Derandomizing Matrix Concentration Inequalities from Free Probability

Robert Wang
University of Waterloo
Waterloo, Ontario, Canada
robert.wang2@uwaterloo.ca

Lap Chi Lau
University of Waterloo
Waterloo, Ontario, Canada
lapchi@uwaterloo.ca

Hong Zhou*
Fuzhou University
Fuzhou, Fujian, China
hong.zhou@fzu.edu.cn

Abstract

Recently, sharp matrix concentration inequalities were developed using the theory of free probability. In this work, we design polynomial time deterministic algorithms to construct outcomes that satisfy the guarantees of these inequalities. As direct consequences, we obtain polynomial time deterministic algorithms for the matrix Spencer problem and for constructing near-Ramanujan graphs. Our proofs show that the concepts and techniques in free probability are useful not only for mathematical analyses but also for efficient computations.

CCS Concepts

• **Theory of computation** → **Pseudorandomness and derandomization**; **Expander graphs and randomness extractors**; • **Mathematics of computing** → *Stochastic processes*; *Functional analysis*; *Combinatorial optimization.*

Keywords

Derandomization, Free Probability, Matrix Discrepancy, Deterministic Expander Construction

ACM Reference Format:
Robert Wang, Lap Chi Lau, and Hong Zhou. 2026. Derandomizing Matrix Concentration Inequalities from Free Probability. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC '26), June 22–26, 2026, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3798129.3800763

1 Introduction

Matrix concentration inequalities [18, 19, 21] consider the sum of independent matrix-valued random variables $X = \sum_i X_i$ and show that its spectral statistics such as its operator norm $\|X\|$ or the moments of its eigenvalues $\mathrm{tr}(X^p)^{\frac{1}{p}}$ are close to their expected values. These inequalities have found various important applications in theoretical computer science and mathematics, such as spectral sparsification of graphs [17], randomized numerical linear algebra [26], analysis of semidefinite programs [1], probabilistic constructions of expander graphs [7], and many more [19]. Derandomization of these concentration inequalities lead to deterministic algorithms for these problems and deterministic constructions of pseudorandom objects [1, 14, 25]. From the technical perspective, matrix concentration inequalities are more challenging to prove than the classical concentration inequalities for real-valued random variables because of the non-commutative nature of matrices.

Consider a general Gaussian model $X = \sum_{i=1}^n g_i A_i$, where $A_i \in \mathbb{C}^{d\times d}$ are arbitrary self-adjoint matrices and g_i are i.i.d. standard real Gaussian variables. Define the standard derivation $\sigma(X)$ as

$$\sigma(X)^2 := \left\|\mathbb{E}[X^2]\right\| = \left\|\sum_{i=1}^n A_i^2\right\|. \tag{1}$$

The non-commutative Khintchine inequality of Lust-Piquard and Pisier implies that

$$\sigma(X) \lesssim \mathbb{E}[\|X\|] \lesssim \sqrt{\log d}\cdot\sigma(X). \tag{2}$$

It is known that both the lower and upper bounds are asymptotically tight: The lower bound is of the correct order when X has independent entries, such that each A_i is supported on a distinct entry. The upper bound is of the correct order when X is a diagonal matrix, or more generally when all A_i commute so that they can be simultaneously diagonalized. The dimension-dependent logarithmic factor between the lower and upper bounds is suboptimal and undesirable in some applications. This motivates the question of whether there is a more refined quantity, based on the structure of A_i, that provides a tighter approximation to $\mathbb{E}[\|X\|]$.

The progress to this question was initiated by Tropp [20] and significantly advanced by Bandeira, Boedihardjo, and van Handel [2]. Informally, they demonstrated a new phenomenon that when the matrices A_i are sufficiently non-commutative, then the logarithmic factor can be removed. To formalize this, [2] defined the $d^2\times d^2$ covariance matrix $\mathrm{Cov}(X)$ where $\mathrm{Cov}(X)_{ij,kl} = \mathbb{E}[X_{ij}\overline{X_{kl}}]$ and

$$v(X)^2 := \|\mathrm{Cov}(X)\| = \left\|\sum_{i=1}^n \mathrm{vec}(A_i)\,\mathrm{vec}(A_i)^\top\right\|, \tag{3}$$

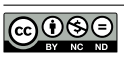
and proved a strong matrix concentration inequality that implies

$$\mathbb{E}[\|X\|] \le 2\sigma(X) + \sigma(X)^{\frac{1}{2}}\cdot v(X)^{\frac{1}{2}}\cdot(\log d)^{\frac{3}{4}}. \tag{4}$$

Note that the parameter $v(X)$ is much smaller than $\sigma(X)$ in the case where X has i.i.d. entries, and more generally $v(X)$ is small when the randomness of X is more well-spread over its entries. In [2, Section 3], it was shown that $v(X)\cdot(\log d)^{\frac{3}{2}} \lesssim \sigma(X)$ in many interesting settings, for which the new inequality provides an asymptotic sharp upper bound that $\mathbb{E}[\|X\|] \lesssim \sigma(X)$.

Notably, this new inequality was at the heart of a recent major progress towards proving the matrix Spencer conjecture [4], with applications in improved spectral sparsification for directed graphs [9].

*Corresponding author: Hong Zhou, School of Mathematics and Statistics, Fuzhou University, 350108 Fuzhou, Fujian, China.



STOC '26, Salt Lake City, UT, USA

ACM ISBN 979-8-4007-2536-4/2026/06
https://doi.org/10.1145/3798129.3800763

More recently, Brailovskaya and van Handel [6] extended these matrix concentration inequalities to a much larger class of random matrices. The random matrix model is $Z = A_0 + \sum_{i=1}^n Z_i$, where $Z_1, \dots, Z_n$ are independent Hermitian random matrices with $\mathbb{E}[Z_i] = 0$ and each $\|Z_i\|$ bounded. This captures many commonly studied random graph models, with arbitrary dependency pattern in their entries. They proved a general non-commutative universality principle, that the spectral statistics of Z are closely approximated by those of a Gaussian matrix X.

1.1 Technical Review

Free Probability: The inequality in (4) was obtained through a novel connection to the concepts and results in free probability – a theory for non-commutative random variables developed by Voiculescu [13, 15]. The key idea in [2] is to compare the Gaussian model $X = \sum_i g_i A_i$ with the free model

$$X_{\text{free}} := \sum_i A_i \otimes s_i. \tag{5}$$

The elements $s_1, \dots s_n$ are freely independent semicircular random variables in a non-commutative probability space, $(\mathcal{A}, \tau)$, where τ is a faithful trace acting on the algebra $\mathcal{A}$. A concrete way to understand the free model is through the Gaussian models

$$X_N := \sum_i A_i \otimes G_i^N, \tag{6}$$

where each G_i^N is an independent $N \times N$ standard Wigner matrix (self-adjoint matrix with independent Gaussian variables). The free model X_{free} can be understood as the limiting object of the Gaussian models X_N: The weak asymptotic freeness property proven by Voiculescu [23] states that for any $p \in \mathbb{N}$,

$$\lim_{N\to\infty} \mathbb{E}[\operatorname{tr}(X_N^p)] = \operatorname{tr} \otimes \tau(X_{\text{free}}^p), \tag{7}$$

where τ is an abstract trace acting on the algebra generated by $s_1, \dots, s_n$. The strong asymptotic freeness property of Haagerup and Thorbjornsen [8] states that

$$\lim_{N\to\infty} \mathbb{E}\big[\|X_N\|\big] = \|X_{\text{free}}\|.$$

Despite being infinite-dimensional and more abstract, an important advantage of the free model is that its spectral statistics is easier to analyze using tools from operator theory: Lehner derived an exact formula for $\lambda_{\max}(X_{\text{free}})$ and Pisier showed that

$$\sigma(X) \le \|X_{\text{free}}\| \le 2\sigma(X). \tag{8}$$

This provides an avenue to establish that $\mathbb{E}[\|X\|]$ is close to $\sigma(X)$ if one can compare $\mathbb{E}[\|X\|]$ to $\|X_{\text{free}}\|$. See Section 2 of the full version [24] for more background on free probability.

Interpolation: The comparison approach in [2] is to interpolate between X_N and X by considering

$$X_t^N := \sum_{i=1}^N A_i \otimes \Big(\sqrt{t}\cdot D_i^N + \sqrt{1-t}\cdot G_i^N\Big), \tag{9}$$

where D_i^N are independent $N \times N$ diagonal matrices with i.i.d. standard Gaussians on the diagonal. Note that $X_0^N = X^N$ whose moments converge to those of X_{free} as $N \to \infty$ by (7), and X_1^N has the same moments as X such that $\mathbb{E}[\operatorname{tr}(X^p)] = \mathbb{E}[\operatorname{tr}((X_1^N)^p)]$ for every $p, N \in \mathbb{N}$. Therefore, the difference of the $2p$-moments of the eigenvalues of X and X_{free} can be written as

$$\Big|\mathbb{E}\big[\operatorname{tr}(X^{2p})\big] - \operatorname{tr}\otimes\tau(X_{\text{free}}^{2p})\Big| = \Big|\lim_{N\to\infty}\int_0^1 \frac{d}{dt}\mathbb{E}\big[\operatorname{tr}\big((X_t^N)^{2p}\big)\big]dt\Big|.$$

By a direct calculation of the derivative and some subtle observations, the integrand $\frac{d}{dt}\mathbb{E}[\operatorname{tr}((X_t^N)^{2p})]$ can be written as a weighted sum of differences of the form

$$\mathbb{E}\Big[\operatorname{tr}\Big(Y(X_t^N)^k Y(X_t^N)^{2p-2-k}\Big)\Big] - \operatorname{tr}\Big(Y\mathbb{E}\Big[(X_t^N)^k\Big]Y\mathbb{E}\Big[(X_t^N)^{2p-2-k}\Big]\Big), \tag{10}$$

for some matrix Y, where the second term is similar to the first term but with the expectation "pushed inside". This allows the use of the Gaussian covariance identity and some complex analysis to relate the sum of these differences to Tropp's matrix alignment parameter in [20], defined as

$$w(X_0^N, X_1^N) := \sup_{U,V,W \text{ unitary}} \big\|\mathbb{E}[X_0^N U X_1^N V X_0^N W X_1^N]\big\|^{\frac{1}{4}}.$$

Finally, some linear algebraic arguments are used to bound the matrix alignment parameter by

$$w(X_0^N, X_1^N) \le \nu(X_0^N)\cdot\sigma(X_0^N)\cdot\nu(X_1^N)\cdot\sigma(X_1^N).$$

Putting together these key steps, the conclusion in [2, Theorem 2.7] is that

$$\begin{aligned}\Big|\mathbb{E}\big\|X\big\|_{2p} - \big\|X_{\text{free}}\big\|_{2p}\Big| &= \Big|\mathbb{E}\big[\operatorname{tr}(X^{2p})\big]^{\frac{1}{2p}} - \operatorname{tr}\otimes\tau(X_{\text{free}}^{2p})^{\frac{1}{2p}}\Big| \\ &\le 2p^{\frac{3}{4}}\cdot\sigma(X)^{\frac{1}{2}}\cdot\nu(X)^{\frac{1}{2}}.\end{aligned} \tag{11}$$

Spectrum: To extend the arguments to bound the spectrum of X (not just the moments of the eigenvalues of X), the resolvent $R_X(z) := (zI - X)^{-1}$ of X at $z \in \mathbb{C}$ is considered. It is relatively straightforward to show that if $\|(zI - X_0^N)^{-1}\| \approx \|(zI - X_1^N)^{-1}\|$ for every point z in a sufficiently fine net, then $\operatorname{spec}(X_0^N) \approx \operatorname{spec}(X_1^N)$. To establish that $\|(zI - X_0^N)^{-1}\| \approx \|(zI - X_1^N)^{-1}\|$ for a fixed $z \in \mathbb{C}$, the large moments of the resolvents $\mathbb{E}[\operatorname{tr}((zI - X_0^N)^{-2p})]^{\frac{1}{2p}}$ and $\mathbb{E}[\operatorname{tr}((zI - X_1^N)^{-2p})]^{\frac{1}{2p}}$ are compared using the same key steps as above. In a slightly simplified form, the main result in [2, Theorem 2.1] is that

$$\begin{aligned}&\mathbb{P}\Big[\operatorname{spec}(X) \subseteq \operatorname{spec}(X_{\text{free}}) + C\sigma(X)^{\frac{1}{2}}\nu(X)^{\frac{1}{2}}\big((\log d)^{\frac{3}{4}} + \alpha\big)[-1,1]\Big] \\ &\ge 1 - e^{-\alpha^2},\end{aligned} \tag{12}$$

for all $\alpha \ge 0$, where C is a universal constant. The inequality in (4) follows as a corollary of (12).

