Matrix Discrepancy for Representations of Finite Groups

Afonso S. Bandeira* Helmut Bölcskei†

Abstract

We prove the group version of the Matrix Spencer conjecture. For every finite group G, there exist signs $\varepsilon \in \{\pm 1\}^G$ such that

$$\left\| \sum_{g \in G} \varepsilon_g \rho(g) \right\| \leq C \sqrt{|G|},$$

where ρ is the left regular representation of G and C is a universal constant. This conjecture was posed in [BKMZ24], which settled it for simple groups; we establish it for all finite groups, combining the Peter–Weyl decomposition with the intrinsic-freeness inequalities of [BBvH23] in an iterated partial-coloring argument.

1 Introduction

The Matrix Spencer conjecture is a far-reaching generalization of Spencer's celebrated "Six standard deviations suffice" theorem [Spe85, Glu89].

Conjecture 1. *[Zou12, Mek14, Ban15] There exists a universal constant C such that for any integer n and any set of self-adjoint[1] $n \times n$ matrices $A_1, \ldots, A_n$ each with spectral norm at most one,*

$$\min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{k=1}^{n} \varepsilon_k A_k \right\| \leq C\sqrt{n}.$$

The Noncommutative Khintchine inequality of Lust-Piquard and Pisier [LPP91, Pis03] shows that a uniformly random draw of ε yields the desired bound, in expectation, up to a logarithmic factor. When the matrices commute, Conjecture 1 reduces to Spencer's Theorem [Spe85], where a uniformly random choice does not suffice and the argument is entropic rather than based on concentration of measure. At the opposite extreme, when the matrices

*`bandeira@math.ethz.ch`, Department of Mathematics, ETH Zürich.

†`hboelcskei@ethz.ch`, Chair for Mathematical Information Science, ETH Zürich.

[1]Note that the self-adjoint requirement is cosmetic, as the Hermitian dilation $A \mapsto \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ reduces the general case to the self-adjoint one.

behave very noncommutatively (more precisely, freely), the intrinsic-freeness matrix concentration inequalities of [BBvH23, Ban25] show that a uniformly random choice of signs suffices. Notable partial progress includes establishing the conjecture for block matrices with small blocks [LRR17] and for low-rank matrices (at first up to rank $\sqrt{n}$ by [HRS22, DJR22] and later up to rank $n/\mathrm{polylog}(n)$ by [BJM23]), as well as for rank-one matrices [KLS20], established via the interlacing-polynomial method of Marcus, Spielman, and Srivastava [MSS15] underlying the resolution of the Kadison–Singer problem. Resolving the conjecture in the general case remains an open problem.

Motivated by the dichotomy of behaviors for the commutative and "very noncommutative" cases, [BKMZ24] proposed a group version of this conjecture in the hope that noncommutativity could be better understood in such an algebraically structured context. In this setting, G is a finite group (with $n = |G|$) and $A_1, \ldots, A_n$ are the matrices of the regular representation of G. The same reference resolved this question for simple groups and left the general finite-group case as a conjecture (see also [BKMR25, Conjecture 2]), which we prove in this paper.[2]

Theorem 2. *There exists a universal constant C such that, for any finite group G,*

$$\min_{\varepsilon \in \{\pm 1\}^G} \left\| \sum_{g \in G} \varepsilon_g \rho(g) \right\| \leq C \sqrt{|G|},$$

where ρ is the left regular representation of G.

In fact, ρ can be any unitary representation of G, since the left regular representation contains every irreducible representation of G, and any unitary representation can be decomposed into irreducibles. Our proof proceeds by simply controlling all irreducible representations simultaneously and does not explicitly use that ρ is the regular representation of G.

The proof of Theorem 2 proceeds via rounds of partial coloring following Gluskin's method. The authors of [BKMZ24] (who overlap with those of the present paper) wrote "The authors were able to adapt Gluskin's volume argument [...] to produce a partial coloring, but we were not able to iterate the procedure." That difficulty is overcome here by exploiting intrinsic freeness [BBvH23] to iterate the procedure, even though the underlying sets will in general no longer be groups after the first iteration.

