n]^k$, where $r = \log(n) + O(k)$ s.t. for all functions $f : [n] \to \mathbb{C}^{d\times d}$ satisfying $\mathbb{E}f = 0$ and for each $v \in [n]$, $f(v)$ is Hermitian and $\|f(v)\| \leq 1$, we have that*

$$\mathbb{P}_{w\in_R\{0,1\}^r}\left[\left\|\frac{1}{k}\sum_{i=1}^{k} f(\sigma(w)_i)\right\| \geq \epsilon\right] \leq 2d\exp\left(-\Omega\left(-\epsilon^2 k\right)\right)$$

We remark that while the derandomization applications studied in [WX05] were later recovered in [WX08] using the method of pessimistic estimators, that method requires additional assumptions to be efficiently implementable (specifically, computability of the matrix moment generating function, which is problem-dependent) and therefore does not constitute a truly black box derandomization of the matrix Chernoff bound, whereas Theorem 3 does. Given the increasing ubiquity of applications of this bound, we therefore suspect that it will find further applications in the study of derandomization and expander graphs, beyond the ones mentioned in [WX05].

Techniques

To describe the ideas that go into the proof of Theorem 2, let us begin by recalling how the usual scalar Chernoff bound is proved, in the case when the random variables have mean zero. The key observation is that if $X_1, \ldots, X_k$ are independent random variables, 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, but we still have the algebraic fact that

$$\exp\left(t\sum_{i=1}^{k} X_i\right) = \prod_{i=1}^{k} \exp(tX_i), \tag{3}$$

which allows one to decompose the sum as a product. The latter allows one to consider the steps of the random walk separately and analyze the change in the expectation inductively.

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

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

and the main difficulty is that (3) no longer holds if the matrices X_i do not commute. A substitute for this fact is given by the Golden-Thompson inequality [Gol65, Tho65], which states that for any Hermitian A, B:

$$\mathrm{tr}[\exp(A+B)] \leq \mathrm{tr}[\exp(A)\exp(B)]. \tag{4}$$

The latter expression may further be bounded by $\|\exp(A)\|\mathrm{tr}[\exp(B)]$, and this is sufficient to prove (2) in the independent case as is done in [AW02], where an inductive application of it yields

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

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 (4):

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

which turns out to be false for $k > 2$. To see why, observe that the left hand side is always nonnegative, whereas the right hand side can be the trace of a product of any three positive semidefinite matrices, which can be negative (and this is not the case for two matrices). This led to a fatal gap in their proof.

The main ingredient in our proof is a new multi-matrix generalization of (4), which is inspired by the following statement that was recently proven in [SBT17] (see also [HKT16]).

Theorem 4 (Corollary 3.3 in [SBT17])**.** *Let $H_1, \ldots, H_k \in \mathbb{C}^{d\times d}$ be Hermitian matrices. Then*

$$\log\left[\operatorname{tr}\left(\exp\left(\sum_{j=1}^{k} H_j\right)\right)\right] \leq \int_{-\infty}^{\infty} \log\left[\operatorname{tr}\left(\prod_{j=1}^{k}\exp\left(\frac{H_j(1+\mathbf{i}b)}{2}\right)\prod_{j=k}^{1}\exp\left(\frac{H_j(1-\mathbf{i}b)}{2}\right)\right)\right] \mathrm{d}\mu(b)$$

where μ is some probability distribution on $(-\infty, \infty)$.

The above inequality successfully relates the matrix exponential of a sum to a product of matrix exponentials, but is not adequate for proving an optimal Chernoff bound. The reason is that all known arguments require a Taylor expansion, and Theorem 4 involves integration over an unbounded region (this region can be truncated, but this introduces a loss which leads to a suboptimal bound). To remedy this, we prove a new multi-matrix Golden-Thompson inequality, which only involves integration over a bounded region instead of a line.

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

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

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

We present the proof in Section 3. Theorem 5 is likely to be of independent interest and could have further applications, e.g. in quantum information theory. We draw attention to the following notable features of the above two theorems:

(a) Since $\exp(H(1+\mathbf{i}b)) = \exp(H(1-\mathbf{i}b))^*$ and $\exp\left(\frac{e^{\mathbf{i}\phi}}{2}H_j\right) = \exp\left(\frac{e^{-\mathbf{i}\phi}}{2}H_j\right)^*$ for Hermitian H, the right hand always considers the trace of a matrix times its adjoint, which is always positive semidefinite, ruling out the bad example described above.

(b) They are *average case* inequalities, where the averaging is done over specific distributions. We remark that the first inequality is known to be false in the worst case (i.e., with $b = 0$ and with other small values of b; see [SBT17] for a discussion).

The main point is that Theorem 5 allows one to relate the exponential of a sum of matrices to a (two-sided) product of bounded $d \times d$ matrices and their adjoints. In order to prove Theorem 2, we rewrite this as a one-sided matrix product of $d^2 \times d^2$ matrices acting on $\mathbb{C}^{d\times d}$, by encoding left and right multiplication on this space via a tensor product. However, these matrices are no longer Hermitian (or even normal), so it is difficult to analyze the moment generating function of their product over the random walk using the perturbation-theoretic approach of [WX05]. We surmount this difficulty by employing a variant of the more robust linear algebraic proof technique of Healy [Hea08]. The proof of Theorem 2 is presented in Section 4.

1.2 Martingale Approximation of Expander Walks

While Theorem 2 provides a satisfactory generalization of the expander Chernoff bound to the case when one is interested in the spectral norm of a matrix-valued function on V, one could ask what happens for other matrix norms (such as Schatten norms), or even more generally, for functions taking values in an arbitrary Banach space. Our second contribution is a generic reduction from this problem, of proving concentration for random variables sampled using a Markov chain, to the much more well-studied problem (see e.g. [CL06]) of concentration for sums of *martingale* random variables.

Theorem 6. *Suppose $G = (V, E)$ is a regular graph whose transition matrix has second eigenvalue λ and $f : V \to \mathbb{R}^N$ is a vector-valued function satisfying $\sum_{v \in V} f(v) = 0$ with $F := \sqrt{\sum_{v \in V} \|f(v)\|_2^2}$, where $\|\cdot\|_2$ denotes the Frobenius norm. If $v_1, \ldots, v_k$ is a stationary random walk on G, then for every $\epsilon > 0$, there is a martingale difference sequence $Z_1, \ldots, Z_k$ with respect to the filtration generated by initial segments of $v_1, \ldots, v_k$ such that*

$$\frac{1}{k} \sum_{i=1}^{k} f(v_i) = W + \frac{1}{k} \sum_{i=1}^{k} Z_i$$

where

(i) W is a random vector satisfying $\|W\|_2 \le \epsilon$.

(ii) Each term Z_i satisfies

$$\|Z_i\|_* \le \frac{2 \log(F/\epsilon)}{1 - \lambda} \cdot \max_{v \in V} \|f(v)\|_*$$

for every norm $\|\cdot\|_*$.

Thus, the empirical sums of any bounded (in any norm) function on a graph are well-approximated by a martingale whose increments are also bounded, with a loss in the bound depending on the ℓ_2 norm F of the function and the spectral gap of the graph. Since F will typically scale with the number of vertices, the ratio above is typically comparable to the mixing time.

To see the theorem in action, consider the case when $f(v)$ is matrix-valued in $d \times d$ Hermitian matrices and $\|\cdot\|_*$ is the operator norm. If $\|f(v)\| \le 1$ then we have the bound

$$F^2 = \sum_{v \in V} \|f(v)\|_F^2 \le dn$$

Suppose we are interested in obtaining an estimate on the probability:

$$\mathbb{P}\left[\left\|\frac{1}{k} \sum_{v \in V} f(v)\right\| > \epsilon\right] \tag{5}$$

Applying Theorem 6 with parameter $\epsilon/2$ and noting that $\|W\| \le \|W\|_F$, we have that (5) is at most

$$\mathbb{P}\left[\left\|\frac{1}{k} \sum_{i=1}^{k} Z_i\right\| > \epsilon/2\right]$$

where Z_i is a martingale with bound

$$\|Z_i\| \le \frac{\log(nd/\epsilon)}{1 - \lambda}$$

We now appeal to the existing martingale generalization of (2) (see e.g. [Tro12]) and find that this probability is at most

$$2d \cdot \exp\left(-\Omega\left(\frac{k\epsilon^2(1-\lambda)^2}{\log^2(nd)}\right)\right)$$

While this theorem is much weaker than the previous one in terms of parameters (depending on the square of the mixing time rather than on the spectral gap), it shows qualitatively that concentration for Markov chains is a generic phenomenon rather than something specific to matrices. It also allows one to instantly import the wealth of results regarding concentration for martingales in various Banach spaces (see e.g., [LT13]) to the random walk setting, albeit with suboptimal parameters. The simple proof of Theorem 6 is presented in Section 5.

2 Preliminaries

For an $n \in \mathbb{N}_+$, let $[n]$ denote the set $\{1, 2, \cdots, n\}$. Let $\mathbf{i}$ denote $\sqrt{-1}$. For $z = a + \mathbf{i}b$, where $a, b \in \mathbb{R}$, we define the complex conjugate of z to be $\overline{z} = a - \mathbf{i}b$ and $|z| = \sqrt{a^2 + b^2}$. Then $|z|^2 = z\overline{z}$. We define real part $\mathrm{Re}(z) = a$ and imaginary part $\mathrm{Im(z)} = b$. Then $\mathrm{Re}(z) = \frac{z+\overline{z}}{2}$ and $\mathrm{Im}(z) = \frac{z-\overline{z}}{2\mathbf{i}}$.