Universality: Brailovskaya and van Handel [6] consider the general model $X := A_0 + \sum_{i=1}^n Z_i$, where A_0 is a deterministic Hermitian $d \times d$ matrix and $Z_1, \dots, Z_n$ are independent $d \times d$ random self-adjoint matrices with $\mathbb{E}[Z_i] = 0$ for $1 \le i \le n$. They defined the corresponding Gaussian matrix model as $G := A_0 + \sum_{i=1}^m g_i A_i$ for some $d \times d$ (deterministic) self-adjoint matrices $A_1, \dots, A_m$ such that $\mathbb{E}[G] = \mathbb{E}[X]$ and $\operatorname{Cov}(G) = \operatorname{Cov}(X)$. They proved the universality principle that, informally, if each $\|Z_i\|$ is small, then the spectrum of X is close to that of G such that

$$\begin{aligned}&\mathbb{E}[\|X\|_{2p}] \approx \mathbb{E}[\|G\|_{2p}],\ \mathbb{E}[(zI - X)^{-1}] \approx \mathbb{E}[(zI - G)^{-1}],\ \text{and} \\ &\operatorname{spec}(X) \approx \operatorname{spec}(G).\end{aligned} \tag{13}$$

This allows them to extend the results in [2] to the general model, with many applications beyond the Gaussian model as the general model captures discrete random variables.

An interpolation as in (9) is used to compare the spectral statistic of the Gaussian and the general model, but the calculations are considerably more involved with several new technical ingredients such as the moment-cumulant formula, a new trace inequality [6, Proposition 5.1], and Poincare's lemma.

1.2 Our Results

We derandomize the general results in [2, 6] by designing deterministic polynomial time algorithms to find outcomes that satisfy the guarantees in the probabilistic statements. We show applications in designing deterministic algorithms for matrix discrepancy, spectral sparsification, and constructions of expander graphs.

1.2.1 Norms of Eigenvalues. Our first result is a one-sided derandomization of (11). By standard Gaussian concentration inequalities, $\|g\|^2 \asymp n$ with high probability. In the context of randomized algorithms, (11) guarantees that with high probability, we can sample a vector $x \in \mathbb{R}^n$ with $\|x\|^2 \asymp n$ such that

$$\begin{aligned}\left\|\sum_{i=1}^n x(i)\cdot A_i\right\|_{2p} &\le \|X_{\text{free}}\|_{2p} + 2p^{\frac{3}{4}}\cdot\sigma(X)^{\frac{1}{2}}\cdot v(X)^{\frac{1}{2}}\\ &\le 2\sigma(X) + 2p^{\frac{3}{4}}\cdot\sigma(X)^{\frac{1}{2}}\cdot v(X)^{\frac{1}{2}},\end{aligned}$$

where the last inequality follows from (8). Such a result can be useful in designing randomized algorithms, but often there are additional requirements such as $x \in \{\pm1\}^n$ or x satisfies some linear constraints. The following theorem provides a derandomization of the upper bound of (11) incorporating additional constraints.

Theorem 1.1 (Deterministic Partial Coloring, Simplified Version of Theorem 3.2). *Let $A_1,\dots,A_n$ be $d\times d$ Hermitian matrices. Let $\mathcal{H}\subseteq\mathbb{R}^n$ be a linear subspace of dimension $(1-\varepsilon)n$. For any $p\ge 4$, there is a deterministic polynomial time algorithm to find a vector $x\in[-1,1]^n\cap\mathcal{H}$ with $|\{i \mid x(i)\in\{\pm1\}\}|\ge\delta n$ such that*

$$\left\|\sum_{i=1}^n x(i)\cdot A_i\right\|_{2p} \le K_{\delta,\varepsilon}\cdot\|X_{\text{free}}\|_{2p} + O\big(p^{\frac{3}{4}}\cdot\sigma(X)^{\frac{1}{2}}\cdot v(X)^{\frac{1}{2}}\big),$$

where $K_{\delta,\varepsilon} := (1-(\varepsilon+\delta)+2/n)^{-\frac{1}{2}}$ and $\sigma(X), v(X)$ are defined as in (1), (3) respectively.

In the full version of Theorem 1.1 in Theorem 3.2, when there are no box constraints $x\in[-1,1]^n$ and linear subspace constraints $x\in\mathcal{H}$, the algorithm returns a vector $x\in\mathbb{R}^n$ with $\|x\|^2\asymp n$ and no multiplicative constant $K_{\delta,\varepsilon}$, matching the upper bound of [2, Theorem 2.7] stated in (11).

1.2.2 Deterministic Matrix Discrepancy. Theorem 1.1 can be interpreted as a general "partial coloring" result in matrix discrepancy theory. We demonstrate that it can be directly applied to obtain a deterministic algorithm for the matrix Spencer problem, matching the randomized algorithm in [4].

Theorem 1.2 (Deterministic Matrix Spencer). *Given $n\times n$ symmetric matrices $A_1,\dots,A_n$ with $\|A_i\|\le 1$ and $\|A_i\|_F^2\le r^2$ for $1\le i\le n$, there is a deterministic polynomial time algorithm to compute a coloring $x:[n]\to\{-1,1\}$ such that*

$$\left\|\sum_{i=1}^n x(i)\cdot A_i\right\|_{2p} \lesssim \sqrt{n} + p^{\frac{3}{4}}\cdot n^{\frac{1}{4}}\cdot r^{\frac{1}{4}}.$$

In particular, when $r\lesssim n/\log^3 n$ and $p\asymp\log n$, then

$$\left\|\sum_{i=1}^n x(i)\cdot A_i\right\| \lesssim \left\|\sum_{i=1}^n x(i)\cdot A_i\right\|_{2p} \lesssim \sqrt{n}.$$

In [4], the inequality in (4) was used to lower bound the Gaussian measure of the norm ball $\mathcal{K} := \{x\in\mathbb{R}^n \mid \left\|\sum_{i=1}^n x(i)\cdot A_i\right\|\le 1\}$, so that Rothvoss' result [16] can be applied to obtain a partial coloring. In Theorem 1.2, we apply Theorem 1.1 to obtain a partial coloring without using Rothvoss' result, providing a simpler and more direct approach to matrix discrepancy.

The matrix discrepancy result in [4] has found an interesting application in spectral sparsification of directed graphs [9], and Theorem 1.2 implies a deterministic algorithm for this application. See Section 4.4 of the full version [24] for more details.

1.2.3 Full Spectrum. We also derandomize the result in (12), deterministically finding an outcome with the full spectrum close to that of X_{free}.

Theorem 1.3 (Deterministic Full Spectrum, informal). *Let $A_1,\dots,A_n$ be $d\times d$ Hermitian matrices. There is a deterministic polynomial time algorithm to compute a vector $x\in\mathbb{R}^n$ with $\|x\|^2\le n$ such that*

$$\operatorname{spec}\Big(\sum_{i=1}^n x(i)\cdot A_i\Big)\subseteq\operatorname{spec}(X_{\text{free}})+[-\eta,\eta] \quad \text{and}$$

$$\operatorname{spec}(X_{\text{free}})\subseteq\operatorname{spec}\Big(\sum_{i=1}^n x(i)\cdot A_i\Big)+[-\eta,\eta],$$

where $\eta\lesssim\sigma(X)^{\frac{1}{2}}\cdot v(X)^{\frac{1}{2}}\cdot(\log d)^{\frac{3}{4}}$ and $\sigma(X), v(X)$ are defined as in (1), (3) respectively.

As an application of this result, we give a deterministic analog of the spiked detection model. Given a set of unit vectors $v_1,\dots,v_n$ with signal values $\theta_1,\dots,\theta_n>0$, we determinstically construct a pesudorandom matrix W such that for all j, the maximum eigenvector of $\theta_j v_j v_j^\top + W$ has nontrivial correlation with v_j if and only if $\theta_j>1$. This is a derandomization of [3, Theorem 3.1]. When $\theta_j\le 1$, the vectors v_j are "efficiently hidden" by W with respect to the spectral detection algorithm (analogous to how a planted clique of size $\sqrt{d}$ is "hidden" by a random graph of d vertices). This application shows that Theorem 1.3 can be used to construct pseudorandom matrices that capture the behavior of not only the eigenvalues of random matrices but also their eigenvectors. Due to the space limit, we omit the proof of Theorem 1.3 and related applications. See the full version [24] for details.

1.2.4 General Model. Let $X := A_0 + \sum_{i=1}^n Z_i$, where A_0 is a deterministic Hermitian $d\times d$ matrix and $Z_1,\dots,Z_n$ are independent $d\times d$ random self-adjoint matrices with $\mathbb{E}[Z_i]=0$ for $1\le i\le n$. For derandomization, we assume that each Z_i has discrete support so that it can be enumerated efficiently, which is satisfied in random graph models.

The following are two related models. Let $G := A_0 + \sum_{i=1}^m g_i A_i$ be the corresponding Gaussian model of X with $\mathbb{E}[G] = \mathbb{E}[X]$ and $\mathrm{Cov}(G) = \mathrm{Cov}(X)$. Note that the distribution of G is uniquely defined but the representation as $A_0 + \sum_{i=1}^m g_i A_i$ is not unique. Let $X_{\mathrm{free}} := A_0 \otimes 1 + \sum_{i=1}^m A_i \otimes s_i$ be the corresponding free model of G as defined in (5), where $s_1, \ldots, s_m$ is a free semicircular family and 1 is the identity element.

We provide one-sided derandomization of the norm bounds in [6].

Theorem 1.4 (Deterministic Moment Universality, informal). *Let $X := A_0 + \sum_{i=1}^n Z_i$, where A_0 is a deterministic Hermitian $d \times d$ matrix and $Z_1, \ldots, Z_n$ are independent $d \times d$ random self-adjoint matrices with $\mathbb{E}[Z_i] = 0$ for $1 \le i \le n$. Assume that $\|Z_i\| \le r$ with probability one for $1 \le i \le n$, and the support size of each Z_i is polynomially bounded. Then, for any $p \in \mathbb{N}$, there is a deterministic polynomial time algorithm to find a matrix $Z_i' \in \mathrm{supp}(Z_i)$ for $1 \le i \le n$ such that*[1]

$$\mathrm{tr}\left(\left(A_0 + \sum_{i=1}^n Z_i'\right)^{2p}\right)^{\frac{1}{2p}} \le \mathrm{tr}\otimes\tau\left(X_{\mathrm{free}}^{2p}\right)^{\frac{1}{2p}} + \tilde{O}\left(p^{\frac{3}{4}} \cdot \sigma(X)^{\frac{1}{2}} \cdot \nu(X)^{\frac{1}{2}} + p^{\frac{2}{3}} \cdot \sigma(X)^{\frac{2}{3}} \cdot r^{\frac{1}{3}}\right).$$

The following is a stronger bound for the operator norm, with the leading constant being one, which is important for applications such as constructing near-Ramanujan graphs.