Outline and Notation: Section 2 contains the crux of the proof, Lemmata 3 and 4, while Section 3 uses these lemmata to prove Theorem 2. We use lowercase c_x to denote universal constants that can be replaced by smaller constants, and C_x for universal constants that can be taken larger. We make no effort to keep track of the values of universal constants.

[2] After this paper was posted, Akbas and Sra [AS26] independently obtained, by a different argument based on multiscale entropy nets rather than intrinsic freeness, a more general algebraic Matrix Spencer theorem for families contained in a finite-dimensional C^*-algebra of dimension $O(n)$, of which the group case is a special case.

2 A Gaussian measure lower bound over subsets

The goal of this section is to prove the following lemma, which is the crux of the argument and the stage at which the group structure is exploited.

Lemma 3. *For every $c_2 > 0$, there exists C_1 (not depending on G) such that, for all non-empty $S \subseteq G$,*

$$\mathbb{P}\left\{\left\|\sum_{g\in S} x_g \rho(g)\right\| \le C_1 \sqrt{|S| \log\left(e\frac{|G|}{|S|}\right)}\right\} \ge \exp(-c_2|S|),$$

where x_g are i.i.d. (real) $\mathcal{N}(0,1)$ random variables.

The idea is to use the Peter–Weyl theorem to decompose ρ into irreducible representations. Before proving Lemma 3, we first establish a matrix concentration result for the spectral norm of the blocks arising from this decomposition.

Lemma 4. *Let λ_π denote an irreducible representation of G of dimension d_π. For any $C_3 > 0$ there exists C_1 (not depending on G or λ_π) such that, for any non-empty subset $S \subseteq G$, we have the following estimate:*

$$\mathbb{P}\left\{\left\|\sum_{g\in S} x_g \lambda_\pi(g)\right\| > C_1 \sqrt{|S| \log\left(e\frac{|G|}{|S|}\right)}\right\} \le \exp\left[-C_3 \log\left(e\frac{|G|}{|S|}\right)\right], \tag{1}$$

where the x_g are i.i.d. (real) $\mathcal{N}(0,1)$ random variables.

Proof of Lemma 4. Consider the Gaussian random matrix $X_{\pi,S} = \sum_{g\in S} x_g \lambda_\pi(g)$. The relevant quantities are the matrix variance $\sigma(X_{\pi,S})$, the weak variance $\sigma_*(X_{\pi,S})$, and the so-called v-parameter $v(X_{\pi,S})$, given by

$$\sigma(X_{\pi,S})^2 := \max\left\{\left\|\mathbb{E}X_{\pi,S}X_{\pi,S}^*\right\|, \left\|\mathbb{E}X_{\pi,S}^*X_{\pi,S}\right\|\right\} = |S|, \tag{2}$$

$$\sigma_*(X_{\pi,S})^2 := \max_{u,v\in\mathbb{S}^{d_\pi-1}} \mathbb{E}|u^*X_{\pi,S}v|^2 \le |S|, \tag{3}$$

$$v(X_{\pi,S})^2 := \|\mathrm{Cov}(X_{\pi,S})\| \le \|\mathrm{Cov}(X_{\pi,G})\| = \frac{|G|}{d_\pi}. \tag{4}$$

A direct computation using $\lambda_\pi(g)\lambda_\pi(g)^* = \lambda_\pi(g)^*\lambda_\pi(g) = I$ establishes (2). Since $\|\lambda_\pi(g)\| = 1$, we have $\sigma_*(X_{\pi,S})^2 \le |S|$, which is (3). It remains to prove (4), which we use in the high-dimensional case below. To this end, note that

$$\max_{\|B\|_F=1} \sum_{g\in G} |\mathrm{Tr}(B^*\lambda_\pi(g))|^2 = \sum_{g\in G}\left|\sum_{i,j=1}^{d_\pi} \overline{B_{ij}}\lambda_\pi(g)_{ij}\right|^2 = \sum_{i,j,k,\ell=1}^{d_\pi} \overline{B_{ij}} B_{k\ell} \sum_{g\in G} \lambda_\pi(g)_{ij}\overline{\lambda_\pi(g)_{k\ell}}.$$