We will be working with D-regular undirected graphs $G = (V, E)$. The number of vertices of the graph, V will be denoted by n. A will denote the adjacency matrix of the graph and $P = A/D$ will denote its normalized adjacency matrix. A regular graph G will be called a λ-expander ($0 < \lambda < 1$) if $\|Px\| \leq \lambda \cdot \|x\|$ for all vectors $x \in \mathbb{C}^n$ s.t. $\sum_{i=1}^n x_i = 0$.

We will use $e_i \in \mathbb{C}^d$ to denote the standard basis vector with 1 in i^{th} position and 0 everywhere else.

2.1 Linear Algebra

Matrices and Norms. For matrix A, we use $A^\top$ to denote the transpose of A, we use $\overline{A}$ to denote the entry-wise complex conjugate of A. For square matrix $A \in \mathbb{C}^{n \times n}$, we use A^* to denote the conjugate transpose of matrix A. It is obvious that $A^* = \overline{A}^\top = \overline{A^\top}$. We say a complex square matrix A is Hermitian, if $A = A^*$, unitary if $AA^* = A^*A = I$, and positive-semidefinite (psd) if $A = A^*$ and $x^*Ax \geq 0$ for all $x \in \mathbb{C}^n$. We use $\succeq, \preceq$ to denote the semidefinite ordering, e.g. $A \succeq 0$ means that A is psd.

For $p \in [1, \infty)$, define the Schatten p-norm of A as

$$\|A\|_p = \left(\sum_{n \geq 1} s_n^p(A) \right)^{1/p}$$

for $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A) \geq \cdots \geq 0$ the singular values of A, i.e., the eigenvalues of the Hermitian matrix $|A| = \sqrt{(A^*A)}$. Then $\|A\|_p^p = \mathrm{tr}[|T|^p]$.

For matrix $A \in \mathbb{C}^{n \times n}$, we define $\|A\|$ to be the spectral norm of A, i.e.,

$$\|A\| = \max_{\|x\|_2 = 1, x \in \mathbb{C}^n} x^*Ax$$

Tensor Products. Given two vectors $v \in \mathbb{C}^{d_1}$ and $w \in \mathbb{C}^{d_2}$, their tensor product $v \otimes w \in \mathbb{C}^{d_1 d_2}$ is the vector whose $(i, j)^{\text{th}}$ entry is $v(i)w(j)$ (for concreteness, assume the entries are in lexicographic order). Given two matrices $A_1 \in \mathbb{C}^{d_1 \times d_1}$ and $A_2 \in \mathbb{C}^{d_2 \times d_2}$, their tensor product $A_1 \otimes A_2 \in \mathbb{C}^{d_1 d_2 \times d_1 d_2}$ is the matrix whose $((i, k), (j, l))^{\text{th}}$ entry is $A_1(i, j)A_2(k, l)$. It is easy to see that

$$(A \otimes B)(v \otimes w) = Av \otimes Bw$$

and

$$(A \otimes B)(C \otimes D) = AB \otimes CD$$

For a matrix $X \in \mathbb{C}^{d\times d}$, $\mathrm{vec}(X) \in \mathbb{C}^{d^2}$ will denote the vectorized version of the matrix X. That is

$$\mathrm{vec}(X) = \sum_{i,j=1}^{d} X(i,j) e_i \otimes e_j$$

We have the following relationship between matrix multiplication and the tensor product:

$$\mathrm{vec}(AXB) = (A \otimes B^\top)\mathrm{vec}(X)$$

Exponential and Logarithm. All logarithms will be taken with the base e, and $\exp(x)$ will denote e^x. The matrix exponential of a complex matrix $A \in \mathbb{C}^{d\times d}$ is defined by the Taylor expansion:

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

which converges for all matrices A. We will use the fact that

$$\exp(A) \otimes \exp(B) = \exp(A \otimes I + I \otimes B)$$

which may be checked by expanding both sides and comparing terms.

The matrix logarithm of a positive definite matrix $A = UDU^*$ with D diagonal and positive is defined by

$$\log(A) := U \log(D) U^*$$

where the logarithm of D is taken entrywise. For such matrices we have

$$\log(\exp(A)) = \exp(\log(A)) = A$$

For positive definite A and complex z, we define $A^z := \exp(z \log(A))$.

Polar Decomposition. The polar decomposition of a square complex matrix A is a matrix decomposition of the form

$$A = UV$$

where U is a unitary matrix and V is a psd matrix. The polar decomposition separates matrix A into a component that stretches the space along a set of orthogonal axes, represented by V, and a rotation (with possible reflection) represented by U. The decomposition of the complex conjugate of matrix A can written as

$$\overline{A} = \overline{UV}$$

The decomposition of the conjugate transpose of matrix A can be written as

$$A^* = V^* U^*$$

The following simple proposition will be useful in our proofs.

Proposition 1. *Let A and B be Hermitian psd matrices. Then*

$$\mathrm{tr}[AB] \leq \|A\| \cdot \mathrm{tr}[B]$$

Proof. Let $B = \sum_{j=1}^{n} \sigma_j v_j v_j^\dagger$ be the eigenvalue decomposition of B. Then

$$\mathrm{tr}[AB] = \sum_{j=1}^{n} \sigma_j \mathrm{tr}\left[A v_j v_j^\dagger\right] = \sum_{j=1}^{n} \sigma_j v_j^\dagger A v_j \leq \sum_{j=1}^{n} \sigma_j \cdot \|A\| = \|A\| \cdot \mathrm{tr}[B]$$

□

2.2 Complex analysis

A function $f : U \to \mathbb{C}$ on a domain $U \subseteq \mathbb{C}$ is *holomorphic* if it has a complex derivative in a neighborhood of every point $z \in U$. The existence of a complex derivative in a neighborhood is a very strong condition, for it implies that any holomorphic function is actually infinitely differentiable and equal to its own Taylor series (i.e., *analytic*) at every point in U. A *biholomorphic* function is a bijective holomorphic function whose inverse is also holomorphic. It follows from the definition that sums, products, and compositions of holomorphic functions are holomorphic. We will also talk about matrix-valued holomorphic functions $f : U \to \mathbb{C}^{d\times d}$, which just means that every entry is holomorphic.

The main property that we will use is that the value of a holomorphic function at a point $z \in U$ can be related to values that it takes on the boundary of U, in the following way. A function $f : U \to \mathbb{R} \cup \{-\infty\}$ is called *subharmonic* if it is upper semicontinuous and

$$f(z) \leq \frac{1}{2\pi}\int_0^{2\pi} f(z + re^{i\theta})d\theta$$

for all $z \in U$ and $r > 0$ such that the closed disk $D(z, r)$ is contained in U, and all of the above integrals converge.

We will make frequent use of the following standard fact.

Proposition 2. *If f is analytic on a domain $U \subset \mathbb{C}$ then $\log |f(z)|$ is subharmonic on U.*

Our main tool will be the Poisson Integral Formula for subharmonic functions.

Lemma 1 (Poisson integral formula on unit disk [Gra08, Eq 1.3.35])**.** *For any subharmonic function U defined on the unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$, we have that*

$$U(z) \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} U(e^{\mathbf{i}\varphi})\frac{1-|z|^2}{|e^{\mathbf{i}\varphi} - z|^2}\mathrm{d}\varphi, \qquad \forall\ |z| < 1$$

3 New Golden-Thompson inequality

We begin by giving an outline of the proof of Theorem 5. The proof of Theorem 4 in [SBT17] relies on the multivariate Lie-Trotter product formula (e.g. see [Bha97]), which states that:

$$\exp\left(\sum_{j=1}^{k} L_j\right) = \lim_{\theta\to 0^+}\left(\prod_{j=1}^{k}\exp(\theta L_j)\right)^{\frac{1}{\theta}}.$$

For Hermitian L_j, a judicious application of the above allows one to rewrite the trace of the exponential as a limit of Schatten norms:

$$\log\left(\mathrm{tr}\left[\exp\left(\sum_{j=1}^{k} L_j\right)\right]\right) = \lim_{\theta\to 0^+}\frac{2}{\theta}\log\left\|\prod_{j=1}^{k}\exp\left(\frac{\theta}{2}L_j\right)\right\|_{2/\theta}$$

Thus, understanding the matrix exponential of a sum is the same as understanding the behavior of a certain norm of the product as $\theta \to 0$. The idea of [SBT17] is to use *complex interpolation*, along the lines of the Stein-Hirschman theorem in complex analysis: for every fixed real θ near zero, find