Theorem 1.5 (Deterministic Norm Universality, informal). *Under the same setting in Theorem 1.4, there is a deterministic polynomial time algorithm to find a matrix $Z_i' \in \mathrm{supp}(Z_i)$ for $1 \le i \le n$ such that*

$$\left\|A_0 + \sum_{i=1}^n Z_i'\right\| \le \|X_{\mathrm{free}}\| + \tilde{O}\left(\sigma(X)^{\frac{1}{2}} \cdot \nu(X)^{\frac{1}{2}} + \sigma(X)^{\frac{2}{3}} \cdot r^{\frac{1}{3}}\right).$$

1.2.5 Deterministic Expander Constructions. Brailovskaya and van Handel [6, Section 3.2] showed that their results for the general model can be applied to give new probabilistic constructions of expander graphs. Our deterministic algorithms in Theorem 1.4 and Theorem 1.5 imply the following consequences:

(1) (Edge-Signing Model, see Section 7.2.1 in [24]): For any k-regular graph $G = (V, E)$, with $k \ge \mathrm{polylog}\,|V|$, there is a deterministic algorithm to find an edge signing of G whose signed adjacency matrix has eigenvalues bounded by $2\sqrt{k} \cdot (1 + \tilde{O}(k^{-\frac{1}{6}}))$.
(2) (Permutation Model, see Section 7.2.2 in [24]): For any $k \ge \mathrm{polylog}\, d$, there is a deterministic algorithm to compute $2k$ perfect matchings over d vertices such that their union has spectral radius $2\sqrt{2k} \cdot (1 + \tilde{O}(k^{-\frac{1}{6}}))$.
(3) (Lift Model, see Section 7.2.3 in [24]): Let $G = (V, E)$ be a simple (possibly nonregular) graph with maximum degree $k_{\max} \ge \mathrm{polylog}(|V|)$. There is a deterministic algorithm to compute an m-lift of G whose new eigenvalues are bounded by $\lambda \cdot (1 + \tilde{O}(k_{\max}^{-\frac{1}{6}}))$ where λ is the spectral radius of the universal cover of G.
(4) (Group-Labeled Lift Models, see Section 7.2.3 in [24]): Let $G = (V, E)$ be a k-regular simple graph and Γ be a finite group with $k \ge \mathrm{polylog}(|\Gamma|, |V|)$. There is a deterministic algorithm to compute a Γ-lift of G whose new eigenvalues are bounded by $2\sqrt{k} \cdot (1 + \tilde{O}(k)^{-\frac{1}{6}})$

To our knowledge, these are the first polynomial time deterministic algorithms for all these settings when $k \ge \mathrm{polylog}(|V|)$. See Section 7.2 of the full version [24] for more background and details.

1.3 Proof Overview

Our proofs show that the concepts and techniques in the theory of free probability are not only useful for mathematical analyses, but also useful for efficient computations.

A classical technique for derandomization is the method of conditional expectation. Given a set of variables $X_1, \ldots, X_n$, and a function $f(x_1, \ldots, x_n)$, we wish to find an outcome $x_1, \ldots, x_n \in \mathrm{supp}(X_1, \ldots, X_n)$ such that $f(x_1, \ldots, x_n) \le \mathbb{E}[f(X_1, \ldots, X_n)]$. The method of conditional expectation allows us to find such an outcome as long as we can compute the expectation of f conditioned on the outcomes of any subset of variables. However, for moments or norms of general random matrices, this is typically difficult to compute in polynomial time. For example, if X is a $d \times d$ random matrix with independent entries, computing $\mathbb{E}[\mathrm{tr}(X^{\log d})]$ would take $d^{\log d}$ time. Thus, previous derandomizations of matrix concentration inequalities (e.g., [25]) use "pessimistic estimators" to estimate the conditional expectation of various spectral statistics of random matrices. By nature, these cannot recover the sharp matrix concentration results in [2] and [6], as these involve inequalities such as Golden-Thompson, which, just like (2), cannot distinguish between the commutative and non-commutative settings.

In this work, we observe that using the "non-crossing" structure in the theory of free probability, the moments of the free model X_{free} can be computed efficiently in polynomial time via a natural recursive formula; see Section 4 for computational aspects. This suggests the following framework for de-randomizing the new matrix concentration inequalities: start with the operator X_{free}, and at each iteration, replace a "little bit" of it with a finite-dimensional random matrix. If this random matrix is simple enough (e.g., with pairwise independent entries), then we can deterministically find an outcome to replace the random matrix. This method can be viewed as a free version of the method of conditional expectation, where instead of computing the expectation of a spectral statistic on X, we compute the spectral statistic on the corresponding operator X_{free}, which is a main theme in this work.

1.3.1 Derandomizing Random Gaussian Matrix Models by Brownian Walks. We elaborate more technical ideas in the derandomization of the random Gaussian matrix model. A natural derandomization strategy is to interpolate from X_0^N to X_1^N in (9) as was done in [2], where in each iteration t we replace a little bit of each G_i^N by a deterministic diagonal matrix $(D_i^N)_t$. The choice of lifting each Gaussian variable g_i to a random matrix D_i^N with independent diagonal entries was used in [2] to apply the multivariate Gaussian integration-by-parts formula in a nice way to compute the derivative $\frac{d}{dt}\mathbb{E}[\mathrm{tr}(f(X_{N,t}))]$. However, for the purposes of derandomization, this lifting technique introduces some issues. The major one is that the expectation of the Gaussian matrix moments are not easy to compute, as previously noted. It is also not clear what

[1] The notation $\tilde{O}(\cdot)$ hides some logarithmic term.

should be the scalar value of $x(i)$ given a deterministic sampling of a diagonal matrix D_i^N.[2]

Interpolation with $\mathbf{X}_{\text{free}}$: In our analysis, we use an alternative interpolation method that avoids lifting the random $d \times d$ matrix X to a $dN \times dN$ random matrix. Instead, we directly interpolate between X and $\bar{X}_{\text{free}} = \sum_{i=1}^n A_i \otimes s_i$. For any vector $x \in \mathbb{R}^n$, we define $A(x) := \sum_{i=1}^n x(i) \cdot A_i$, and the mixed operator

$$A_t(x) := A_0 \otimes 1 + A(x) \otimes 1 + \sqrt{1-t} \cdot \bar{X}_{\text{free}}.$$

By definition, $A_0(\vec{0}) = X_{\text{free}}$ and $A_1(g) = X$ when $g \sim \mathcal{N}(0, I)$. Given a potential function, say $\Phi(t,x) = \operatorname{tr} \otimes \tau (A_t(x)^{2p})^{\frac{1}{2p}}$, we can evaluate it at any point using dynamic programming, because the "non-crossing structure" in free probability gives recursive formulas for computing moments (while such formulas do not exist for random matrices). See Section 4 for these computational aspects.

Brownian Walks and Pairwise Independent Updates: To interpolate between $\vec{0}$ and g, we consider the stochastic process $x_t \sim \mathcal{N}(0, \sqrt{t}I)$, where $x_0 = \vec{0}$ and $x_1 = g$. Rather than integrating only the derivative with respect to t as in [2], we integrate with respect to the stochastic process $x_t \sim \mathcal{N}(0, \sqrt{t}I)$. Given the potential function $\Phi(t,x)$, we evaluate

$$\mathbb{E}[\Phi(1,x_1)] - \Phi(0,x_0) = \int_0^1 \mathbb{E}[d\Phi(t,x_t))].$$

The main task is to bound $\mathbb{E} d\Phi(t,x_t)$. By Ito's lemma,

$$\mathbb{E}[d\Phi(t,x_t)] = \frac{\partial}{\partial t}\Phi(t,x_t)dt + \mathbb{E}\left[\nabla\Phi(t,x_t)^\top dx_t\right] + \frac{1}{2}\mathbb{E}\left[dx_t^\top \nabla^2 \Phi(t,x_t) dx_t\right],$$

where $\nabla\Phi(t,x)$ and $\nabla^2\Phi(t,x)$ are the gradient and Hessians of $\Phi(t,x)$ as a function of $x \in \mathbb{R}^n$. The key to derandomizing this process is to observe that dx_t need not be a Gaussian vector. Since the update formula only depends on the first and second order statistics of the entries of dx_t, it suffices to take dx_t to be a vector with *pairwise independent* entries.

We derandomize this interpolation process by discretizing the stochastic integral. In particular, we break the interval $[0,1]$ into steps $\eta, 2\eta, \ldots, T\eta$ where $\eta = 1/T$. At each step t, we deterministically find an update vector y_t such that $\Phi(t+\eta, x_t + \sqrt{\eta}\cdot y_t) \le \Phi(t,x_t)$. Such an update can be computed in polynomial time because 1) the potential function Φ is in terms of X_{free}, and so can be evaluated efficiently, and 2) the *expected* potential change is small over a pairwise independent distribution of update vectors, so we only need to search over polynomially many possibilities for y_t. In addition to being easily derandomizable, this method of "Brownian interpolation"[3] can also be used to handle linear constraints on x_t using the sticky Brownian walk method of Lovett and Meka [12]. These linear constraints are required for many applications, especially for matrix discrepancy problems. The sticky Brownian walk method can be derandomized using the eigenspace methods in [11]. This is similar to pairwise independent updates but with stronger guarantees, which will be elaborated in Section 3.

Bounding the Expected Change: The key to bounding the expected change in potential lies in the property that that the sum of freely independent semicircular random variables is also semicircular. This means that

$$\sqrt{1-t}\cdot\bar{X}_{\text{free}} = \sqrt{1-(t+\eta)}\cdot\bar{X}_{\text{free}} + \sqrt{\eta}\cdot\bar{X}'_{\text{free}},$$

where $\bar{X}'_{\text{free}}$ is a freely independent copy of $\bar{X}_{\text{free}}$. Under this discretization,

$$\begin{aligned}\frac{\partial}{\partial t}\Phi(t,x_t)dt &\approx \Phi(t+\eta,x_t) - \Phi(t,x_t)\\ &\approx -\operatorname{tr}\otimes\tau\left(\left(A_t(x_t) + \sqrt{\eta}\cdot X'_{\text{free}}\right)^{2p}\right)^{\frac{1}{2p}} + \operatorname{tr}\otimes\tau\left(A_t(x_t)^{2p}\right)^{\frac{1}{2p}}.\end{aligned}$$

Using second-order Taylor expansion, the derivatives in terms of dx_t can be expressed as

$$\begin{aligned}&\nabla\Phi(t,x_t)^\top dx_t + \frac{1}{2}dx_t^\top\nabla^2\phi(t,x_t)dx_t \approx \Phi(t,x_t+\sqrt{\eta}y_t) - \Phi(t,x_t)\\ &= \operatorname{tr}\otimes\tau\left(\left(A_t(x_t) + \sqrt{\eta}\cdot A(y_t)\right)^{2p}\right)^{\frac{1}{2p}} - \operatorname{tr}\otimes\tau\left(A_t(x_t)^{2p}\right)^{\frac{1}{2p}}\end{aligned}$$

Thus, we can interpret this step as transferring one unit of mass from the free part of the mixed operator $\sqrt{1-t}\cdot X_{\text{free}}$ to the finite-dimensional part $A(x_t)$. The term

$$\nabla\Phi(t,x_t)^\top dx_t + \frac{1}{2}dx_t^\top\nabla^2\phi(t,x_t)dx_t$$

is the loss incurred by adding more mass to the finite dimensional part, while the term $\frac{\partial}{\partial t}\Phi(t,x_t)dt$ is the gain incurred by taking away mass from the free part.