Schur orthogonality yields $\sum_{g\in G}\lambda_\pi(g)_{ij}\overline{\lambda_\pi(g)_{k\ell}} = \delta_{ik}\delta_{j\ell}\frac{|G|}{d_\pi}$, which establishes (4).[3] Lipschitz Gaussian concentration (e.g. [BSS25, Theorem 8.9]) gives

$$\mathbb{P}\left\{\|X_{\pi,S}\| > \mathbb{E}\|X_{\pi,S}\| + t\sigma_*(X_{\pi,S})\right\} \le \exp\left(\frac{-t^2}{2}\right),$$

for any $t \ge 0$. Recall that $\sigma_*(X_{\pi,S}) \le \sqrt{|S|}$. Thus, setting $t = \sqrt{2C_3\log\left(e\frac{|G|}{|S|}\right)}$, we have

$$\mathbb{P}\left\{\|X_{\pi,S}\| > \mathbb{E}\|X_{\pi,S}\| + \sqrt{2C_3\log\left(e\frac{|G|}{|S|}\right)}\sqrt{|S|}\right\} \le \exp\left[-C_3\log\left(e\frac{|G|}{|S|}\right)\right]. \tag{5}$$

It remains only to establish the upper bound $\mathbb{E}\|X_{\pi,S}\| \le C_4\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}$ for some constant C_4. Here, we treat low- and high-dimensional representations differently. For low-dimensional representations, the Noncommutative Khintchine inequality [BSS25, Theorem 9.1] gives

$$\mathbb{E}\|X_{\pi,S}\| \le \sigma(X_{\pi,S})\sqrt{2\lceil\log(d_\pi)\rceil + 1} = \sqrt{(2\lceil\log(d_\pi)\rceil + 1)|S|}, \tag{6}$$

which provides the desired upper bound as long as $d_\pi \le \left(e\frac{|G|}{|S|}\right)^{C_5}$ (and $C_5 = 4$ will suffice).

For high-dimensional representations, we use the intrinsic-freeness inequalities in [BBvH23]. Namely, when $d_\pi > \left(e\frac{|G|}{|S|}\right)^4$, [BBvH23, Theorem 1.2] gives

$$\begin{aligned}
\mathbb{E}\|X_{\pi,S}\| &\le 2\sigma(X_{\pi,S}) + C_6\log(d_\pi)^{\frac{3}{4}}\sqrt{v(X_{\pi,S})\sigma(X_{\pi,S})} \\
&= \sqrt{|S|}\left(2 + C_6\log(d_\pi)^{\frac{3}{4}}\left(\frac{|G|}{d_\pi|S|}\right)^{\frac{1}{4}}\right) \\
&= \sqrt{|S|}\left(2 + C_6\left(\frac{\log(d_\pi)^3}{d_\pi}\frac{|G|}{|S|}\right)^{\frac{1}{4}}\right) \\
&\le \sqrt{|S|}\left(2 + C_6\left(\frac{4^3\log\left(e\frac{|G|}{|S|}\right)^3}{(e|G|/|S|)^4}\frac{|G|}{|S|}\right)^{\frac{1}{4}}\right) \\
&\le C_7\sqrt{|S|},
\end{aligned} \tag{7}$$

for some constant C_7. □

Proof of Lemma 3. Let $C_3 > \log(2)$ be large enough (to be fixed later, depending on c_2) and let C_1 be the constant that is guaranteed to exist by Lemma 4 (for the particular choice of C_3).