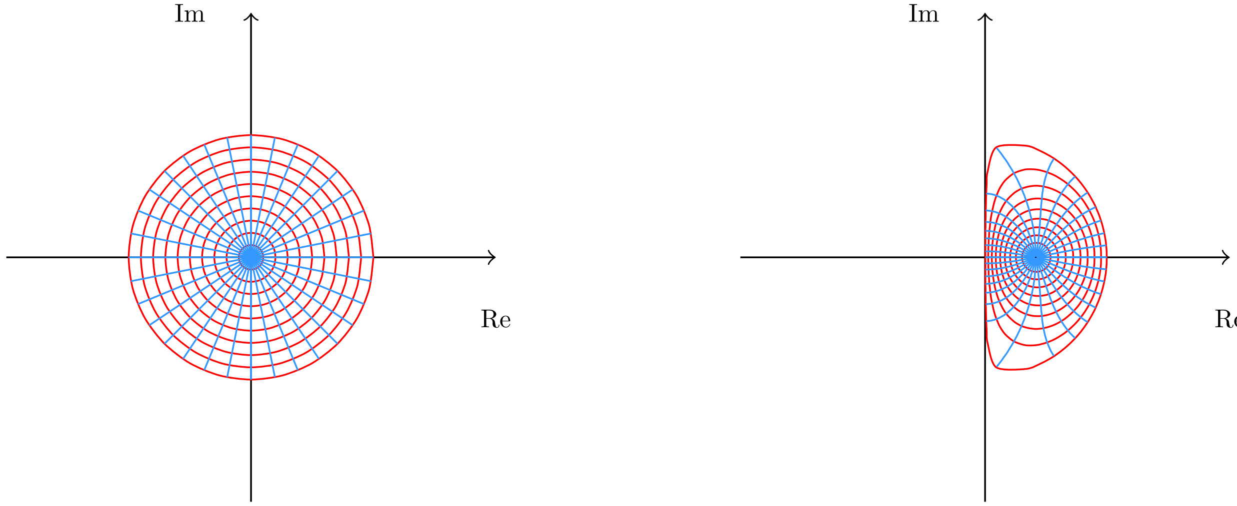

Figure 1. The function $h(z) = -\frac{1+z}{1-z} + \sqrt{\left(\frac{1+z}{1-z}\right)^2 + 1}$ maps the unit disk $\{z \in \mathbb{C} : |z| \le 1\}$ to the half disk $\{z \in \mathbb{C} : |z| \le 1 \text{ and } \mathrm{Re}(z) \ge 0\}$.

a complex function $F_\theta(z)$ that agrees with the right hand side at $z = \theta$ and is holomorphic on the strip $\{0 \le \Re(z) \le 1\}$. Since the value of a holomorphic function at any point can be related to an integral of its values on the boundary, this allows one to relate $F_\theta(\theta)$ to its integrals on $\{\Re(z) = 0\}$ and $\{\Re(z) = 1\}$, which are easy to understand. Taking the limit in θ yields Theorem 4.

To avoid integration on the whole vertical line $\{1 + \mathbf{i}b : b \in \mathbb{R}\}$, we observe that the above strategy only relies on the fact that θ is enclosed by the two vertical lines $\{\mathbf{i}b : b \in \mathbb{R}\}$ and $\{1 + \mathbf{i}b : b \in \mathbb{R}\}$, and we could have used any other region enclosing a neighborhood of real positive θ near zero, provided we can define the required holomorphic functions F_θ. We choose the half-circle (which is easy to work with because the Riemann map to the unit disk is explicit) and use it to derive a variant of the Riesz–Thorin theorem (Theorem 7), from which our new multi-matrix Golden-Thompson inequality follows by mimicking the remainder of the proof of 4 given in [SBT17].

3.1 Complex estimate on the half disk

In general, we can upper bound the value of any subharmonic function on a simply connected domain by mapping the domain to the unit disk via Riemann mapping theorem and applying the Poisson integral formula. In this section, we will give such estimate on a unit half disk.

The following lemma follows from the biholomorphic map from unit disk onto the half disk. defined in Figure 1.

Lemma 2. *[Poisson Integral Formula on the Half-Disk] For any analytic function F on the half disk $\{z \in \mathbb{C} : |z| \le 1$ and $\mathrm{Re}(z) \ge 0\}$, we have that*

$$\log |F(x)| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left|F \circ h(e^{\mathbf{i}\varphi})\right| \frac{1-\rho^2}{1 - 2\rho\cos(\varphi) + \rho^2} \mathrm{d}\varphi \tag{6}$$

for any $0 \le x \le 1$ where

$$h(z) = -\frac{1+z}{1-z} + \sqrt{\left(\frac{1+z}{1-z}\right)^2 + 1} \quad \text{and} \quad \rho = \frac{x^2 + 2x - 1}{x^2 - 2x - 1}$$

Proof. Note that the function $h(z)$ is a biholomorphic map from the unit disk $\{z \in \mathbb{C} : |z| \le 1\}$ to the half disk $\{z \in \mathbb{C} : |z| \le 1 \text{ and } \mathrm{Re}(z) \ge 0\}$. (We provide a proof for completeness, see Lemma 6 [2])

Since F and h are holomorphic, $\log|F \circ h(z)|$ is subharmonic. Therefore, we can apply the Poisson integral formula for subharmonic functions (Lemma 1) and get

$$\begin{aligned}\log|F \circ h(z)| &\le \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|F \circ h(e^{\mathbf{i}\varphi})\right| \frac{1-|z|^2}{|e^{\mathbf{i}\varphi} - z|^2}\mathrm{d}\varphi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|F \circ h(e^{\mathbf{i}\varphi})\right| \frac{1-\rho^2}{1-2\rho\cos(\theta-\varphi)+\rho^2}\mathrm{d}\varphi\end{aligned}$$

for $z = \rho e^{\mathbf{i}\theta}$.

Setting $x = h(z)$, we obtain that

$$z = \frac{x^2+2x-1}{x^2-2x-1}$$

Therefore, we have that

$$\theta = 0 \text{ and } \rho = \frac{x^2+2x-1}{x^2-2x-1}$$

This gives the desired result. $\square$

The following lemma allows us to conveniently study the behavior of certain analytic functions near zero. The idea is that when $|F(z)|$ is at most 1 on the imaginary axis, $\log|F(\theta)|$ should be close to 0 for small θ, and the value of $\log|F(\theta)|/\theta$ can be upper bounded by a suitable average of the values on the boundary of the half disk.

Lemma 3. *Given any analytic function F on the half disk $\{z \in \mathbb{C} : |z| \le 1$ and $\mathrm{Re}(z) \ge 0\}$. Suppose that $|F(\mathbf{i}y)| \le 1$ for all $y \in [-1,1]$. Then, for any $0 \le \theta \le 1/4$, we have*

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

where μ_θ is some probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ depending only on θ, and $\mu_\theta \to$ some probability distribution as $\theta \to 0^+$.

Proof. Since $\log|F(\mathbf{i}y)| \le 0$ for all $y \in [-1,1]$ and since $h(e^{\mathbf{i}\varphi})$ is imaginary with modulus at most 1 whenever $|\varphi| \le \pi/2$, we can ignore these φ in the integral (6), namely,

$$\begin{aligned}\log|F(x)| &\le \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|F \circ h(e^{\mathbf{i}\varphi})\right| \frac{1-\rho^2}{1-2\rho\cos(\varphi)+\rho^2}\mathrm{d}\varphi \\ &\le \frac{1}{2\pi}\int_{\pi/2 \le |\varphi| \le \pi} \log\left|F \circ h(e^{\mathbf{i}\varphi})\right| \frac{1-\rho^2}{1-2\rho\cos(\varphi)+\rho^2}\mathrm{d}\varphi. \end{aligned} \tag{7}$$

[2] The similar version is an exercise 4 in page 163 (Section VII) of [Con78], and also can be found here, https://math.stackexchange.com/questions/882147/find-a-conformal-map-from-semi-disk-onto-unit-disk

To bound the right hand side, for $\pi/2 \leq |\varphi| \leq \pi$ and $0 \leq \rho \leq 1$, we can prove the following statement using elementary calculations (see Lemma 7),

$$\frac{1-\rho^2}{1-2\rho\cos(\varphi)+\rho^2} = \frac{1-\rho}{1-\cos(\varphi)} \pm O(1-\rho)^2$$

Note that $\rho = \frac{\theta^2+2\theta-1}{\theta^2-2\theta-1} \geq 1-4\theta$ for all $0 \leq \theta \leq 1/4$. Therefore, we have that $0 \leq \rho \leq 1$ and

$$\frac{1-\rho^2}{1-2\rho\cos(\varphi)+\rho^2} \leq \frac{4\theta}{1-\cos(\varphi)} + O(\theta^2)$$

Putting this inequality into (7), we have that

$$\log|F(x)| \leq \frac{1}{2\pi}\int_{\pi/2\leq|\varphi|\leq\pi}\left(\frac{4\theta}{1-\cos(\varphi)} + O(\theta^2)\right)\log|F \circ h(e^{\mathbf{i}\varphi})|\mathrm{d}\varphi$$

We now observe that

$$\int_{\pi/2\leq|\varphi|\leq\pi}\frac{1}{1-\cos(\varphi)}\mathrm{d}\varphi = \int_{-\pi}^{-\pi/2}\frac{1}{1-\cos(\varphi)}\mathrm{d}\varphi + \int_{\pi/2}^{\pi}\frac{1}{1-\cos(\varphi)}\mathrm{d}\varphi = 2$$

Note that h maps $e^{\mathbf{i}\varphi}$ for $\pi/2 \leq |\varphi| \leq \pi$ to the boundary of the half disk $([-\pi/2, \pi/2])$, and let μ_θ be some probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ depending only on θ. Then we can have

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

Note that $\mu_\theta \to$ some probability distribution as $\theta \to 0^+$. $\square$

3.2 Bounded Multimatrix Golden-Thompson type inequality

Plugging our new complex estimate into the proof of Theorem 3.1 in [SBT17], we obtain the following Riesz-Thorin-type inequality. We give a complete proof for completeness.