Free Probability and Gaussian Analysis: To show that the loss and gain terms approximately cancel each other out, there are two steps. First, we apply Taylor expansion on the potential function to show that both expressions are dominated by the second order term in the Taylor expansion. From this second order expansion, using basic properties of free independence, we show that the difference of the loss term and the gain term is exactly a weighted sum of differences of the form in (10). This provides a better insight on the subtle step where the expectation is "pushed inside" in [2]. Once we reached this form, we can reuse the Gaussian analysis and the linear algebraic arguments proven in [2] to bound the difference.

To summarize, the Brownian walk method uses techniques developed in discrepancy theory to ensure that the outcome satisfies box constraints and linear constraints, while the Gaussian and complex analysis developed in [2] is used in bounding the potential increase in each infinitesimal step where we replace a bit of X_{free} by a bit of $A(y_t)\otimes 1$.

Full Spectrum and Multiplicative Weight Update: Given this Brownian interpolation method, the proof of Theorem 1.3 has a very similar structure, where the potential function is changed to the moments of the resolvent such that

$$\Phi_t^z(x_t) = \operatorname{tr}\otimes\tau\left(|zI\otimes 1 - A_0\otimes 1 - A(x_t)\otimes 1 - \sqrt{1-t}\cdot\bar{X}_{\text{free}}|^{-2p}\right)^{\frac{1}{2p}}.$$

To control the full spectrum, the probabilistic analysis in [2] bounds the moments of the resolvent for many different points $z \in \mathbb{C}$ and

[2] If the potential function f is convex, then Jensen's inequality gives $f(\operatorname{tr}((D_i^N))) \le \operatorname{tr}(f(D_i^N))$, so one could potentially take $x(i)$ to be the average of the entries in D_i^N. However, if f is not convex, then it becomes unclear how $x(i)$ should be sampled. We note that the proof of Theorem 1.3 requires interpolating non-convex functions to control the full spectrum.

[3] We note that this method can be interpreted as a geometric version of the method of conditional expectation, since at each time t, the operator $\sqrt{1-t}\cdot X_{\text{free}}$ is an infinite dimensional approximation of the random matrix $\sqrt{1-t}\sum_i A_i g_i$, which captures the randomness that has yet to be derandomized. The $\operatorname{tr}\otimes\tau$ operator then approximates the expectation over this randomness. The selection of an update to minimize the potential function is analogous to selecting a deterministic outcome that minimizes the expectation of the random part conditioned on the outcome we selected.

applies a union bound. For derandomization, an extra ingredient in this proof is to use the multiplicative weight update method to combine many potential functions into one.

1.3.2 Derandomizing General Random Matrix Models by Random Swap. In [6], Brailovskaya and van Handel showed, via interpolation arguments, that the spectral statistics of the general random matrix model $X = A_0 + \sum_{i=1}^n Z_i$ is close to those of the corresponding Gaussian matrix model. Such a method does not directly lead to a polynomial time derandomization scheme for two reasons. First, as in the previous setting, the expected Gaussian moments are not easy to compute in general. To address this issue, we will bypass the Gaussian approximation of Z, and instead directly compare the spectral statistics of Z with those of the free model

$$X_{\text{free}} := A_0 + X_1 + X_2 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are freely independent that satisfy $\mathrm{Id} \otimes \tau[X_i] = \mathbb{E}[Z_i]$ and $\mathrm{Cov}(X_i) = \mathrm{Cov}(Z_i)$ for all $i \in [n]$. This operator X_{free} is exactly the free approximation of the Gaussian matrix corresponding to the general model Z. See Section 7.1 in the full version [24] for the construction of this free model.

The second issue is that unlike in the Gaussian model, where we can approximately decompose a Gaussian vector into the sum of pairwise independent random vectors, the general random matrix model Z admits no such simple decomposition. Thus, we make the assumption that each Z_i is a discrete random matrix with polynomial support size. This still encapsulates many important applications such as expander graph constructions. For example, if Z is the signing of the adjacency matrix of a graph, then each Z_i corresponds to the signing of an edge, with only two outcomes.

Random Swap: The non-Gaussian structure of Z makes it difficult to apply the Brownian walk method to the general matrix model. Our assumption of polynomial support size of each Z_i suggests a more direct application of the method of conditional expectation. Consider the following "random swap" procedure, in which we randomly selects an index $i \in [n]$ and perform the swap $X_i \leftarrow Z_i \otimes 1$. Then, we find an outcome of Z_i that minimizes the desired potential function. For example, if our potential is the $2p$-th moment, then we want to bound the quantity

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{Z_i}\left[\operatorname{tr}\otimes\tau\left((X_{\text{free}} + Z_i \otimes 1 - X_i)^{2p}\right)\right] - \operatorname{tr}\otimes\tau(X_{\text{free}}^{2p}) \qquad (14)$$

There are two new technical ingredients in our proofs. An important step in our analysis is to derive a "*semicircular integration by parts formula*" (see Proposition 6.5 in the full version [24]). This is crucially used to compute the potential update for our interpolation with the free model, where the formula allows us to replace "dependent" random variables by freely independent random variables in the Taylor expansion of (14). One can view this as a replacement of the moment-cumulant formula used in [6] to carry out the interpolation with the Gaussian model. A technical remark is that we exploited the property that $\max_i \|X_i\|$ is bounded, while it does not hold in the Gaussian interpolation setting where $\|G_i\|$ is unbounded.

Another technical ingredient is the use of the barrier method, developed by [5], to derandomize the norm universality result. We analyze the potential function $\operatorname{tr}\otimes\tau((\lambda I - X)^{-2p})$ where $\lambda > \lambda_{\max}(X)$. In each iteration of the random swap algorithm, we update the barrier $\lambda \leftarrow \lambda + \delta$ and show that the potential function does not increase. This allows us to bypass the step of having to simultaneously control the resolvent norm at many points in [6], which was much more challenging than in the Gaussian setting of [2] and required proving new concentration inequalities for the general model. Again, we exploited the property that $\max_i \|X_i\|$ is bounded for the barrier method to be well-defined. These considerations further highlight the advantages of directly interpolating between Z and the free model in our analysis.

1.4 Organization of the Paper

In the remainder of this extended abstract, we start by presenting some background on free probability and random matrix theory in Section 2. Then, we provide the analysis of the deterministic partial coloring algorithm (Theorem 1.1) together with the application to matrix Spencer problem (Theorem 1.2) in Section 3, which shows a flavor of the derandomization using free probability. The computational issues of the derandomization algorithms are addressed in Section 4.

The proof of the full spectrum result in Theorem 1.3 and related applications can be found in Section 5 of the full version [24]. The proof of deterministic moment/norm universality results in Theorem 1.4 and Theorem 1.5, and their applications to deterministic expander constructions can be found in Section 7 of [24].

2 Preliminaries

Free probability is a theory for non-commutative random variables, which are often characterized by the spectral distributions of matrices and operators. We refer the readers Section 2 of [24] or Nica and Speicher's textbook [15] for more background on free probability. In this section, give a brief survey of some connections between free probability and random matrix theory.

A fundamental result in random matrix theory is Wigner's semicircle law, which states that the limiting distribution of the eigenvalues of many random symmetric matrices is the semicircle distribution, a fundamental object in free probability. This motivated Voiculescu to build a connection between the theory of free probability and random matrix theory.

2.1 Gaussian Random Matrices and Semicircular Family

We begin by defining the Gaussian Orthogonal Ensemble and outlining its relationship to the family of freely independent semicircular elements. This will allow us to formally define the free model X_{free} and study its relationship to the Gaussian model.

Definition 2.1 (Gaussian Orthogonal Ensemble (GOE)). A GOE matrix of dimension d can be expressed as

$$G = \sum_{1\le i\le j\le d} \frac{g_{i,j}}{\sqrt{d}} E_{i,j}, \quad \text{where } E_{i,j} = \chi_i\chi_j^\top + \chi_j\chi_i^\top.$$

In other words, a GOE matrix is a random symmetric matrix with independent Gaussian entries on the upper-diagonal. Each diagonal entry has variance $2/d$ and each off-diagonal entry has variance $1/d$. Under this normalization, we have $\mathbb{E}[\operatorname{tr}(G)] = 0$ and $\mathbb{E}[\operatorname{tr}(G^2)] = 1$.

A classical result in random matrix theory is that if $G^{(N)}$ is a Gaussian GOE matrix of dimension N, then for all $p \in \mathbb{N}$,

$$\lim_{N\to\infty} \mathbb{E}\,\mathrm{tr}((G^{(N)})^p) = \begin{cases} C_{p/2} & p \text{ is even} \\ 0 & p \text{ is odd,} \end{cases}$$

where C_k is the k^{th} Catalan number. This coincides with the moments of a standard semicircular element. It follows that in the limit as $N \to \infty$, the empirical eigenvalue distribution of $G^{(N)}$ tends towards that of a semicircle distribution.

An important result by Voiculescu is that independent GOE matrices are asymptotically freely independent. More formally, let $G_1, \ldots, G_m$ be independent $N \times N$ GOE matrices, $p_1, \ldots, p_k \in \mathbb{N}$, and $i_1, \ldots, i_k \in [m]$ be such that $i_1 \neq i_2, i_2 \neq i_3, \ldots, i_{k-1} \neq i_k$. Then,

$$\lim_{N\to\infty} \mathbb{E}\,\mathrm{tr}\left[\left(G_{i_1}^{p_1} - C_{p_1} I\right)\left(G_{i_2}^{p_2} - C_{p_2} I\right)\cdots\left(G_{i_k}^{p_k} - C_{p_k} I\right)\right] = 0.$$

As free independence determines the joint distribution of a set of random variables (see Section 2.1.3 in [24]), it follows that independent GOE matrices converge in distribution to freely independent semicircular elements.

Theorem 2.2 (Weak Convergence of GOE Matrices). *Let $G_1^{(N)}, \ldots, G_n^{(N)}$ be independent $N \times N$ GOE matrices. Let $(\mathcal{A}, \tau)$ be a probability space where $\mathcal{A}$ is generated by a family of freely independent semicircular elements $s_1, \ldots, s_n$. Then, for any non-commutative polynomial $p \in \mathbb{C}\langle x_1, \ldots, x_n\rangle$,*

$$\lim_{N\to\infty} \mathbb{E}\,\mathrm{tr}(p(G_1^{(N)}, \ldots, G_n^{(N)})) = \tau(p(s_1, \ldots, s_n)).$$

For concreteness, we can think of the infinite-dimensional operators $s_1, s_2, \ldots$ from Fock space (see Theorem 2.18 in [24]) as a representation of the limit objects $\{\lim_{N\to\infty}(G_i^{(N)})\}_{i\in\mathbb{N}}$. We refer to the C^*-algebra generated by $s_1, s_2, \ldots$ as the semicircular algebra.

Using the weak convergence in Theorem 2.2 and properties of Gaussian matrices, we can establish analogous properties of semicircular elements. For example, the following property follows as the sum of independent Gaussian matrices is also a Gaussian matrix.

Fact 1 (Sum of Freely Independent Semicircular Elements). *Suppose s_1 and s_2 are freely independent semicircular elements with zero mean. Then $s_1 + s_2$ is a semicircular element with variance $\tau(s_1^2) + \tau(s_2^2)$.*

2.2 Semicircular Matrices

We have seen that a semicircular operator is a good approximation for a large GOE matrix. What should be the free operator approximation of general Gaussian matrices? The main result of [2] is that under certain "intrinsic freeness" conditions, a random matrix with Gaussian entries can be approximated by a matrix with semicircular elements.

Formally, let $(\mathcal{A}, \tau)$ be the semicircular algebra, and $\mathcal{M}_d(\mathbb{C})$ be the set of $d \times d$ complex matrices. We consider the algebra $\mathcal{M}_d(\mathbb{C}) \otimes \mathcal{A}$, which can be represented as the space of $d \times d$ matrices whose entries are elements in $\mathcal{A}$. We say that an operator $X \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{A}$ is a "semicircular matrix" if each of its matrix entries has the semicircular distribution. This is analogous to a Gaussian matrix in the classical setting. These operators can be characterized as sums of freely independent semicircular elements with matrix coefficients.