[3] There are two other ways of viewing this computation: (i) Schur orthogonality implies that $\{\lambda_\pi(g)\}_{g\in G}$ is a tight frame with respect to the Hilbert–Schmidt inner product; or (ii) we are, in essence, computing the v-parameter for a Wigner matrix of size $d_\pi \times d_\pi$ with entrywise variance $\frac{|G|}{d_\pi}$ (see Lemma 7 in [BKMZ24]).

Let Σ denote the set of isomorphism classes of irreducible representations of G, and let d_π denote the degree of $\pi \in \Sigma$. Pick a representative $\lambda_\pi \in \pi$ per isomorphism class. By the Peter–Weyl theorem (cf. proof of Lemma 7 in [BKMZ24]), there exists a unitary matrix U (depending only on the group) such that

$$\rho(g) = U\left[\bigoplus_{\pi\in\Sigma}(I_{d_\pi} \otimes \lambda_\pi(g))\right]U^*, \qquad g \in G.$$

Thus,

$$\mathbb{P}\left\{\left\|\sum_{g\in S} x_g\rho(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\} = \mathbb{P}\bigcap_{\pi\in\Sigma}\left\{\left\|\sum_{g\in S} x_g\lambda_\pi(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\}.$$

Since this corresponds to the Gaussian measure of the intersection of centrally symmetric convex sets, the Gaussian Correlation Inequality [Roy14, LM17] implies

$$\begin{aligned}
&\mathbb{P}\left\{\left\|\sum_{g\in S} x_g\rho(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\} \\
&\geq \prod_{\pi\in\Sigma}\mathbb{P}\left\{\left\|\sum_{g\in S} x_g\lambda_\pi(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\} \\
&= \prod_{\pi\in\Sigma}(1-p_\pi) = \exp\left(\sum_{\pi\in\Sigma}\log(1-p_\pi)\right), \qquad (8)
\end{aligned}$$

where we define $p_\pi := \mathbb{P}\left\{\left\|\sum_{g\in S} x_g\lambda_\pi(g)\right\| > C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\}$. By Lemma 4 we have $p_\pi \leq \exp\left[-C_3\log\left(e\frac{|G|}{|S|}\right)\right] \leq \exp(-C_3) \leq \frac{1}{2}$. Thus, we have

$$\mathbb{P}\left\{\left\|\sum_{g\in S} x_g\rho(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\} \geq \prod_{\pi\in\Sigma}(1-p_\pi) \geq \exp\left(-2\sum_{\pi\in\Sigma}p_\pi\right), \qquad (9)$$

where the last inequality used $p_\pi \leq \frac{1}{2}$. Note that

$$\frac{2}{|S|}\sum_{\pi\in\Sigma}p_\pi \leq \frac{2}{|S|}|\Sigma|\exp\left[-C_3\log\left(e\frac{|G|}{|S|}\right)\right] \leq 2\frac{|G|}{|S|}\left(e\frac{|G|}{|S|}\right)^{-C_3} = 2e^{-C_3}\left(\frac{|S|}{|G|}\right)^{C_3-1}.$$

For any $c_2 > 0$, there exists C_3 large enough that

$$2e^{-C_3}\left(\frac{|S|}{|G|}\right)^{C_3-1} \leq c_2 \qquad \text{for all } |S| \leq |G|.$$

Picking such C_3 yields

$$\mathbb{P}\left\{\left\|\sum_{g\in S} x_g\rho(g)\right\| \leq C_1\sqrt{|S|\log\left(e\frac{|G|}{|S|}\right)}\right\} \geq \exp\left(-2\sum_{\pi\in\Sigma}p_\pi\right) \geq \exp(-c_2|S|).$$

□

3 Rounds of Partial Coloring

With Lemma 3 established, the rest of the proof relies on the now-standard partial-coloring strategy dating back to the work of Gluskin [Glu89, Gia97, Ban23].