Theorem 7 (Riesz-Thorin-type inequality)**.** *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. 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.

Proof. To apply Lemma 3, we want to define a holomorphic function $F(z)$ such that

$$|F(\mathbf{i}t)| \leq 1\ ,\ |F(z)| \leq \|G(z)\|_{p_1} \text{ and } F(\theta) = \|G(\theta)\|_{p_\theta}$$

Now, we describe how to define such $F(z)$. For $x \in [0,1]$, define q_x as the Hölder conjugate of p_x such that $p_x^{-1} + q_x^{-1} = 1$. Hence, using the definition of p_x in the statement, we have

$$\frac{1}{q_x} = \frac{1-x}{q_0} + \frac{x}{q_1}$$

Now for our fixed $\theta \in (0,1)$, let $G(\theta) = UV$ be the polar decomposition of $G(\theta)$, where V is positive definite since $G(\theta)$ is always invertible, and U is unitary. Finally, we define $X(z)$ and $F(z)$ by

$$\begin{aligned} X(z)^* &= (V/c)^{p_\theta(\frac{1-z}{q_0}+\frac{z}{q_1})} U^*, \text{ where } c = \|V\|_{p_\theta} = \|G(\theta)\|_{p_\theta} \\ F(z) &= \mathrm{tr}[X(z)^* G(z)] \end{aligned}$$

Note that $F(z)$ is holomorphic. Due to the renormalization c, we can show $\|X(x+\mathbf{i}y))\|_{q_x}^{q_x} = 1$ for all $x \in [0,1]$:

$$\begin{aligned} \|X(x+\mathbf{i}y)\|_{q_x}^{q_x} &= \mathrm{tr}\left[\sqrt{X^*(x+\mathbf{i}y)X(x+\mathbf{i}y)}^{q_x}\right] && \text{by definition of } \|\cdot\|_p \\ &= \mathrm{tr}\left[(V/c)^{q_x p_\theta(\frac{1-x}{q_0}+\frac{x}{q_1})}\right] && \text{by } U^*U = I \\ &= \mathrm{tr}\left[(V/c)^{p_\theta}\right] && \text{by } \frac{1}{q_x} = \frac{1-x}{q_0} + \frac{x}{q_1} \\ &= 1. && \text{by } c = \|V\|_{p_\theta} \end{aligned}$$

Therefore, $F(z)$ is bounded on S as follows

$$|F(x+\mathbf{i}y)| \leq \|X(x+\mathbf{i}y)\|_{q_x} \cdot \|G(x+\mathbf{i}y)\|_{p_x} \leq \|G(x+\mathbf{i}y)\|_{p_x}$$

Using this, it is obvious that

$$|F(\mathbf{i}t)| \leq \|G(\mathbf{i}t)\|_{p_0} \leq 1, \text{ and } |F(z)| \leq \|G(z)\|_{p_{\mathrm{Re}(z)}} \leq \|G(z)\|_{p_1}$$

for all $z \in S$ where we used that $p_0 \geq p_{\mathrm{Re}(z)} \geq p_1$.

Finally, we verify that $F(\theta) = \|G(\theta)\|_{p_\theta}$:

$$\begin{aligned} F(\theta) &= \mathrm{tr}[X(\theta)^* G(\theta)] \\ &= \mathrm{tr}[(V/c)^{p_\theta(\frac{1-\theta}{q_0}+\frac{\theta}{q_1})} U^* \cdot UV] && \text{by definition of } X \text{ and } G \\ &= \mathrm{tr}[(V/c)^{p_\theta(\frac{1-\theta}{q_0}+\frac{\theta}{q_1})} V] && \text{by } U^*U = I \\ &= \mathrm{tr}[c^{-p_\theta/q_\theta} V^{1+p_\theta/q_\theta}] && \text{by } \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1} \\ &= \mathrm{tr}[c^{1-p_\theta} V^{p_\theta}] && \text{by } p_\theta/q_\theta = p_\theta(1-1/p_\theta) = p_\theta - 1 \\ &= c^{1-p_\theta} c^{p_\theta} && \text{by } (\mathrm{tr}[V^{p_\theta}])^{1/p_\theta} = c \\ &= \|G(\theta)\|_{p_\theta} && \text{by } c = \|G(\theta)\|_{p_\theta} \end{aligned}$$

Hence, the statement follows from Lemma 3. $\square$

Now, we are ready to prove our variant of multimatrix Golden-Thompson inequality. It follows from plugging in Theorem 7 into the proof of Theorem 3.5 in [SBT17]. We give a complete proof for completeness.

Theorem 8 (Multimatrix Golden-Thompson inequality)**.** *For any k Hermitian matrices $H_1, \cdots, H_k$, we have:*

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

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

Proof. Define

$$G(z) = \prod_{j=1}^{k}\exp\left(\frac{z}{2}H_j\right)$$

Note that $\|G(\mathbf{i}y)\|_\infty = 1$ for all $y \in \mathbb{R}$. We now have:

$$\begin{aligned}
&\log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j\right)\right]\right) \\
= \; &\log\left(\operatorname{tr}\left[\exp\left(\sum_{j=1}^{k} H_j/2 + \sum_{j=k}^{1} H_j/2\right)\right]\right) \\
&\text{by the Lie-Trotter formula} \\
= \; &\log\left(\operatorname{tr}\left[\lim_{\theta\to 0^+}(G(\theta)G(\theta)^*)^{1/\theta}\right]\right) \\
&\text{since the } H_j \text{ are Hermitian} \\
= \; &\lim_{\theta\to 0^+}\frac{2}{\theta}\log\left(\operatorname{tr}\left[(G(\theta)G(\theta)^*)^{1/\theta}\right]^{\theta/2}\right) \\
&\text{by continuity of log away from 0} \\
= \; &\lim_{\theta\to 0^+}\frac{2}{\theta}\log\|G(\theta)\|_{2/\theta} \\
\leq \; &\lim_{\theta\to 0^+} 2\left(\frac{4}{\pi}+O(\theta)\right)\int_{-\pi/2}^{\pi/2}\log\|G(e^{\mathbf{i}\phi})\|_2 \mathrm{d}\mu_\theta(\phi) \\
&\text{by Theorem 7 with } p_0 = \infty \text{ and } p_1 = 2 \\
= \; &\lim_{\theta\to 0^+}\left(\frac{4}{\pi}+O(\theta)\right)\int_{-\pi/2}^{\pi/2}\log\left(\operatorname{tr}\left[G(e^{\mathbf{i}\phi})G(e^{\mathbf{i}\phi})^*\right]\right)\mathrm{d}\mu_\theta(\phi)
\end{aligned}$$

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

Remark. We suspect the constant $4/\pi$ is tight and that any constant larger than one is unavoidable if we consider the maximum over a bounded domain inside the strip $\{z \in \mathbb{C} : 0 \leq \mathrm{Re}(z) \leq 1\}$.

4 Proof of Theorem 2

In this section we present the proof of Theorem 2. We restate it here (with explicit constants) for convenience.

Theorem 9. *Let G be a regular λ-expander on V and let f be a function $f : V \to \mathbb{C}^{d\times d}$ s.t.*

(i) For each $v \in V$, $f(v)$ is Hermitian and $\|f(v)\| \leq 1$.

(ii) $\sum_{v\in V} f(v) = 0$.

If $v_1, \ldots, v_k$ is a stationary random walk on G, and $\epsilon \in (0,1)$,

$$\mathbb{P}\left[\lambda_{\max}\left(\sum_{j=1}^{k} f(v_j)\right) \geq k\epsilon\right] \leq d^{2-\pi/4} \cdot \exp\left(-\epsilon^2(1-\lambda)k/80\right)$$

$$\mathbb{P}\left[\lambda_{\min}\left(\sum_{i=1}^{k} f(v_j)\right) \leq -k\epsilon\right] \leq d^{2-\pi/4} \cdot \exp\left(-\epsilon^2(1-\lambda)k/80\right)$$

Remark. Depsite the exponent of d is different from Theorem 2, we note that since the left hand side (the probability) is at most 1, one can prove the same statement with any positive exponent by changing the constant 80.

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

$$\begin{aligned}\mathbb{P}\left[\lambda_{\max}\left(\sum_{j=1}^{k} f(v_j)\right) \geq k\epsilon\right] &\leq \mathbb{P}\left[\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \geq \exp(tk\epsilon)\right] \\ &\leq \frac{\mathbb{E}\left[\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right]\right]}{\exp(tk\epsilon)} \end{aligned} \tag{8}$$

The second inequality follows from Markov's inequality.