Definition 2.3 (Semicircular Matrix). Let $X \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{A}$. X is called a semicircular matrix if there exist matrices $A_0, A_1, \ldots, A_n$ such that $X = A_0 \otimes 1 + \sum_{i=1}^n A_i \otimes s_i$ where $s_1, \ldots, s_n$ are freely independent semicircular elements. X is called centered if $A_0 = 0$.

The main result in [2], as stated in (12), is that the Gaussian model $X = A_0 + \sum_{i=1}^n g_i A_i$ where $g_1, \ldots, g_n$ are independent Gaussians can be approximated by the free model $X_{\text{free}} = A_0 \otimes 1 + \sum_{i=1}^n A_i \otimes s_i$ where $s_1, \ldots, s_n$ are freely independent semicircular elements. In their approach, X_{free} is approximated by $Nd \times Nd$ random matrix X_N for large enough N, defined as

$$X_N = A_0 \otimes I_N + \sum_{i=1}^n A_i \otimes G_i^N$$

where $G_1^N, \ldots, G_n^N$ are i.i.d. GOE matrices. As $N \to \infty$, the spectral distribution of X^N tends towards that of X_{free}. This follows from Theorem 2.2, as for each $p \in \mathbb{N}$, $(X^N)^p$ and $(X_{\text{free}})^p$ can be expressed as matrices whose entries are polynomials in $G_1^N, \ldots, G_n^N$ and $s_1, \ldots, s_n$ respectively. Since for any continuous function f, there is a sequence of polynomials that pointwise converge to f, it follows from Theorem 2.2 that

$$\mathrm{tr} \otimes \tau(f(X_{\text{free}})) = \lim_{N\to\infty} \mathbb{E}\,\mathrm{tr}[f(X_N)]. \tag{15}$$

A key advantage of working with the free model is that there are formulas for bounding and calculating the norms of semicircular matrices. These formulas are derived using the concrete realizations of the free semicircular random variables as operators in a Fock space (see Theorem 2.18 in [24]).

Theorem 2.4 (Pisier's Inequality). *Let $X_{\text{free}} = A_0 \otimes 1 + \sum_{i=1}^n A_i \otimes s_i$ be a centered semicircular matrix. Then*

$$\|X_{\text{free}}\| \leq \|A_0\| + \left\|\sum_{i=1}^n A_i^* A_i\right\|^{\frac{1}{2}} + \left\|\sum_{i=1}^n A_i A_i^*\right\|^{\frac{1}{2}}.$$

Theorem 2.5 (Lehner's Formula [10]). *Let $X_{\text{free}} = A_0 + \sum_{i=1}^n A_i \otimes s_i$ be a semicircular matrix. Then*

$$\|X_{\text{free}}\| = \sup_{\substack{Y \succ 0 \\ \varepsilon \in \{\pm 1\}}} \lambda_{\max}\left(\varepsilon A_0 + Y^{-1} + \sum_{i=1}^n A_i^* Y A_i\right).$$

A corollary [3, Eq. (1.6)] is that if X is a random matrix with the same covariance profile as X_{free}, then

$$\|X_{\text{free}}\| = \sup_{\substack{Y \succ 0 \\ \varepsilon \in \{\pm 1\}}} \lambda_{\max}\left(\varepsilon \mathbb{E}[X] + Y^{-1} + \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot Y \cdot (X - \mathbb{E}[X])\right]\right). \tag{16}$$

Finally, we state a lemma that will be useful for matrix discrepancy.

Lemma 2.6 (Trace of Projection). *Let $A_0, A_1, \ldots, A_n$ be symmetric matrices and $s_1, \ldots, s_n$ be freely independent semicircular elements. Let $f : \mathbb{C} \to \mathbb{C}$ be a continuous convex function. Let P is an $n \times n$ projection matrix such that $P^2 = P$, and $\tilde{A}_1, \ldots, \tilde{A}_n$ be matrix projections such that $\tilde{A}_i = \sum_{j=1}^n P(i,j) A_j$. Then*

$$\mathrm{tr} \otimes \tau\left(f\left(A_0 \otimes 1 + \sum_{i=1}^n \tilde{A}_i \otimes s_i\right)\right) \leq \mathrm{tr} \otimes \tau\left(f\left(A_0 \otimes 1 + \sum_{i=1}^n A_i \otimes s_i\right)\right).$$

3 Moment Concentration Inequalities with Linear Constraints

The goal in this section is to prove Theorem 1.1 and its application to matrix discrepancy in Theorem 1.2.

3.1 Technical Statements

Let $X = \sum_{i=1}^n g_i A_i$ be the Gaussian model, where each $A_i \in \mathbb{C}^{d\times d}$ is an arbitrary self-adjoint matrix. The non-commutative Khintchine inequality of Lust-Piquard and Pisier states that

$$\mathbb{E}\big[\operatorname{Tr}(X^{2p})\big]^{\frac{1}{2p}} \le \sqrt{2p-1}\cdot \operatorname{Tr}\left(\left(\sum_{i=1}^n A_i^2\right)^p\right)^{\frac{1}{2p}}.$$

Since $\mathbb{E}[\|X\|] \le \mathbb{E}[\|X\|^{2p}]^{\frac{1}{2p}} \asymp \mathbb{E}[\operatorname{Tr}[X^{2p}]]^{\frac{1}{2p}}$ for $p \asymp \log d$, this implies that

$$\mathbb{E}[\|X\|] \lesssim \sqrt{\log d}\cdot \sigma(X) \quad \text{where} \quad \sigma(X)^2 := \left\|E[X^2]\right\| = \left\|\sum_{i=1}^n A_i^2\right\|.$$

This upper bound can be achieved for diagonal matrices, where $A_1,\dots,A_n$ are commutative with each other. However, it is far from tight for non-commutative matrices, e.g., $\mathbb{E}[\|X\|] \asymp \sigma(X)$ when X is a GOE matrix. A main result in [2] quantities the non-commutativeness of X by a new parameter $\nu(X)$ and demonstrates that the $\sqrt{\log d}$ factor can be removed when $\nu(X)$ is sufficiently small. Their approach is by comparing the Gaussian model to the free matrix model in (5). The following is the formal statement of (11) in the introduction.

Theorem 3.1 ([2, Theorem 2.7]). *Let $A_0, A_1, \dots, A_n$ be $d\times d$ self-adjoint matrices. Let $X = A_0 + \sum_{i=1}^n g_i A_i$ be the Gaussian model and $X_{\text{free}} = A_0 + \sum_{i=1}^n A_i \otimes s_i$ be the corresponding free model. Then, for any $p \ge 4$,*

$$\left|\mathbb{E}\|X\|_{2p} - \|X_{\text{free}}\|_{2p}\right| \le 2p^{\frac{3}{4}}\cdot \sigma(X)^{\frac{1}{2}}\cdot \nu(X)^{\frac{1}{2}},$$

where $\operatorname{Cov}(X)$ is the $d^2\times d^2$ matrix with

$$\operatorname{Cov}(X)_{ij,kl} = \mathbb{E}[X_{ij}\overline{X_{kl}}] \quad \text{and}$$
$$\nu(X)^2 := \|\operatorname{Cov}(X)\| = \left\|\sum_{i=1}^n \operatorname{vec}(A_i)\operatorname{vec}(A_i)^\top\right\|.$$

This result implies that $\mathbb{E}[\|X\|] \lesssim \sigma(X)$ when $\nu(X)\cdot(\log d)^{\frac{3}{2}} \lesssim \sigma(X)$. This consequence is the key of the recent major progress in the matrix Spencer problem [4], which is used to lower bound the Gaussian measure of the norm ball $\mathcal{K} := \{x\in\mathbb{R}^n \mid \|\sum_{i=1}^n x_i\cdot A_i\| \le 1\}$ so that Rothvoss' result [16] can be applied to obtain a partial coloring.

Using techniques from both free probability and algorithmic discrepancy theory, we provide a direct and simpler approach to obtain a partial coloring that satisfies additional box constraints and linear constraints. The following is the full version of Theorem 1.1.

Theorem 3.2 (Deterministic Partial Coloring). *Suppose that we are given $d\times d$ Hermitian matrices $A_0, A_1, \dots, A_n$ with matrix parameters[4]:*

$$\sigma^2 = \left\|\sum_{i=1}^n A_i^2\right\|, \quad \nu^2 = \left\|\sum_{i=1}^n \operatorname{vec}(A_i)\operatorname{vec}(A_i)^\top\right\|,$$
$$\sigma_*^2 = \sup_{\|y\|,\|z\|=1}\sum_{i=1}^n \langle y, A_i z\rangle^2.$$

Let $\bar{X}_{\text{free}} := \sum_{i=1}^n A_i\otimes s_i$ be the centered free matrix model for our input matrices. Let $b, c\in\mathbb{R}^n$ be constraint vectors such that $b\le 0\le c$. Let $\varepsilon,\delta\ge 0$ be constants such that $\varepsilon+\delta<1$. Let $\mathcal{H}\subseteq\mathbb{R}^n$ be a linear subspace of dimension greater than $(1-\varepsilon)n$. Then, for any $p\ge 4$, there is a deterministic algorithm with running time $\text{poly}(n, \frac{\sigma\nu}{\sigma^*}) \le \text{poly}(n,d)$ that finds a vector $x\in\mathbb{R}^n$ satisfying

1. (Box constraints and linear constraints:) $b\le x\le c$ *and* $x\in\mathcal{H}$.
2. (Partial coloring:) *Either* $|\{i \mid x_i = b_i \text{ or } x_i = c_i\}| \ge \delta n$ *or* $\|x\|^2 = n$.
3. ($2p$-norm bound:) *Let* $A(x) := \sum_{i=1}^n x_i A_i$. *Then*
$$\|A_0 + A(x)\|_{2p} \le \left\|A_0\otimes 1 + \sqrt{K_{\delta,\varepsilon}}\cdot\bar{X}_{\text{free}}\right\|_{2p} + O\big(p^{\frac{3}{4}}\sqrt{\sigma\nu}\big),$$
where $K_{\delta,\varepsilon} := 1/(1-\varepsilon-\delta-\frac{3}{n})$.

In the case where there are no box constraints and no linear constraints (i.e., $b=-\infty\cdot 1$, $c=+\infty\cdot 1$, and $\varepsilon=\delta=0$), Theorem 3.2 always returns a vector x with $\|x\|^2 = n$ such that

$$\|A_0 + A(x)\|_{2p} \le \left\|A_0\otimes 1 + \bar{X}_{\text{free}}\right\|_{2p} + O(p^{\frac{3}{4}}\sqrt{\sigma\nu} + \sigma/\sqrt{n}).$$

This provides a one-sided derandomization of Theorem 3.1, with an essentially negligible additional error term of $\sigma/\sqrt{n}$.

In the general case when $\varepsilon,\delta>0$, there is a multiplicative factor loss of $K_{\varepsilon,\delta}$ on top of the additive loss of $O(p^{3/4}\sqrt{\sigma\nu})$ in Theorem 3.1. In many applications such as for matrix discrepancy, a constant multiplicative loss is tolerable. In exchange, we can handle additional linear and box constraints (e.g., $x\in[\pm 1]^n$), which broadens the scope of where such a bound can be applied.

3.2 Brownian Walk Algorithm

We provided an overview of the algorithm in Section 1.3, using a modified interpolation from X to $\bar{X}_{\text{free}}$ and the Brownian walk approach in algorithmic discrepancy theory to implement the interpolation.