Lemma 5 ([Glu89]; see [Ban23] for this formulation)**.** *Let $n \geq 1$ be an integer. There is a small constant $\delta > 0$ such that any symmetric convex body $K \subseteq \mathbb{R}^n$ with $\mathbb{P}(g \in K) \geq 2^{-\delta n}$, where $g \sim \mathcal{N}(0, I_n)$, contains a point $\gamma \in \{-1, 0, 1\}^n$ with at least δn coordinates equal to ± 1.*

Proof of Theorem 2. We adopt the nomenclature of partial colorings and coloring rounds: the goal is to assign a value ± 1, referred to as a color, to each entry of ε. Let S_t be the set of uncolored coordinates after round t, with $S_0 = G$. At round t we use Lemmata 3 and 5 (with $c_2 = \delta \log 2$) for $S_t \subseteq G$ to obtain a partial coloring of S_t in which at least $\delta |S_t|$ coordinates are colored. Thus, for all $t \geq 0$, $|S_{t+1}| \leq (1-\delta)|S_t|$, and since the $|S_t|$ are all integers, the process must terminate with $S_T = \emptyset$ for some T.

By the triangle inequality, the resulting assignment satisfies

$$\left\| \sum_{g \in G} \varepsilon_g \rho(g) \right\| \leq \sum_{t=0}^{T-1} \left\| \sum_{g \in S_t \setminus S_{t+1}} \varepsilon_g \rho(g) \right\| \leq \sum_{t=0}^{T-1} C_1 \sqrt{|S_t| \log \left(e \frac{|G|}{|S_t|} \right)}. \tag{10}$$

The derivative of $x \mapsto x \log(e/x)$ equals $\log(1/x)$, so $x \log(e/x)$ is increasing on $(0,1)$; hence $\sqrt{|S_t| \log\left(\frac{e|G|}{|S_t|}\right)} \leq \sqrt{(1-\delta)^t |G| \log\left(\frac{e}{(1-\delta)^t}\right)}$ since $|S_t| \leq (1-\delta)^t |G|$. Substituting into (10), we have

$$\left\| \sum_{g \in G} \varepsilon_g \rho(g) \right\| \leq C_1 \sum_{t=0}^{T-1} \sqrt{(1-\delta)^t |G| \log \frac{e}{(1-\delta)^t}} = C_1 \sqrt{|G|} \sum_{t=0}^{T-1} \sqrt{(1-\delta)^t \log \frac{e}{(1-\delta)^t}}.$$

The proof is completed by noting that, since $\delta > 0$, there is a constant C_δ such that $\sum_{t=0}^{T-1} \sqrt{(1-\delta)^t \log \frac{e}{(1-\delta)^t}} \leq \sum_{t=0}^{\infty} \sqrt{(1-\delta)^t \log \frac{e}{(1-\delta)^t}} \leq C_\delta$. □

Acknowledgments and use of AI

We used modern AI tools in this work, primarily ChatGPT Pro 5.5 and Claude Opus (versions 4.7 and 4.8). One of the key arguments—the one that broke the $\sqrt{\log(|G|)}$ bottleneck, reducing it to a $\sqrt{\log \log(|G|)}$ factor that we subsequently removed—was found by ChatGPT Pro 5.5. That argument no longer appears in the paper as it was subsumed by the later improvement, but in our view it was the decisive step in this work. The same AI tools were also used to organize partial results, run numerical experiments, and efficiently test ideas and approaches. For instance, once we understood how low- and high-dimensional representations could be handled simultaneously in Lemma 3, we asked ChatGPT Pro 5.5 to check whether the two regimes together covered the full range or left a gap "in the middle"; in the argument as written, this split appears only at the very end, but in earlier versions it occupied a more central place.

The paper was written entirely by the authors. AI was used to scan the manuscript for typos and inconsistencies in style and notation, and to obtain general suggestions, some of which we adopted. All errors are ours.