Now the question is how to bound $\mathbb{E}_{v_1,\cdots,v_k}[\operatorname{tr}[\exp(t\sum_{j=1}^{k} f(v_j))]]$. Using Theorem 8 and note that $\mu(\phi)$ is a probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, we have

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

where the the second step follows by concavity of log function. This implies that

$$\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \leq \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\mathbf{i}\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\mathbf{i}\phi}}{2}tf(v_j)\right)\right]\mathrm{d}\mu(\phi)\right)^{\frac{4}{\pi}} \tag{9}$$

Note that $\|x\|_p \le d^{1/p-1}\|x\|_1$ for $p \in (0,1)$, choosing $p = \pi/4$ we have

$$\left(\operatorname{tr}\left[\exp\left(\frac{\pi}{4}t\sum_{j=1}^{k} f(v_j)\right)\right]\right)^{\frac{4}{\pi}} \le d^{4/\pi-1}\operatorname{tr}\left[\exp\left(t\sum_{j=1}^{k} f(v_j)\right)\right] \tag{10}$$

Combining Eq. (9) and Eq. (10), we have

$$\operatorname{tr}\left[\exp\left(\frac{\pi}{4}t\sum_{j=1}^{k} f(v_j)\right)\right] \le d^{1-\pi/4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\mathbf{i}\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\mathbf{i}\phi}}{2}tf(v_j)\right)\right]\mathrm{d}\mu(\phi) \tag{11}$$

The core of the proof is the following bound on the moment generating function-like expression that appears above, thinking of $\mathbf{i}\phi$ as $\gamma + \mathbf{i}b$ with $\gamma^2 + b^2 = 1$:

Lemma 4. *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 $t > 0$, $\gamma \ge 0$, $b > 0$, $t^2(\gamma^2 + b^2) \le 1$, and $t\gamma \le \frac{1-\lambda}{4\lambda}$ we have*

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

Assuming this lemma, we can easily complete the proof of the theorem as:

$$\begin{aligned}
&\mathbb{E}_{v_1,\cdots,v_k}\left[\operatorname{tr}\left[\exp\left(\frac{\pi}{4}t\sum_{j=1}^{k} f(v_j)\right)\right]\right] \\
&\le d^{1-\pi/4}\mathbb{E}_{v_1,\cdots,v_k}\left[\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\mathbf{i}\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\mathbf{i}\phi}}{2}tf(v_j)\right)\right]\mathrm{d}\mu(\phi)\right] \\
&= d^{1-\pi/4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbb{E}_{v_1,\cdots,v_k}\left[\operatorname{tr}\left[\prod_{j=1}^{k}\exp\left(\frac{e^{\mathbf{i}\phi}}{2}tf(v_j)\right)\prod_{j=k}^{1}\exp\left(\frac{e^{-\mathbf{i}\phi}}{2}tf(v_j)\right)\right]\right]\mathrm{d}\mu(\phi) \\
&\le d^{1-\pi/4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\exp\left(kt^2|e^{\mathbf{i}\phi}|^2(1+\frac{8}{1-\lambda})\right)\mathrm{d}\mu(\phi) \\
&= d^{2-\pi/4}\exp\left(kt^2(1+\frac{8}{1-\lambda})\right)\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\mathrm{d}\mu(\phi) \\
&= d^{2-\pi/4}\exp\left(kt^2(1+\frac{8}{1-\lambda})\right)
\end{aligned} \tag{12}$$

where the first step follows by the Equation (11), the second step follows by swapping $\mathbb{E}$ and $\int$, the third step follows by Lemma 4 , the fourth step follows by $|e^{\mathbf{i}\phi}| = 1$, and the last step follows by $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\mathrm{d}\mu(\phi) = 1$.

Finally, putting it all together,

$$\mathbb{P}_{v_1,\cdots,v_k}\left[\lambda_{\max}\left(\sum_{j=1}^{k} f(v_j)\right) \ge k\epsilon\right] \le d^{2-\pi/4}\cdot\exp\left((4/\pi)^2kt^2\frac{9}{1-\lambda} - kt\epsilon\right)$$

$$
\begin{aligned}
&= d^{2-\pi/4} \cdot \exp\left((4/\pi)^2 k\epsilon^2 (1-\lambda)^2 \frac{1}{36^2}\frac{9}{1-\lambda} - k\frac{(1-\lambda)\epsilon}{36}\epsilon\right) \\
&\le d^{2-\pi/4} \cdot \exp\left(-k\epsilon^2(1-\lambda)/72\right)
\end{aligned}
$$

where the first step follows by Eq. (12) the second step follows by choosing $t = (1-\lambda)\epsilon/36$. $\square$

We now give the proof of Lemma 4

Proof of Lemma 4. We start by writing the expected trace expression in terms of the transition matrix of the random walk. This is an analogue of a step which is common to most of the expander chernoff bound proofs in the scalar case. Let P be the normalized adjacency matrix of G and let $\widetilde{P} = P \otimes I_{d^2}$. Let E denote the $nd^2 \times nd^2$ block diagonal matrix where the v^{th} diagonal block is the matrix

$$
M_v = \exp\left(\frac{tf(v)(\gamma + \mathbf{i}b)}{2}\right) \otimes \exp\left(\frac{tf(v)(\gamma - \mathbf{i}b)}{2}\right) \tag{13}
$$

Then $\left(E\widetilde{P}\right)^k$ is an $nd^2 \times nd^2$ block matrix whose $(u,v)^{\text{th}}$ $(d^2 \times d^2)$ block is given by the matrix

$$
\sum_{v_1,\ldots,v_{k-1}} P_{u,v_1} \cdot \left(\prod_{j=1}^{k-2} P_{v_j,v_{j+1}}\right) \cdot P_{v_{k-1},v} \cdot M_u \cdot \left(\prod_{j=1}^{k-1} M_{v_j}\right) \tag{14}
$$

Let $z_0 \in \mathbb{C}^{nd^2}$ be the vector $\frac{\mathbf{1}}{\sqrt{n}} \otimes \mathrm{vec}(I_d)$. Here $\mathbf{1}$ is the all 1's vector and $\mathrm{vec}(I_d)$ is the vector form of the identity matrix. Then, by applying

$$
\langle \mathrm{vec}(I_d), A_1 \otimes A_2 \,\mathrm{vec}(I_d)\rangle = \mathrm{tr}\left[A_1 A_2^T\right]
$$

it follows that for a stationary random walk $v_1, \ldots, v_k$:

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{tr}\left[\prod_{j=1}^{k} \exp\left(\frac{tf(v_j)(\gamma+\mathbf{i}b)}{2}\right)\prod_{j=k}^{1}\exp\left(\frac{tf(v_j)(\gamma-\mathbf{i}b)}{2}\right)\right]\right] &= \mathbb{E}\left[\left\langle \mathrm{vec}(I_d), \prod_{i=1}^{k} M_{v_i} \mathrm{vec}(I_d)\right\rangle\right] \\
&= \left\langle z_0, \left(E\widetilde{P}\right)^k z_0\right\rangle
\end{aligned}
$$

Hence we can focus our attention on $\left\langle z_0, \left(E\widetilde{P}\right)^k z_0\right\rangle$.

Let $\mathbf{1}$ be the all 1's vector. For a vector $z \in \mathbb{C}^{nd^2}$, let $z^{\|}$ denote the component of z which lies in the subspace spanned by the d^2 vectors $\mathbf{1} \otimes e_i$, $1 \le i \le d^2$. Let $z^{\perp}$ denote the component in the orthogonal space. Letting $z_j = \left(E\widetilde{P}\right)^j z_0$, we are interested in bounding

$$
\langle z_0, z_k\rangle = \langle z_0, z_k^{\|}\rangle \le \|z_0\| \cdot \|z_k^{\|}\| = \sqrt{d} \cdot \|z_k^{\|}\|
$$

The following lemma is the analogue of the main lemma in Healy's proof for the scalar valued expander Chernoff bound [Hea08]. Roughly speaking, it tracks how much a vector can move in and out of the subspace we are interested in as the operator $E\widetilde{P}$ is applied.

Lemma 5. *Given four parameters $\lambda \in [0,1]$, $\gamma \geq 0$, $\ell \geq 0$, and $t > 0$. Let G be a regular λ-expander on V. Suppose each vertex v is assigned a matrix $H_v \in \mathbb{C}^{d^2 \times d^2}$ s.t. $\|H_v\| \leq \ell$ and $\sum_v H_v = 0$. Let P be the normalized adjacency matrix of G and let $\widetilde{P} = P \otimes I_{d^2}$. Let E denote the $nd^2 \times nd^2$ block diagonal matrix where the v-th diagonal block is the matrix $\exp(tH_v)$. Also suppose that $\|E\| = \max_{v\in V} \|\exp(tH_v)\| \leq \exp(\gamma t)$. Then for any $z \in \mathbb{C}^{nd^2}$, we have:*

1. $\|(E\widetilde{P}z^{\|})^{\|}\| \leq \alpha_1 \|z^{\|}\|$ *where* $\alpha_1 = \exp(t\ell) - t\ell$
2. $\|(E\widetilde{P}z^{\|})^{\perp}\| \leq \alpha_2 \|z^{\|}\|$ *where* $\alpha_2 = \exp(t\ell) - 1$
3. $\|(E\widetilde{P}z^{\perp})^{\|}\| \leq \alpha_3 \|z^{\perp}\|$ *where* $\alpha_3 = \lambda \cdot (\exp(t\ell) - 1)$
4. $\|(E\widetilde{P}z^{\perp})^{\perp}\| \leq \alpha_4 \|z^{\perp}\|$ *where* $\alpha_4 = \lambda \cdot \exp(t\gamma)$

Proof of Lemma 5. Part 1.