In the algorithm, we use the following potential function to govern the moments of our mixed operators. Given a vector $x\in\mathbb{R}^n$, define the potential function[5]

$$\Phi(t,x) := \operatorname{tr}\otimes\tau\left(\left(A_0\otimes 1 + A(x)\otimes 1 + \sqrt{K_{\delta,\varepsilon}\cdot(1-t)}\cdot\bar{X}_{\text{free}}\right)^{2p}\right)^{\frac{2}{p}}$$
$$= \left\|A_0\otimes 1 + A(x)\otimes 1 + \sqrt{K_{\delta,\varepsilon}\cdot(1-t)}\cdot\bar{X}_{\text{free}}\right\|_{2p}^4.$$

The idea of the Brownian walk algorithm is to divide the interval $[0,1]$ into small discrete steps of size η. We start with $t=0$, $m=0$,

[4] We assume that $1/\text{poly}(d) \le \max_{i=0}^n \|A_i\| \le \text{poly}(d)$. This assumption ensures that the matrix parameters σ,ν,σ^* are all upper bounded by $\text{poly}(n,d)$ and lower bounded by $1/\text{poly}(n,d)$.

[5] The $2/p$ in the exponent normalizes the potential to be the 4-th power of the $2p$-norm. This choice of taking the 4-th power is only to make calculations more convenient. The analysis would still work if we use $1/2p$ in the exponent, as what we do in Section 7.4 in the full version [24].

and $x = 0$. At the m-th step, we update $t \leftarrow t + \eta/n$ and $x_{m+1} \leftarrow x_m + y_m$, where the update y_m is chosen so that $\|x_m\|^2$ increases by a small but non-trivial amount, while subject to the box and linear-subspace constraints required to satisfy conditions (1) and (2) in Theorem 3.2. The multiplicative factor $K_{\delta,\varepsilon}$ ensures that the decrease of the potential function due to the shrinking of the free part is sufficiently large enough to offset the increase of the potential function caused by the update y_m. The key in the analysis is to control the increase of the potential function.

Proposition 3.3 (Potential Increase). *Let $\eta > 0$ be a sufficiently small step size. Given the setting in Theorem 3.2, for any $x \in \mathbb{R}^n$ with $\|x\|^2 \le n$ and any $t \in [0, 1-\eta)$, there exists a subspace $\mathcal{H}' \subseteq \mathbb{R}^n$ of dimension at least $(\varepsilon+\delta)n+2$, such that for any $y \in \mathcal{H}'$ with $\|y\|^2 = 1$,*

$$\Phi\Big(t + \frac{\eta}{n},\ x + \sqrt{\eta}\cdot y\Big) \le \Phi(t,x) + O\Big(\frac{\eta}{n}\cdot p^3\cdot\sigma^2\cdot\nu^2\Big).$$

In particular, the inequality holds as long as

$$\eta \le \min\left\{\frac{\sigma^4\cdot\nu^4}{n^3\cdot\rho^6\cdot\rho_0^2},\ \frac{\sigma^2\cdot\nu^2}{n^3\cdot\rho^4},\ \frac{1}{n^2}\right\} \le \frac{1}{\mathrm{poly}(n,d)}, \tag{17}$$

where $\rho := \max\left\{\frac{2\cdot\sigma}{\sqrt{n}}, \sigma_*\right\}$ and $\rho_0 = \|A_0\| + \rho\sqrt{n}$.

Deterministic Sticky Brownian Walk Algorithm

- Initialize $m = 0$, $x_0 = 0$, and η is a small enough number as defined in (17).
- Initialize $|\mathcal{F}_0| = [n]$ as the set of alive/active coordinates.
- For $m \in 0, 1, 2, \ldots, \lfloor \frac{n}{\eta} \rfloor$,
  (1) Let $\mathcal{H}_m$ be the subspace defined by the linear constraints

$$\mathcal{H}_m := \mathcal{H} \cap \{\langle y, x_m\rangle = 0 \text{ and } y(i) = 0 \text{ for all } i \notin \mathcal{F}_m\}.$$

  (2) Apply Proposition 3.3 to find a unit vector $y \in \mathcal{H}_m$ such that

$$\begin{aligned} &\Phi\Big(\tfrac{(m+1)\cdot\eta}{n},\ x_m + \sqrt{\eta}\cdot y\Big) \\ \le\ &\Phi\Big(\tfrac{m\cdot\eta}{n}, x_m\Big) + O\Big(\tfrac{\eta}{n}\cdot p^3\cdot\sigma^2\cdot\nu^2\Big). \end{aligned} \tag{18}$$

  (3) Update $x_{m+1} \leftarrow x_m + \sqrt{\eta}\cdot y$.
  (4) For any entry i satisfies $x(i) \le b(i) + \frac{1}{n}$ or $x(i) \ge c(i) - \frac{1}{n}$, remove i from $\mathcal{F}_m$.
  (5) If $|\mathcal{F}_{m+1}| < (1-\delta)n$, then round each $i \notin \mathcal{F}_m$ to $b(i)$ or $c(i)$ depending on which one is closer to $x(i)$. Then return x_{m+1} and terminate the algorithm.
- Return $x_{\lfloor \frac{n}{\eta} \rfloor}$.

We first prove Theorem 3.2 by analyzing this algorithm assuming Proposition 3.3, the proof of which can be found in [24].

Proof of Theorem 3.2. First, we argue that Step (2) of the algorithm always succeeds. Since the algorithm has not terminated, the number of alive variables is at least $|\mathcal{F}_m| \ge (1-\delta)n$. Thus, the dimension of the (bad) subspace $\mathcal{H}_m^{\perp}$ is at most $(\delta+\varepsilon)n+1$, with at most εn constraints from $\mathcal{H}$, at most δn constraints from $\{y(i) = 0 \text{ for all } i \in \mathcal{F}_m\}$, and one constraint from $\langle y, x_m\rangle = 0$. By Proposition 3.3, there exists a (good) subspace $\mathcal{H}'$ of dimension at least $(\varepsilon+\delta)n+2$ such that any $y \in \mathcal{H}'$ with $\|y\|_2 = 1$ satisfies (18). Therefore, the subspace $\mathcal{H}' \cap \mathcal{H}_m$ is non-empty, and hence Step (2) always succeeds.

Now, we check that each of the three conditions in Theorem 3.2 are satsified. For the box constraints, when a coordinate is still alive such that $x(i) > b(i) + \frac{1}{n}$ and $x(i) < c(i) - \frac{1}{n}$, the choice of η in (17) ensures that the coordinate will not change by more than $1/n$, hence the box constraints will never be violated. For the linear subspace constraint, each update step $y \in \mathcal{H}$ by Step (1) of the algorithm, and hence each $x_m \in \mathcal{H}$.

For the partial coloring condition, since $y \perp x_m$, it follows that $\|x_{m+1}\|^2 = \|x_m\|^2 + \eta$. If the algorithm terminates after the for-loop finished, then $\left\|x_{\lfloor n/\eta \rfloor}\right\|^2 \ge n - \eta$, and we can round it to a vector x such that $\|x\|^2 = n$. Note that the increase of the potential function due to the rounding is negligible, as $\left\|A(x) - A(x_{n/\eta})\right\|_{2p} \le \sigma^*\sqrt{\eta} \le \sqrt{\sigma\cdot\nu}$, using the facts that $\|A(y)\|_{2p} \le \sigma^*\|y\|$ and $\sigma^* \le \min\{\sigma, \nu\}$. If the algorithm terminates before the for-loop finished, then there are at least δn frozen coordinates, so the rounding in Step (5) of the algorithm ensures that the values of these coordinates are either $b(i)$ or $c(i)$, satisfying the partial coloring condition. The increase of the potential function is again negligible, as $\|A(x) - A(x_T)\|_{2p} \le \sigma_*\|x - x_T\| \le \sigma_* \le \sqrt{\sigma\nu}$ since $|x(i) - x_T(i)| \le 1/n$ for all i.

For the $2p$-norm bound condition, let $T \le n/\eta$ be the iteration in which the algorithm terminated. Then

$$\begin{aligned} &\|A_0 + A(x_T)\|_{2p}^4 \\ \le\ &\left\|A_0\otimes 1 + A(x_T)\otimes 1 + \sqrt{K_{\varepsilon,\delta}\cdot\left(1 - \frac{T\eta}{n}\right)}\cdot\bar{X}_{\mathrm{free}}\right\|_{2p}^4 \\ \le\ &\left\|A_0\otimes 1 + \sqrt{K_{\varepsilon,\delta}}\cdot\bar{X}_{\mathrm{free}}\right\|_{2p}^4 + O(p^3\sigma^2\nu^2), \end{aligned}$$

where the first inequality follows from Lemma 2.6, and the second inequality follows by repeatedly applying Proposition 3.3 for at most n/η iterations. Finally, we use the inequality that $y - x \le (y^4 - x^4)^{\frac{1}{4}}$ for real numbers $y > x > 0$ to conclude that

$$\|A_0 + A(x_T)\|_{2p} \le \left\|A_0\otimes 1 + \sqrt{K_{\varepsilon,\delta}}\cdot\bar{X}_{\mathrm{free}}\right\|_{2p} + O(p^{\frac{3}{4}}\sqrt{\sigma\nu}).$$

□

3.3 Applications to Matrix Discrepancy

An important tool in discrepancy theory is Rothvoss's algorithmic partial coloring theorem for convex bodies. The partial coloring theorem states that if a symmetric convex body has large enough Gaussian volume, then it contains a point with a constant fraction of entries in ±1. Rothvoss proved a constructive version of this result using a randomized Gaussian projection algorithm, with wide-ranging applications in algorithmic discrepancy theory, spectral sparsification, and numerical linear algebra.

Theorem 3.4 (Randomized Partial Coloring for Convex Body [16]). *Let $\varepsilon < \frac{1}{60000}$, $\delta \asymp \log\frac{1}{\varepsilon}$ be constants and $U \subseteq \mathbb{R}^n$ be a subspace of dimension at least $(1-\delta)n$. Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a symmetric, convex set, and $x_0 \in [-1,1]^n$. Suppose the Gaussian measure of $\mathcal{K}$ is at least $e^{-\varepsilon n}$. There is a randomized polynomial time algorithm to find a vector x satisfying*

(1) $x \in \mathcal{K}$,

(2) $x + x_0 \in [-1,1]^n$,
(3) *at least* εn *coordinates of* $x + x_0$ *are in* $\{-1,1\}$.

Using Theorem 3.2, we provide a derandomization of Rothvoss's result when $\mathcal{K}$ is the p-norm ball of matrices (with constant Gaussian measure). The proof is based on the idea in [4] to apply partial coloring on projected matrices.

Proposition 3.5 (Deterministic Partial Coloring for Matrix Norms). *Let* $A_1, \ldots, A_n$ *be symmetric matrices with matrix parameters* $\left\|\sum_{i=1}^n A_i^2\right\| = \sigma^2$ *and* $\sum_{i=1}^n \|A_i\|_F^2 = nf^2$. *Let* $x_0 \in [-1,1]^n$ *be an initial vector. For any* $p \geq 2$, *there is a deterministic polynomial time algorithm to find a partial coloring* x *satisfying*

(1) $\|A(x)\|_{2p} \lesssim \sigma + p^{\frac{3}{4}}\sqrt{\sigma f}$,
(2) $x \perp x_0$ *and* $x_0 + x \in [-1,1]^n$,
(3) *at least* $\frac{n}{4}$ *coordinates of* $x_0 + x$ *are in* $\{-1,1\}$.