References

[AS26] E. Akbas and S. Sra. An algebraic Matrix Spencer theorem. https://arxiv.org/abs/2606.16005, 2026.

[Ban15] A. S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science, 2015. Lecture notes, MIT course 18.S096 (Topics in Mathematics of Data Science), Fall 2015.

[Ban23] N. Bansal. Discrepancy theory and related algorithms. In D. Beliaev and S. Smirnov, editors, *International Congress of Mathematicians 2022 (ICM 2022)*. EMS Press, Berlin, 2023. Sectional lecture, Control Theory and Optimization.

[Ban25] A. S. Bandeira. Random matrices, intrinsic freeness, and sharp non-asymptotic inequalities, 2025. To appear in the Proceedings of the International Congress of Mathematicians 2026.

[BBvH23] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel. Matrix concentration inequalities and free probability. *Inventiones mathematicae*, 234(1):419–487, 2023.

[BJM23] N. Bansal, H. Jiang, and R. Meka. Resolving Matrix Spencer conjecture up to poly-logarithmic rank. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 1814–1819, New York, NY, USA, 2023. Association for Computing Machinery.

[BKMR25] A. S. Bandeira, A. Kireeva, A. Maillard, and A. Rödder. Randomstrasse 101: Open problems of 2024, 2025. Includes Entry 1, "Welcome and Matrix Discrepancy," and Conjecture 2 (Group Spencer).

[BKMZ24] A. S. Bandeira, D. Kunisky, D. G. Mixon, and X. Zeng. On the concentration of Gaussian Cayley matrices. *Applied and Computational Harmonic Analysis*, 73:101694, 2024.

[BSS25] A. S. Bandeira, A. Singer, and T. Strohmer. Topics in mathematics of data science. Preprint, version 1.0-, September 2025.

[DJR22] D. Dadush, H. Jiang, and V. Reis. A new framework for matrix discrepancy: Partial coloring bounds via mirror descent. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2022.

[Gia97] A. A. Giannopoulos. On some vector balancing problems. *Studia Mathematica*, 122(3):225–234, 1997.

[Glu89] E. D. Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85–96, 1989. Russian original: Mat. Sb. (N.S.) **136(178)** (1988), no. 1, 85–96.

[HRS22] S. B. Hopkins, P. Raghavendra, and A. Shetty. Matrix discrepancy from quantum communication. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2022.

[KLS20] R. Kyng, K. Luh, and Z. Song. Four deviations suffice for rank 1 matrices. *Advances in Mathematics*, 375:107366, 2020.

[LM17] R. Latała and D. Matlak. Royen's proof of the Gaussian correlation inequality. In B. Klartag and E. Milman, editors, *Geometric Aspects of Functional Analysis*, volume 2169 of *Lecture Notes in Mathematics*, pages 265–275. Springer, 2017.

[LPP91] F. Lust-Piquard and G. Pisier. Non commutative Khintchine and Paley inequalities. *Ark. Mat.*, 29:241–260, 1991.

[LRR17] A. Levy, H. Ramadas, and T. Rothvoss. Deterministic discrepancy minimization via the multiplicative weight update method. In *Integer Programming and Combinatorial Optimization (IPCO)*, pages 380–391, 2017.

[Mek14] R. Meka. Discrepancy and beating the union bound. `https://windowsontheory.org/2014/02/07/discrepancy-and-beating-the-union-bound/`, 2014.

[MSS15] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem. *Annals of Mathematics*, 182(1):327–350, 2015.

[Pis03] G. Pisier. *Introduction to Operator Space Theory*. Cambridge University Press, 2003.

[Roy14] T. Royen. A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East Journal of Theoretical Statistics*, 48(2):139–145, 2014.

[Spe85] J. Spencer. Six standard deviations suffice. *Trans. Amer. Math. Soc.*, 289:679–706, 1985.

[Zou12] A. Zouzias. A matrix hyperbolic cosine algorithm and applications. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 846–858. Springer, 2012.