Note that $(E\widetilde{P}z^{\|})^{\|} = (Ez^{\|})^{\|}$. Let $\mathbf{1} \in \mathbb{R}^n$ denote all ones vector, suppose $z^{\|} = \mathbf{1} \otimes \mathbf{w}$ for some $\mathbf{w} \in \mathbb{C}^{d^2}$. Then $\|z^{\|}\| = \sqrt{n} \cdot \|\mathbf{w}\|$ and

$$(Ez^{\|})^{\|} = \mathbf{1} \otimes \left(\frac{1}{n} \sum_{v \in V} \exp(H_v) \mathbf{w} \right)$$

We can upper bound $\|\frac{1}{n} \sum_{v\in V} \exp(tH_v)\|$ in the following way,

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{v \in V} \exp(tH_v) \right\| &= \left\| \frac{1}{n} \sum_{v \in V} \sum_{j=0}^{\infty} \frac{t^j H_v^j}{j!} \right\| \\
&= \left\| I + \frac{1}{n} \sum_{v \in V} \sum_{j=2}^{\infty} \frac{t^j H_v^j}{j!} \right\| \\
&\leq 1 + \frac{1}{n} \sum_{v \in V} \sum_{j \geq 2} \frac{t^j}{j!} \|H_v\|^j \\
&= 1 + \sum_{j \geq 2} \frac{(t\ell)^j}{j!} \\
&= \exp(t\ell) - t\ell
\end{aligned}$$

where the first step follows by Taylor expansion, the second step follows by $\sum_{v\in V} H_v = 0$, the third step follows by triangle inequality, the fourth step follows by $|V| = n$ and $\|H_v\| \leq \ell$, and last step follows by Taylor expansion.

Thus,

$$\|(Ez^{\|})^{\|}\| = \sqrt{n} \left\| \frac{1}{n} \sum_{v \in V} \exp(tH_v) w \right\| \leq \sqrt{n} \|w\| (\exp(t\ell) - t\ell) = \|z^{\|}\| (\exp(t\ell) - t\ell)$$

Part 2. Note that $(E\widetilde{P}z^{\|})^{\perp} = (Ez^{\|})^{\perp} = ((E-I)z^{\|})^{\perp}$. and $(z^{\|})^{\perp} = 0$. We can upper bound $\|((E-I)z^{\|})^{\perp}\|$ in the following way,

$$\begin{aligned}
\|((E-I)z^{\|})^{\perp}\| &\leq \|(E-I)z^{\|}\| \\
&\leq \|E - I\| \cdot \|z^{\|}\|
\end{aligned}$$

$$
\begin{aligned}
&= \max_{v\in V} \|\exp(tH_v) - I\| \cdot \|z^{\|}\| \\
&= \max_{v\in V} \left\| \sum_{j=1}^{\infty} \frac{t^j}{j!} H_v^j \right\| \cdot \|z^{\|}\| \\
&\leq \left(\sum_{j=1}^{\infty} \frac{t^j \ell^j}{j!} \right) \cdot \|z^{\|}\| \\
&= (\exp(t\ell) - 1) \cdot \|z^{\|}\|
\end{aligned}
$$

where the second step follows by $\|Ax\| \leq \|A\| \cdot \|x\|$, the third step follows by definition of E, the fourth step follows by Taylor expansion, the fifth step follows by triangle inequality and $\|H_v\| \leq \ell$, and the last step follows by Taylor expansion.

Part 3. Note that

$$
(E\widetilde{P}z^{\perp})^{\|} = ((E-I)\widetilde{P}z^{\perp})^{\|}
$$

This is because $(\widetilde{P}z^{\perp})^{\|} = 0$ since $\widetilde{P}$ preserves the property of being orthogonal to the space spanned by the vectors $\mathbf{1} \otimes e_i$ (these are the top eigenvectors of $\widetilde{P}$). Hence we can bound

$$
\begin{aligned}
\left\| ((E-I)\widetilde{P}z^{\perp})^{\|} \right\| &\leq \left\| (E-I)\widetilde{P}z^{\perp} \right\| \\
&\leq \|E - I\| \cdot \left\| \widetilde{P}z^{\perp} \right\| \\
&\leq (\exp(t\ell) - 1) \cdot \lambda \cdot \left\| z^{\perp} \right\|
\end{aligned}
$$

Third inequality follows from the fact that $\|E - I\| \leq \exp(t\ell) - 1$ and that G is a λ-expander.

Part 4. We can bound

$$
\|(E\widetilde{P}z^{\perp})^{\perp}\| \leq \|E\widetilde{P}z^{\perp}\| \leq \exp(t\gamma) \cdot \lambda \|z^{\perp}\|
$$

where the second step follows by $\|E\| \leq \exp(\gamma t)$ and G is a λ-expander.

□

We now use the above Lemma 5 to analyze the evolution of $z_j^{\|}$ and $z_j^{\perp}$. Recall the definition of H_v,

$$
H_v = \frac{f(v)(\gamma + \mathbf{i}b)}{2} \otimes I_d + I_d \otimes \frac{f(v)(\gamma - \mathbf{i}b)}{2}
$$

which means

$$
\begin{aligned}
\exp(tH_v) = &\ \exp\left(\frac{tf(v)(\gamma + \mathbf{i}b)}{2} \otimes I_d + I_d \otimes \frac{tf(v)(\gamma - \mathbf{i}b)}{2} \right) \\
= &\ \exp\left(\frac{tf(v)(\gamma + \mathbf{i}b)}{2} \right) \otimes \exp\left(\frac{tf(v)(\gamma - \mathbf{i}b)}{2} \right) \\
= &\ M_v
\end{aligned}
$$

where the the first step follows by definition of H_v, the second step follows by $\exp(A \otimes I_d + I_d \otimes B) = \exp(A) \otimes \exp(B)$, and the last step follows by Eq. (13).

We can upper bound $\|H_v\| \leq \sqrt{\gamma^2 + b^2}$ and then set $\ell = \sqrt{\gamma^2 + b^2}$. We can also upper bound $\|\exp(tH_v)\|$

$$\|\exp(tH_v)\| = \|\exp(t \cdot \mathrm{Re}(H_v))\| = \left\|\exp\left(\gamma t \left(\frac{f(v)}{2} \otimes I_d + I_d \otimes \frac{f(v)}{2}\right)\right)\right\| \leq \exp(\gamma t)$$

Note that $\sum_{v \in V} H_v = 0$ since $\sum_{v \in V} f(v) = 0$.

Claim 10. $\|z_i^{\perp}\| \leq \frac{\alpha_2}{1-\alpha_4} \max_{j<i} \|z_j^{\|}\|$.

Proof.

$$\begin{aligned}
\|z_i^{\perp}\| &= \|(E\widetilde{P} z_{i-1})^{\perp}\| \\
&\leq \|(E\widetilde{P} z_{i-1}^{\|})^{\perp}\| + \|(E\widetilde{P} z_{i-1}^{\perp})^{\perp}\| \\
&\leq \alpha_2 \|z_{i-1}^{\|}\| + \alpha_4 \|z_{i-1}^{\perp}\| \\
&\leq (\alpha_2 + \alpha_2\alpha_4 + \alpha_2\alpha_4 + \cdots) \cdot \max_{j<i} \|z_j^{\|}\| \\
&\leq \frac{\alpha_2}{1-\alpha_4} \max_{j<i} \|z_j^{\|}\|
\end{aligned}$$

where the first step follows by definition of z_i, the second step follows by triangle inequality, the third step follows by part 2 and 4 of Lemma 5. $\square$

Claim 11. $\|z_i^{\|}\| \leq (\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4}) \max_{j<i} \|z_j^{\|}\|$.

Proof.

$$\begin{aligned}
\|z_i^{\|}\| &= \|(E\widetilde{P} z_{i-1})^{\|}\| \\
&\leq \|(E\widetilde{P} z_{i-1}^{\|})^{\|}\| + \|(E\widetilde{P} z_{i-1}^{\perp})^{\|}\| \\
&\leq \alpha_1 \|z_{i-1}^{\|}\| + \alpha_3 \|z_{i-1}^{\perp}\| \\
&\leq \alpha_1 \|z_{i-1}^{\|}\| + \alpha_3 \frac{\alpha_2}{1-\alpha_4} \max_{j<i-1} \|z_j^{\|}\| \\
&\leq (\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4})
\end{aligned}$$

where the first step follows by definition of z_i, the second step follows by triangle inequality, the third step follows by part 1 and 3 of Lemma 5, the fourth step follows by Claim 10. $\square$

Combining Claim 10 and Claim 11 gives

$$\|z_k^{\|}\| \leq (\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4})^k \|z_0^{\|}\| = \sqrt{d} \cdot (\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4})^k$$

which implies that

$$\left\langle z_0, (E\widetilde{P})^k z_0 \right\rangle \leq d \cdot (\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4})^k$$

Now the question is how to bound $(\alpha_1 + \frac{\alpha_2\alpha_3}{1-\alpha_4})^k$. We can upper bound α_1, $\alpha_2\alpha_3$ and α_4 in the following sense,

$$\alpha_1 = \exp(t\ell) - t\ell \leq 1 + t^2\ell^2 = 1 + t^2(\gamma^2 + b^2)$$

and

$$\alpha_2\alpha_3 = \lambda(\exp(t\ell)-1)^2 \le \lambda(2t\ell)^2 = 4\lambda t^2(\gamma^2+b^2)$$

where the second step follows by $t\ell < 1$ (because $\exp(x) \le 1+2x, \forall x \in [0,1]$),

$$\alpha_4 = \lambda \cdot \exp(t\gamma) \le \lambda(1+2t\gamma) \le \frac{1}{2} + \frac{1}{2}\lambda$$

where the second step follows by $t\gamma < 1$, and the third step follows by $t\gamma \le (1-\lambda)/4\lambda$.