Proof. Let M be the $n \times n$ matrix where $M(i,j) = \langle A_i, A_j\rangle$. Let $\mathcal{H}$ be the subspace orthogonal to the top $\frac{n}{3}$ eigenspace of M. Let P be the projection matrix into $\mathcal{H}$. Define the projected matrices as

$$\tilde{A}_i = \sum_{j=1}^n P(i,j) \cdot A_j.$$

We apply Theorem 3.2 with input matrices $A_0 = 0, \tilde{A}_1, \ldots, \tilde{A}_n$, linear subspace constraint $\mathcal{H} \cap \{x_0\}^\perp$, box constraints $b(i) = -1 - x_0(i)$, $c(i) = 1 - x_0(i)$, and parameter $\delta = \frac{1}{4}$. Specify the remaining input parameters as follows. Since $\dim\left(\mathcal{H} \cap \{x_0\}^\perp\right) \geq \frac{2}{3}n - 1$, we set $\varepsilon = \frac{1}{3} + \frac{1}{n}$ so that $\varepsilon + \delta < 1$. Let $\tilde{M}$ be the $n \times n$ matrix where $\tilde{M}(i,j) = \langle \tilde{A}_i, \tilde{A}_j\rangle$. Note that $\tilde{M} = PMP \succcurlyeq 0$, with maximum eigenvalue at most $\frac{3}{n}\operatorname{Tr}(M) = 3f^2$. The input matrices satisfy the matrix parameters

$$\sigma^2 = \left\|\sum_{i=1}^n \tilde{A}_i^2\right\| \leq \left\|\sum_{i=1}^n A_i^2\right\| = \sigma^2 \quad \text{and}$$

$$v^2 = \left\|\sum_{i=1}^n \operatorname{vec}(\tilde{A}_i)\operatorname{vec}(\tilde{A}_i)^\top\right\| = \left\|\tilde{M}\right\| \leq 3f^2.$$

Let x be the output vector guaranteed by Theorem 3.2. Since $x \in \mathcal{H}$, we have

$$A(x) = A(Px) = \sum_{i=1}^n \left(\sum_{j=1}^n P(i,j)\cdot x(j)\right)\cdot A_i$$
$$= \sum_{j=1}^n x(j)\cdot\left(\sum_{i=1}^n P(i,j)\cdot A_i\right) = \sum_{j=1}^n x(j)\cdot \tilde{A}_j = \tilde{A}(x),$$

where the second-to-last equality uses that P is a symmetric matrix. Let $\bar{X}_{\text{free}} = \sum_i \tilde{A}_i \otimes s_i$ be the free model in Theorem 3.2. By Theorem 2.4, $\left\|\bar{X}_{\text{free}}\right\|_{2p} \leq \left\|\bar{X}_{\text{free}}\right\| \leq 2\sigma$. Thus, the $2p$-norm bound guarantee in Theorem 3.2 implies that

$$\left\|\tilde{A}(x)\right\| \lesssim \sigma + p^{\frac{3}{4}}\sqrt{\sigma f}.$$

The box constraints that $b \leq x \leq c$ ensures that $x + x_0 \in [-1,1]^n$. The linear subspace constraint ensures that $x \perp x_0$. The partial coloring condition in Theorem 3.2 ensures that there are at least $\frac{n}{4}$ coordinates of $x + x_0$ are in $\{-1,1\}$ (as the possibility that $\|x + x_0\|^2 = n$ implies all coordinates are in $\{-1,1\}$). □

3.3.1 Matrix Spencer Problem. In the matrix Spencer problem, we are given $n \times n$ matrices $A_1, \ldots, A_n$ with $\|A_i\| \leq 1$ and for $i \in [n]$. The goal is to find a coloring $x : [n] \to [-1,1]$ so that $\|A(x)\| \lesssim O(\sqrt{n})$. Bansal, Jiang, and Meka [4] proved that such a coloring exists when each A_i has $\|A_i\|_F^2 \leq O(n/\log^3 n)$, by using (4) from [2] to show that the Gaussian measure of the matrix p-norm ball is large and then applying Rothvoss' partial coloring theorem. We use Proposition 3.5 to provide a deterministic polynomial time algorithm with the same guarantee.

Proof of Theorem 1.2. Consider the following procedure for finding a full coloring with low discrepancy.

- Initialize $\mathcal{F}_0 = [n]$ and $x_0 = 0$.
- For $m = 1$ to $O(\log n)$
  (1) Let $x_0^{|\mathcal{F}_m|}$ be the restriction of x_m to the coordinates in $\mathcal{F}_m$.
  (2) Apply Proposition 3.5 to find a vector $y^{|\mathcal{F}_m|} \in \mathbb{R}^{|\mathcal{F}_m|}$ such that $x_0^{|\mathcal{F}_m|} + y^{|\mathcal{F}_m|} \in [-1,1]^{|\mathcal{F}_m|}$, with at least $\frac{1}{4}|\mathcal{F}_m|$ coordinates in $\{1,-1\}$.
  (3) Update $x_{m+1} \leftarrow x_m + y^{|\mathcal{F}_m|}$. Let $\mathcal{F}_{m+1}$ be the coordinates in x_{m+1} with value in $(-1,1)$.

We apply the partial coloring algorithm iteratively until all coordinates become ± 1. In each iteration m, $\mathcal{F}_m$ is the set of coordinates of x_m with value in $(-1,1)$. Let $\sigma_m^2 = \left\|\sum_{i\in\mathcal{F}_m} A_i^2\right\|$ and $f_m^2 = \frac{1}{|\mathcal{F}_m|}\sum_{i\in\mathcal{F}_m}\|A_i\|_F^2$. Proposition 3.5 guarantees that $|\mathcal{F}_{m+1}| \leq \frac{3}{4}|\mathcal{F}_m|$ and

$$\|A(x_{m+1}) - A(x_m)\|_{2p} = \left\|A(y^{|\mathcal{F}_m|})\right\|_{2p} \lesssim \sigma_m + p^{\frac{3}{4}}\sqrt{\sigma_m f_m}.$$

Since $f_m^2 \leq r^2$ and $\sigma_m^2 \leq |\mathcal{F}_m| \leq (\frac{3}{4})^m \cdot n$, this gives a final discrepancy bound of

$$\sum_m \left(\sigma_m + p^{\frac{3}{4}}\sqrt{\sigma_m f_m}\right) \lesssim \sqrt{n} + p^{\frac{3}{4}}\cdot n^{\frac{1}{4}}\cdot r^{\frac{1}{4}},$$

as this is a geometric sum which is dominated by the first term. □

3.3.2 Spectral Sparsification. Given a weighted directed Eulerian graph $\vec{G} = (V, E, w)$, the goal of the Eulerian sparsification problem is to find a reweighted subgraph $\vec{H}$ of $\vec{G}$ such that $\vec{H}$ is sparse and

$$\left\|L_G^{\dagger/2}(\vec{L}_{\vec{G}} - \vec{L}_{\vec{H}})L_G^{\dagger/2}\right\| \leq \varepsilon,$$

where $\vec{L}_{\vec{G}}$ is the directed Laplacian matrix of $\vec{G}$.

Jambulapati, Sachdeva, Sidford, Tian, and Zhao [9] combine the matrix discrepancy result from [4] with an effective resistance decomposition technique to construct a Eulerian sparsifer of size $O(\varepsilon^{-2}n\log n + \varepsilon^{-4/3}n\log^{5/3} n)$ in randomized polynomial time. Using the deterministic partial coloring result in Theorem 3.2 as a black box, we obtain a deterministic polynomial time algorithm with the same sparsity guarantee (but with a much slower runtime). We refer the reader to [9] for details.

4 Computation on Free Semicircular Matrices

In this section, we discuss the computational aspects of free semicircular matrices. In particular, we provide formulas for computing

moments and resolvents of random variables on $\mathcal{M}_d(\mathbb{C})\otimes\mathcal{A}$, which are essential for the efficiency of our derandomization algorithms.

4.1 Computation of Moments

Lemma 4.1 (Moments Computation). *Let $X_{\text{free}} = A_0 + \sum_{i=1}^n A_i \otimes s_i$ be a general semicircular matrix. For any integer $p \geq 2$,*

$$\varphi[X_{\text{free}}^p] = \varphi[X_{\text{free}}^{p-1}]\cdot A_0 + \sum_{i=1}^n\sum_{k=0}^{p-2}\varphi[X_{\text{free}}^k]\cdot A_i\cdot\varphi[X_{\text{free}}^{p-2-k}]\cdot A_i.$$

Let $B \in \mathcal{M}_d(\mathbb{C})$ be any matrix. For integers $p, q \geq 1$,

$$\begin{aligned}\varphi[X_{\text{free}}^p(B\otimes 1)X_{\text{free}}^q] &= \varphi[X_{\text{free}}^p(B\otimes 1)X_{\text{free}}^{q-1}]\cdot A_0\\ &+\sum_{i=1}^n\sum_{k=0}^{p-1}\varphi[X_{\text{free}}^k]\cdot A_i\cdot\varphi[X_{\text{free}}^{p-1-k}(B\otimes 1)X_{\text{free}}^{q-1}]\cdot A_i\cdot\\ &+\sum_{i=1}^n\sum_{k=0}^{q-2}\varphi[X_{\text{free}}^p(B\otimes 1)X_{\text{free}}^k]\cdot A_i\cdot\varphi[X_{\text{free}}^{q-2-k}]\cdot A_i\cdot 1_{\{q\geq 2\}}.\end{aligned}$$

These formulas can be derived from the basic properties of free semicircular moments. For example, the first formula was used in some proofs in Van Handel's survey (see [22, Lemma 4.4]) but was not stated explicitly. We omit details of the proofs and refer the reader to Section 6 of [24].

Observe that the formulas lead to a polynomial time algorithm for computing moments.

Lemma 4.2 (Efficient Computation for Moments). *Let B_1, B_2 be arbitrary $d\times d$ matrices. For any $p, q \in \mathbb{N}$, there is a polynomial time algorithm to compute the quantity* $\operatorname{tr}\otimes\tau((B_1\otimes 1)X_{\text{free}}^p(B_2\otimes 1)X_{\text{free}}^q)$.

Proof. Note that $\operatorname{tr}\otimes\tau((B_1\otimes 1)X_{\text{free}}^p(B_2\otimes 1)X_{\text{free}}^q) = \operatorname{tr}(B_1\cdot\varphi[(X_{\text{free}})^p(B_2\otimes 1)(X_{\text{free}})^q])$, thus it suffices to evaluate $\varphi[X_{\text{free}}^p(B_2\otimes 1)X_{\text{free}}^q]$. The formulas in Theorem 4.1 provide a natural recurrence for a dynamic programming algorithm to compute this quantity. Specifically, there are two types of subproblems: one of the form $\varphi[X_{\text{free}}^k]$, and one of the form $\varphi[X_{\text{free}}^k(B_2\otimes 1)X_{\text{free}}^\ell]$. Denote the size of each subproblem as $k+\ell$. Then we see from the recursive formulas that each subproblem only relies on solutions to subproblems of a smaller size. The base case for the first kind of subproblem is given by $\varphi[X_{\text{free}}^0] = I$ and $\varphi[X_{\text{free}}^1] = A_0$. The second type of subproblems eventually reduces to the first type when k or ℓ reaches 0. The total number of subproblems is at most $O(pq)$. This leads to a polynomial time dynamic programming algorithm. □

From Lemma 4.2, we can compute all moments of X_{free} in polynomial time by simply taking $B_1, B_2 = I$. The more general formula with arbitrary B_1, B_2 is used for computation with the Hessian matrix of the potential function in Section 3.

4.2 From Moments to Resolvents

Now, we show how the moments of X_{free} can be used to compute the moments of its resolvents. This is done via a re-centering trick that allows us to approximate the resolvent moment by a convergent power series. The efficient computation of moments of resolvents of X_{free} is used in the derandomization algorithms for Theorem 1.3 and Theorem 1.5.