Thus,

$$\begin{aligned}(\alpha_1 + \frac{\alpha_2\cdot\alpha_3}{1-\alpha_4})^k \le & \left(1 + t^2(\gamma^2+b^2) + \frac{4\lambda t^2(\gamma^2+b^2)}{\frac{1}{2}-\frac{1}{2}\lambda}\right)^k \\ \le & \exp\left(kt^2(\gamma^2+b^2)(1+\frac{8}{1-\lambda})\right)\end{aligned}$$

$\square$

Remark. As is the case with Healy's proof [Hea08], our proof also works for the case when there are different mean zero functions $f_1, \dots, f_k$ for the different steps of the walk and also when there are k λ-expanders $G_1, \dots, G_k$ and the j^{th} step of the walk is taken according to G_j.

Remark. We suspect that with appropriate modifications, our proof should generalize to random walks on irregular undirected graphs (or reversible Markov chains) as was done for Healy's proof in [CLLM12].

Remark. Although we have stated the theorem for Hermitian matrices, the same result can be obtained for general matrices by a standard dilation trick, namely replacing every $d \times d$ matrix M that appears with the $2d \times 2d$ Hermitian matrix

$$\begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix}$$

whose norm is always within a factor of two of M.

5 Proof of Theorem 6

Proof. Observe that for every $i = 2, \dots, k$ we have

$$\mathbb{E}[f(v_i) \mid v_{i-1}] = \sum_{u \sim v_{i-1}} P(v_{i-1}, u) f(u) = Pf(v_{i-1})$$

whence the random vectors

$$Y_i^{(1)} := f(v_i) - Pf(v_{i-1})$$

satisfy

$$\mathbb{E}[Y_i^{(1)} \mid v_1, \dots, v_{i-1}] = 0$$

and thus form a martingale difference sequence with respect to the filtration generated by initial segments of $v_1, \ldots, v_k$. Denoting $Y_1^{(1)} := f(v_1)$, we can write the sum of interest as a martingale part plus a remainder, which is a sum of $k-1$ (i.e., one fewer) random variables:

$$S = \sum_{i=1}^{k} Y_i^{(1)} + \sum_{i=1}^{k-1} Pf(v_i)$$

Notice that Pf is also a mean zero function on G, and by Jensen's inequality we have

$$\|(Pf)(v)\|_* \leq M := \max_{v \in V} \|f(v)\|_* \qquad \text{for all } v$$

The key point is that the remainder terms $Pf(v_i)$ are smaller on average than the original terms $f(v_i)$ in squared Euclidean norm, because P is a contraction orthogonal to the constant vector; in particular, by considering the action of P on each coordinate of f separately, we have:

$$\sum_v \|Pf(v)\|_2^2 \leq \lambda \cdot \sum_v \|f(v)\|_2^2 \tag{15}$$

Iterating this construction on the remainder a total of $T \leq k$ times, we obtain a sequence of martingales $1 \leq t \leq T$:

$$Y_1^{(t)} := P^{t-1} f(v_1) \qquad Y_i^{(t)} := P^{t-1} f(v_i) - P^t f(v_{i-1}), \qquad i = 1 \ldots, k-(t-1)$$

which are related to the original sum as:

$$S = \sum_{t=1}^{T} \sum_{i=1}^{k-t+1} Y_i^{(t)} + \sum_{i=1}^{k-T} P^T f(v_i)$$

Interchanging the order of summation, we find that the random matrices

$$Z_i := \sum_{t=1}^{\min\{(k+1-i), T\}} Y_i^{(t)}$$

themselves form a martingale difference sequence, with each

$$\|Z_i\|_* \leq \sum_t \|Y_i^{(t)}\|_* \leq TM$$

We bound the error $W := \frac{1}{k} \sum_{i=1}^{k-T} (P^T f)(v_i)$ crudely as:

$$\|W\|_2 \leq \frac{1}{k} \sum_{i=1}^{k-T} \|P^T f(v_i)\|_2 \leq \frac{k-T}{k} \sum_{v \in V} \|(P^T f)(v)\|_2 \leq \lambda^{T/2} F \leq \exp(-(1-\lambda)T/2) F$$

where $F := \sqrt{\sum_{v \in V} \|f(v)\|_2^2}$, by applying (15). Rearranging and setting $T = 2\log(F/\epsilon)/(1-\lambda)$ yields the advertised bound on $\|Z_i\|_*$. $\square$

6 Conclusion

We have presented new extensions to the classical Chernoff bound in the context of matrix-valued functions, focusing on expander graphs and random walks. Our primary contribution is proving the conjectured *Matrix Expander Chernoff Bound*, which generalizes both scalar Chernoff bounds for random walks and matrix Chernoff bounds for independent variables. This result offers strong concentration guarantees for sums of Hermitian matrix-valued functions sampled along a stationary random walk on an expander graph.

The novelty of our approach lies in the development of an improved multi-matrix generalization of the Golden-Thompson inequality, enabling us to handle non-commutative matrix operations in the context of dependent random variables. By leveraging advanced linear algebraic techniques and martingale approximations, we have achieved tighter bounds that significantly enhance randomness efficiency for samplers.

Our theoretical findings provide important improvements for derandomization in computational settings, particularly in spectral graph theory and quantum information. These results not only push the boundaries of matrix concentration inequalities but also pave the way for new applications in both theory and practice, furthering the study of efficient algorithms and derandomization techniques.

References

[AW02] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[Bha97] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 1997.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[CL06] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.

[CLLM12] Kai-min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff-Hoeffding bounds for Markov chains: Generalized and simplified. *arXiv preprint arXiv:1201.0559*, 2012.

[Con78] John B. Conway. *Functions of one Complex Variable*, volume I of *Graduate Texts in Mathematics (Second Edition)*. Springer New York, Heidelberg, Berlin, 1978.

[Gil98] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[Gol65] Sidney Golden. Lower bounds for the Helmholtz function. *Physical Review*, 137(4B):B1127, 1965.

[Gra08] Loukas Grafakos. *Classical Fourier Analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer New York, 2008.

[Hea08] Alexander D Healy. Randomness-efficient sampling within NC1. *Computational Complexity*, 17(1):3–37, 2008.

[HKT16] Fumio Hiai, Robert Koenig, and Marco Tomamichel. Generalized log-majorization and multivariate trace inequalities. *arXiv preprint arXiv:1609.01999*, September 2016.

[Kah97] Nabil Kahale. Large deviation bounds for Markov chains. *Combinatorics, Probability and Computing*, 6(04):465–474, 1997.

[Lez98] Pascal Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, pages 849–867, 1998.

[LP04] C. A. Leon and F. Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *Annals of Applied Probability*, 14(2), 2004.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LT13] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes.* Springer Science & Business Media, 2013.

[RR17] Shravas Rao and Oded Regev. A sharp tail bound for the expander random sampler. *arXiv preprint arXiv:1703.10205*, 2017.

[Rud99] Mark Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 3–13. IEEE, 2000.

[SBT17] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352(1):37–58, 2017. arXiv:1604.03023.

[Tho65] Colin J Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813, 1965.

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[Tro15] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[Wag08] R. Wagner. Tail estimates for sums of variables sampled by a random walk. *Combinatorics, Probability and Computing*, 17(2), 2008.

[WX] Avi Wigderson and David Xiao. Retraction of Wigderson-Xiao: A randomness-efficient sampler for matrix valued functions, eccc tr05-107. 2006.

[WX05] Avi Wigderson and David Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 397–406. IEEE, 2005.

[WX08] Avi Wigderson and David Xiao. Derandomizing the AW matrix-valued Chernoff bound using pessimistic estimators and applications. *Theory of Computing*, 4(3):53–76, 2008.

A Elementary calculations

Lemma 6. *We define function $h(z) : \mathbb{C} \to \mathbb{C}$ as follows*

$$h(z) = -\frac{1+z}{1-z} + \sqrt{\left(\frac{1+z}{1-z}\right)^2 + 1}$$

Then function $h(z)$ maps the unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ to the half disk $\{z \in \mathbb{C} : |z| \leq 1 \text{ and } \mathrm{Re}(z) \geq 0\}$.

Proof. We first compute the inverse of function $h(z)$, let $f = h^{-1}$. By definition of $h(z)$, we can do the following elementary calculations,

$$\begin{aligned}
& h(z) + \frac{1+z}{1-z} = \sqrt{\left(\frac{1+z}{1-z}\right)^2 + 1} \\
\Longrightarrow\ & \left(h(z) + \frac{1+z}{1-z}\right)^2 = \left(\frac{1+z}{1-z}\right)^2 + 1 \\
\Longrightarrow\ & (h(z))^2 + 2h(z)\frac{1+z}{1-z} = 1 \\
\Longrightarrow\ & (1-z)(h(z))^2 + 2h(z)(1+z) = (1-z) \\
\Longrightarrow\ & (h(z))^2 + 2h(z) - 1 = z((h(z))^2 - 2h(z) - 1)
\end{aligned}$$

Thus, we obtain that

$$f(z) = \frac{z^2 + 2z - 1}{z^2 - 2z - 1}$$

To finish the proof, we need Claim 12 and Claim 13.

Claim 12. *For all $z \in \{z \in \mathbb{C} : |z| \leq 1 \text{ and } \mathrm{Re}(z) \geq 0\}$,*

$$|f(z)| \leq 1$$

Proof. Let $z = re^{\mathbf{i}\phi}$, where $r \in [0,1]$ and $\phi \in [-\pi/2, \pi/2]$. It is easy to observe that $\cos(\phi) \in [0,1]$ We have

$$\begin{aligned}
|f(z)| &= \left|\frac{z^2 + 2z - 1}{z^2 - 2z - 1}\right| \\
&= \left|\frac{r^2 e^{\mathbf{i}2\phi} + 2re^{\mathbf{i}\phi} - 1}{r^2 e^{\mathbf{i}2\phi} - 2re^{\mathbf{i}\phi} - 1}\right| \\
&= \frac{|r^2 e^{\mathbf{i}2\phi} + 2re^{\mathbf{i}\phi} - 1|}{|r^2 e^{\mathbf{i}2\phi} - 2re^{\mathbf{i}\phi} - 1|}
\end{aligned}$$

We can compute the numerator,

$$\begin{aligned}
|r^2 e^{\mathbf{i}2\phi} + 2re^{\mathbf{i}\phi} - 1|^2 &= |r^2 \cos 2\phi + \mathbf{i} r^2 \sin 2\phi + 2r\cos\phi + 2r\mathbf{i}\sin\phi - 1|^2 \\
&= (r^2 \cos 2\phi + 2r\cos\phi - 1)^2 + (r^2 \sin 2\phi + 2r \sin\phi)^2 \\
&= r^4 + 4r^2 + 1 - 2r^2 \cos 2\phi + 4r^3 \cos 2\phi \cos\phi - 4r\cos\phi + 4r^3 \sin 2\phi \sin\phi
\end{aligned}$$

$$= r^4 + 4r^2 + 1 - 2r^2\cos 2\phi + 4r^3\cos\phi - 4r\cos\phi$$

We can compute the denominator,

$$\begin{aligned}
|e^{\mathbf{i}2\phi} - 2e^{\mathbf{i}\phi} - 1|^2 &= |r^2\cos 2\phi + \mathbf{i}r^2\sin 2\phi - 2r\cos\phi - 2r\mathbf{i}\sin\phi - 1|^2 \\
&= (r^2\cos 2\phi - 2r\cos\phi - 1)^2 + (r^2\sin 2\phi - 2r\sin\phi)^2 \\
&= r^4 + 4r^2 + 1 - 2r^2\cos 2\phi - 4r^3\cos 2\phi\cos\phi + 4r\cos\phi - 4r^3\sin 2\phi\sin\phi \\
&= r^4 + 4r^2 + 1 - 2r^2\cos 2\phi - 4r^3\cos\phi + 4r\cos\phi
\end{aligned}$$

Note that, in order to show $|f(z)| \le 1$, it is sufficient to prove

$$4r^3\cos\phi - 4r\cos\phi \le -4r^3\cos\phi + 4r\cos\phi$$

which is equivalent to

$$8r(1-r^2)\cos\phi \ge 0$$

It follows by definition of r and ϕ. Thus, we complete the proof. □

Next, we can show that

Claim 13. *For all z is on the boundary of half disk,*

$$|f(z)| = 1$$

Proof. First, we want to show that $\forall z \in [-\mathbf{i}, \mathbf{i}]$, $|f(z)| = 1$. Let $b \in [0,1]$, let $z = \mathbf{i}b$, then we have

$$\begin{aligned}
|f(z)| &= |f(\mathbf{i}b)| \\
&= \left|\frac{-b^2 + 2b\mathbf{i} - 1}{-b^2 - 2b\mathbf{i} - 1}\right| \\
&= 1
\end{aligned}$$

Second, we want to show that for all z on half circle, $|f(z)| = 1$. We replace z by $e^{\mathbf{i}\phi}$, where $\phi \in [-\pi/2, \pi/2]$. Then we have

$$\begin{aligned}
|f(z)| &= \left|\frac{z^2 + 2z - 1}{z^2 - 2z - 1}\right| \\
&= \left|\frac{e^{\mathbf{i}2\phi} + 2e^{\mathbf{i}\phi} - 1}{e^{\mathbf{i}2\phi} - 2e^{\mathbf{i}\phi} - 1}\right| \\
&= \frac{|e^{\mathbf{i}2\phi} + 2e^{\mathbf{i}\phi} - 1|}{|e^{\mathbf{i}2\phi} - 2e^{\mathbf{i}\phi} - 1|}
\end{aligned}$$

We can compute the numerator,

$$\begin{aligned}
|e^{\mathbf{i}2\phi} + 2e^{\mathbf{i}\phi} - 1| &= |\cos 2\phi + \mathbf{i}\sin 2\phi + 2\cos\phi + 2\mathbf{i}\sin\phi - 1| \\
&= \left((\cos 2\phi + 2\cos\phi - 1)^2 + (\sin 2\phi + 2\sin\phi)^2\right)^{1/2} \\
&= (6 - 2\cos 2\phi + 4\cos 2\phi\cos\phi - 4\cos\phi + 4\sin 2\phi\sin\phi)^{1/2} \\
&= (6 - 2\cos 2\phi + 4\cos\phi - 4\cos\phi)^{1/2}
\end{aligned}$$

$$= (6 - 2\cos 2\phi)^{1/2}$$

We can compute the denominator,

$$\begin{aligned}
|e^{\mathbf{i}2\phi} - 2e^{\mathbf{i}\phi} - 1| &= |\cos 2\phi + \mathbf{i}\sin 2\phi - 2\cos\phi - 2\mathbf{i}\sin\phi - 1| \\
&= \left((\cos 2\phi - 2\cos\phi - 1)^2 + (\sin 2\phi - 2\sin\phi)^2\right)^{1/2} \\
&= (6 - 2\cos 2\phi - 4\cos 2\phi\cos\phi + 4\cos\phi - 4\sin 2\phi\sin\phi)^{1/2} \\
&= (6 - 2\cos 2\phi + 4\cos\phi - 4\cos\phi)^{1/2} \\
&= (6 - 2\cos 2\phi)^{1/2}
\end{aligned}$$

Thus, we have

$$|f(z)| = 1$$

$\square$

Note the biholomorphic is basically follows from the formula of f and h, because they are composition of holomorphic function. $\square$

Lemma 7. *For any $\rho \in [0,1]$ and $\cos\varphi \in [-1,0]$, we have*

$$\frac{1-\rho}{1-\cos\varphi} - (1-\rho)^2 \leq \frac{1-\rho^2}{1-2\rho\cos\varphi+\rho^2} \leq \frac{1-\rho}{1-\cos\varphi} + 2(1-\rho)^2$$

Proof. This directly follows by combining Claim 14 and Claim 15 $\square$

Claim 14. *There exists some sufficiently large constant $c \geq 1$ such that for any $\rho \in [0,1]$ and $\cos\varphi \in [-1,0]$, we have*

$$\frac{1-\rho^2}{1-2\rho\cos\varphi+\rho^2} \leq \frac{1-\rho}{1-\cos\varphi} + c(1-\rho)^2$$

Proof. It is equivalent to

$$\begin{aligned}
\frac{1+\rho}{1-2\rho\cos\varphi+\rho^2} &\leq \frac{1}{1-\cos\varphi} + c(1+\rho) \\
(1+\rho)(1-\cos\varphi) &\leq 1 - 2\rho\cos\varphi + \rho^2 + c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\varphi+\rho^2) \\
\rho - \cos\varphi &\leq -\rho\cos\varphi + \rho^2 + c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\varphi+\rho^2)
\end{aligned}$$

which is equivalent to,

$$-\rho\cos\varphi + \rho^2 - \rho + \cos\varphi + c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\phi+\rho^2) \geq 0$$

Since $1 - 2\rho\cos\varphi + \rho^2 \geq 1$, thus it suffices to show

$$(\rho-1)(\rho-\cos\varphi) + c(1+\rho)(1-\cos\varphi) \geq 0$$

Note that $(\rho-1)(\rho-\cos\varphi) \geq -2$, by choosing $c \geq 2$, we complete the proof. $\square$

Claim 15. *There exists some sufficiently large constant $c \geq 1$ such that for any $\rho \in [0,1]$ and $\cos\varphi \in [-1,0]$, we have*

$$\frac{1-\rho^2}{1-2\rho\cos\varphi+\rho^2} \geq \frac{1-\rho}{1-\cos\varphi} - c(1-\rho)^2$$

Proof. It is equivalent to

$$\begin{aligned}\frac{1+\rho}{1-2\rho\cos\varphi+\rho^2} &\geq \frac{1}{1-\cos\varphi} + c(1+\rho)\\ (1+\rho)(1-\cos\varphi) &\geq 1-2\rho\cos\varphi+\rho^2 - c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\varphi+\rho^2)\\ \rho-\cos\varphi &\geq \ -\rho\cos\varphi+\rho^2 - c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\varphi+\rho^2)\end{aligned}$$

which is equivalent to,

$$-\rho\cos\varphi+\rho^2-\rho+\cos\varphi - c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\phi+\rho^2) \leq 0$$

which is equivalent to

$$(\rho-1)(\rho-\cos\varphi) \leq c(1+\rho)(1-\cos\varphi)(1-2\rho\cos\varphi+\rho^2)$$

It suffices to choose $c=1$ $\square$