Let $X \in \mathcal{M}_d(\mathbb{C})\otimes\mathcal{A}$ be a self-adjoint semicircular matrix. For $z = \lambda + \varepsilon i \notin \operatorname{spec}(X)$, we would like to compute

$$\operatorname{tr}\otimes\tau\left(|z1 - X|^{-2p}\right) = \operatorname{tr}\otimes\tau\left(\left(\varepsilon^2 1 + (\lambda 1 - X)^2\right)^{-p}\right),$$

where we use the shorthand 1 for $I\otimes 1$. Note that $Y = (\lambda 1 - X)^2$ is an element in $\mathcal{M}_d(\mathbb{C})\otimes\mathcal{A}$ with nonnegative spectrum and $\|Y\| \leq 2(\lambda^2 + \|X\|^2)$. Therefore, it suffices to compute the moments of the resolvent of the form

$$\operatorname{tr}\otimes\tau\left(\left(\varepsilon^2 1 + Y\right)^{-p}\right),$$

where Y has nonnegative spectrum and bounded norm. The following lemma shows how to compute this using a convergent power series.

Lemma 4.3 (Power Series for Resolvent). *Let Y be an element in $\mathcal{M}_d(\mathbb{C})\otimes\mathcal{A}$ with nonnegative spectrum with $\|Y\| \leq u$. For any $\varepsilon, \delta > 0$ and $p \in \mathbb{N}$, there is a deterministic algorithm to compute $\operatorname{tr}\otimes\tau((\varepsilon^2 1 + Y)^{-p})$ up to an additive error δ in time $O(kT(k))$, where $k \in \operatorname{poly}(u, 1/\varepsilon^2, p, \log(1/\delta))$ and $T(k)$ is the time required to compute the k-th moment of Y.*

Proof. We re-center the power series at the point $u1 - Y$ such that

$$\begin{aligned}\operatorname{tr}\otimes\tau((\varepsilon^2 1 + Y)^{-p}) &= \operatorname{tr}\otimes\tau((\varepsilon^2 1 + u1 - u1 + Y)^{-p})\\ &= (\varepsilon^2 + u)^{-p}\cdot\operatorname{tr}\otimes\tau\left(\left(1 - \frac{1}{\varepsilon^2 + u}(u1 - Y)\right)^{-p}\right).\end{aligned}$$

Since Y has nonnegative spectrum and $u \geq \|Y\|$, it holds that $\|u1 - Y\| \leq u$ and thus $\frac{\|u1-Y\|}{\varepsilon^2+u} < 1$. This implies that the p-th moment of the resolvent can be expressed as a converging power series using the Taylor expansion for the function $(1-x)^{-p}$ such that

$$\operatorname{tr}\otimes\tau((\varepsilon^2 1 + Y)^{-p}) = \frac{1}{(\varepsilon^2+u)^p}\cdot\sum_{k=0}^\infty\binom{p+k-1}{k}\frac{\operatorname{tr}\otimes\tau((u1-Y)^k)}{(\varepsilon^2+u)^k}.$$

Since $\|u1 - Y\| \leq u$, the k-th term in the series is bounded by

$$\frac{1}{(\varepsilon^2+u)^p}\binom{p+k-1}{k}\frac{u^k}{(\varepsilon^2+u)^k} \leq \left(\frac{k+p}{\varepsilon^2+u}\right)^p\cdot\exp\left(-\frac{k\varepsilon^2}{\varepsilon^2+u}\right),$$

where we used $1 - y \leq e^{-y}$ for the inequality. Thus, the power series is dominated by a geometrically decreasing sequence. In particular, if k satisfies $\frac{k}{\log k} \geq Cp\cdot\frac{\varepsilon^2+u}{\varepsilon^2}$ for a large enough constant C, then the rightmost expression is upper bounded by $\exp\left(-\frac{k\varepsilon^2}{2(\varepsilon^2+u)}\right)$. Truncating the power series at such a k would yield error at most

$$\begin{aligned}\sum_{l\geq k}\exp\left(-\frac{l\varepsilon^2}{2(\varepsilon^2+u)}\right) &\leq \frac{\exp(-\frac{k\varepsilon^2}{2(\varepsilon^2+u)})}{1-\exp(-\frac{\varepsilon^2}{2(\varepsilon^2+u)})}\\ &\leq \frac{4(\varepsilon^2+u)}{\varepsilon^2}\exp\left(-\frac{k\varepsilon^2}{2(\varepsilon^2+u)}\right),\end{aligned}$$

where we used $1 - e^{-y} \geq \frac{y}{2}$ for $y \in (0,1)$ in the last inequality. The truncation error is bounded by δ if k also satisfies

$$k \geq \frac{2(u+\varepsilon^2)}{\varepsilon^2}\ln\left(\frac{4(\varepsilon^2+u)}{\delta\varepsilon^2}\right).$$

Finally, for each of the $j \leq k$ term in the power series, we can compute $\operatorname{tr}\otimes\tau((u1 - Y)^j) = \sum_{l=0}^j\binom{j}{l}\operatorname{tr}\otimes\tau(Y^l)$ in at most $T(k)$ time. □

The power series leads to a polynomial time algorithm to compute the moments of the resolvent, which is required for the algorithm of Theorem 1.3 (see Section 5 in [24]). We also observe that the moments of resolvent with $z > \lambda_{\max} + \varepsilon \in \mathbb{R}$ can be computed efficiently, which will be used in the barrier method for proving the universality of operator norm in Theorem 1.5 (see Section 7.5 in [24]).

Corollary 4.4 (Efficient Computation for Resolvents). *Let $X \in \mathcal{M}_d(\mathbb{C}) \otimes \mathcal{A}$ be a self-adjoint semicircular matrix. Let $z = \lambda + \varepsilon i$ with $\varepsilon > 0$. There is a deterministic algorithm to compute $\mathrm{tr} \otimes \tau(|z1 - X|^{-2p})$ up to an additive error δ in polynomial time with respect to $\sigma(X), d, \lambda, p, 1/\varepsilon^2$ and $\log(1/\delta)$).*

Furthermore, the same conclusion holds when $z = \lambda_{\max}(X) + \varepsilon \in \mathbb{R}$.

Proof. We directly apply Lemma 4.3 with $z = \lambda + \varepsilon i$ and $Y = \lambda 1 - X$. Note that the moments of Y can be computed from Lemma 4.2 in polynomial time via binomial expansion, where we use the upper bound $\|Y\| \le u = 2\lambda^2 + 8\sigma(X)^2$ provided by Pisier's bound in Theorem 2.4.

For the furthermore part, we write $\lambda 1 - X = \frac{\varepsilon}{2} 1 + (\lambda - \frac{\varepsilon}{2}) 1 - X$. By construction, $Y := (\lambda - \frac{\varepsilon}{2}) 1 - X$ has nonnegative spectrum, with norm at most $4\sigma(X) + \frac{\varepsilon}{2}$. Then, the conclusion follows from Lemma 4.3, using Lemma 4.2 to compute the moments of Y. □

Acknowledgments

Hong Zhou was supported in part by National Key R&D Program of China (No. 2023YFA1010202), Natural Science Foundation of China (No. 12571336) and Fuzhou University Research Funding (No. XRC-25098). Lap Chi Lau was supported in part by NSERC Discovery Grant.

References

[1] Sanjeev Arora and Satyen Kale. 2016. A combinatorial, primal-dual approach to semidefinite programs. *J. ACM* 63, 2 (2016), Art. 12, 35. doi:10.1145/2837020

[2] Afonso S. Bandeira, March T. Boedihardjo, and Ramon van Handel. 2023. Matrix concentration inequalities and free probability. *Invent. Math.* 234, 1 (2023), 419–487. doi:10.1007/s00222-023-01204-6

[3] Afonso S. Bandeira, Giorgio Cipolloni, Dominik Schröder, and Ramon van Handel. 2024. Matrix Concentration Inequalities and Free Probability II. Two-sided Bounds and Applications. *arXiv preprint* arXiv:2406.11453 (2024). arXiv:2406.11453 [math.PR] https://arxiv.org/abs/2406.11453

[4] Nikhil Bansal, Haotian Jiang, and Raghu Meka. 2023. Resolving matrix Spencer conjecture up to poly-logarithmic rank. In *STOC'23—Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. ACM, New York, 1814–1819. doi:10.1145/3564246.3585103

[5] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. 2012. Twice-Ramanujan sparsifiers. *SIAM J. Comput.* 41, 6 (2012), 1704–1721. doi:10.1137/090772873

[6] Tatiana Brailovskaya and Ramon van Handel. 2024. Universality and sharp matrix concentration inequalities. *Geom. Funct. Anal.* 34, 6 (2024), 1734–1838. doi:10.1007/s00039-024-00692-9

[7] Joel Friedman. 2008. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.* 195, 910 (2008), viii+100. doi:10.1090/memo/0910

[8] Uffe Haagerup and Steen Thorbjø rnsen. 2005. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)* 162, 2 (2005), 711–775. doi:10.4007/annals.2005.162.711

[9] Arun Jambulapati, Sushant Sachdeva, Aaron Sidford, Kevin Tian, and Yibin Zhao. 2025. Eulerian Graph Sparsification by Effective Resistance Decomposition. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 1607–1650.

[10] Franz Lehner. 1999. Computing norms of free operators with matrix coefficients. *Amer. J. Math.* 121, 3 (1999), 453–486.

[11] Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. 2017. Deterministic Discrepancy Minimization via the Multiplicative Weight Update Method. In *Integer Programming and Combinatorial Optimization (IPCO)*. 380–391.

[12] Shachar Lovett and Raghu Meka. 2015. Constructive discrepancy minimization by walking on the edges. *SIAM J. Comput.* 44, 5 (2015), 1573–1582. doi:10.1137/130929400

[13] James A. Mingo and Roland Speicher. 2017. *Free probability and random matrices*. Fields Institute Monographs, Vol. 35. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON. xiv+336 pages. doi:10.1007/978-1-4939-6942-5

[14] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. 2020. Explicit near-Ramanujan graphs of every degree. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*. 510–523.

[15] Alexandru Nica and Roland Speicher. 2006. *Lectures on the combinatorics of free probability*. London Mathematical Society Lecture Note Series, Vol. 335. Cambridge University Press, Cambridge. xvi+417 pages. doi:10.1017/CBO9780511735127

[16] Thomas Rothvoss. 2017. Constructive discrepancy minimization for convex sets. *SIAM J. Comput.* 46 (2017), 224–234. doi:10.1137/141000282

[17] Daniel A. Spielman and Nikhil Srivastava. 2011. Graph sparsification by effective resistances. *SIAM J. Comput.* 40, 6 (2011), 1913–1926. doi:10.1137/080734029

[18] Joel A. Tropp. 2012. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* 12, 4 (2012), 389–434. doi:10.1007/s10208-011-9099-z

[19] Joel A. Tropp. 2015. An Introduction to Matrix Concentration Inequalities. *Foundations and Trends in Machine Learning* 8, 1-2 (05 2015), 1–230. doi:10.1561/2200000048

[20] Joel A. Tropp. 2018. Second-order matrix concentration inequalities. *Appl. Comput. Harmon. Anal.* 44, 3 (2018), 700–736. doi:10.1016/j.acha.2016.07.005

[21] Ramon van Handel. 2017. Structured random matrices. In *Convexity and concentration*. IMA Vol. Math. Appl., Vol. 161. Springer, New York, 107–156.

[22] Ramon van Handel. 2025. The strong convergence phenomenon. *arXiv preprint* arXiv:2507.00346 (2025).

[23] Dan Voiculescu. 1991. Limit laws for random matrices and free products. *Invent. Math.* 104, 1 (1991), 201–220. doi:10.1007/BF01245072

[24] Robert Wang, Lap Chi Lau, and Hong Zhou. 2026. Derandomizing Matrix Concentration Inequalities from Free Probability. *arXiv preprint* arXiv:2601.08111 (2026).

[25] Avi Wigderson and David Xiao. 2008. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory Comput.* 4 (2008), 53–76. doi:10.4086/toc.2008.v004a003

[26] David P Woodruff. 2014. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science* 10, 1–2 (2014), 1–157.