Resolving Matrix Spencer Conjecture Up to Poly-logarithmic Rank

Nikhil Bansal*   Haotian Jiang†   Raghu Meka‡

Abstract

We give a simple proof of the matrix Spencer conjecture up to poly-logarithmic rank: given symmetric $d \times d$ matrices $A_1, \ldots, A_n$ each with $\|A_i\|_{\mathsf{op}} \leq 1$ and rank at most $n/\log^3 n$, one can efficiently find ± 1 signs $x_1, \ldots, x_n$ such that their signed sum has spectral norm $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O(\sqrt{n})$. This result also implies a $\log n - \Omega(\log \log n)$ qubit lower bound for quantum random access codes encoding n classical bits with advantage $\gg 1/\sqrt{n}$.

Our proof uses the recent refinement of the non-commutative Khintchine inequality in [Bandeira, Boedihardjo, van Handel, 2022] for random matrices with correlated Gaussian entries.

---

*University of Michigan, Ann Arbor. `bansal@gmail.com`.

†University of Washington, Seattle. `jhtdavid@cs.washington.edu`

‡University of California, Los Angeles. `raghum@cs.ucla.edu`

1   Introduction

We study discrepancy minimization in the matrix setting. Let us start with the classical discrepancy setting where given vectors $a_1, \ldots, a_n \in \mathbb{R}^d$ satisfying $\|a_i\|_\infty \leq 1$ for all $i \in [n]$, and the goal is to find signs $x_1, \ldots, x_n \in \{\pm 1\}$ to minimize the discrepancy $\|\sum_{i=1}^n x_i a_i\|_\infty$. In a seminal result, Spencer [Spe85] showed that the $O(\sqrt{n \log d})$ bound obtained by a random coloring is not tight and showed the following bound, which is also the best possible in general.

Theorem 1.1 (Spencer [Spe85])**.** *Given vectors* $a_1, \ldots, a_n \in \mathbb{R}^d$ *each satisfying* $\|a_i\|_\infty \leq 1$, *there exist signs* $x \in \{\pm 1\}^n$ *such that* $\|\sum_{i=1}^n x_i v_i\|_\infty = O(\sqrt{n} \cdot \max\{1, \sqrt{\log(d/n)}\})$.

In particular for $d = O(n)$, this gives an $O(\sqrt{n})$ bound, in contrast to the $O(\sqrt{n \log n})$ bound for random coloring obtained by applying Chernoff and union bounds.

To prove this result, Spencer developed the powerful partial-coloring method via the entropy method, building on the previous work of Beck [Bec81]. Another approach to prove Theorem 1.1 based on convex geometry was developed independently by Gluskin [Glu89]. While these original arguments used the pigeonhole principle and were non-algorithmic, in recent years, there has been a rich line of work [Ban10, BS13, LM15, Rot17, LRR17, ES18, RR20b] on their algorithmic versions.

Matrix Spencer Setting.   A natural generalization of Spencer's problem to matrices is the following. Let $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ be symmetric matrices with maximum singular value, or operator norm, $\|A_i\|_{\mathsf{op}} \leq 1$. Find a coloring $x \in \{\pm 1\}^n$ that minimizes $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}}$. In particular, Spencer's result corresponds to the case when all the $A_i = \mathsf{diag}(a_i)$ are diagonal.

As in the vector case, for a random coloring $x \in \{\pm 1\}^n$, the non-commutative Khintchine inequality of Lust-Piquard and Pisier [LP86, LPP91, Pis03], or the matrix Chernoff bound [Oli10, Tro15], give that

$$\mathbb{E}\Big[\big\|\sum_i x_i A_i\big\|_{\mathsf{op}}\Big] = O\Big(\sqrt{\log d} \cdot \big\|\sum_i A_i^2\big\|_{\mathsf{op}}^{1/2}\Big). \tag{1}$$

This implies a bound of $O(\sqrt{n \log d})$ on the matrix discrepancy. This inequality also holds when one picks $x \in \mathbb{R}^n$ to be standard Gaussians, which will play an important role in our results.

Matrix concentration bounds are powerful and widely used tools in mathematics and computer science, and it is natural to ask when one can beat them. In particular, whether the following natural analog of Spencer's result for matrices holds is a tantalizing open question.

Conjecture 1.2 (Matrix Spencer Conjecture [Zou12, Mek14])**.** *Given* $d \times d$ *symmetric matrices* $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ *with* $\|A_i\|_{\mathsf{op}} \leq 1$, *there exist signs* $x \in \{\pm 1\}^n$ *such that* $\big\|\sum_{i=1}^n x_i A_i\big\|_{\mathsf{op}} \leq O(\sqrt{n} \cdot \max\{1, \sqrt{\log(d/n)}\})$. *In particular, the matrix discrepancy is* $O(\sqrt{n})$ *for* $d = n$.

While this conjecture is still open, there has been exciting progress on important special cases. Recently, Hopkins, Raghavendra and Shetty [HRS22] proved Conjecture 1.2 where each matrix A_i has rank $O(\sqrt{n})$; in a different direction, Levy, Ramadas and Rothvoss [LRR17] and Dadush, Jiang and Reis [DJR22] established Conjecture 1.2 for block-diagonal matrices with block size $h = O(n/d)$. Recently, Bansal, Jiang, and Meka [BJM22] gave an approach based on *barrier functions* to achieve a bound that unifies and slightly strengthens the results of [HRS22, DJR22].

1.1 Our Results

The main result of this paper is the following theorem.

Theorem 1.3 (Matrix Spencer Up to Poly-logarithmic Rank). *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ each with $\|A_i\|_{\mathsf{op}} \leq 1$ and $\|A_i\|_F^2 \leq n/\log^3 n$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i A_i\|_{\mathsf{op}} = O(\sqrt{n})$. Moreover, these signs can be computed efficiently.*

Note that the condition $\|A_i\|_F^2 \leq n/\log^3 n$ is satisfied when each A_i has rank at most $n/\log^3 n$ or in particular if $d \leq n/\log^3 n$. Thus Theorem 1.3 resolves Conjecture 1.2 up to poly-logarithmic dimension or poly-logarithmic rank. We remark that even when assuming the matrices A_i have small rank (even rank 1) or small dimension (even $d = \sqrt{n}$), it is known that one cannot hope for a bound better than $\Theta(\sqrt{n})$ [DJR22]. For instance, let $e_1, \ldots, e_n \in \mathbb{R}^n$ be the standard basis vectors and take $A_i = (1/2)(e_1 + e_i)(e_1 + e_i)^T$. Then, for any $x \in \{\pm 1\}^n$, the first column of $\sum_i x_i A_i$ has norm $\Omega(\sqrt{n})$ so its spectral norm is $\Omega(\sqrt{n})$. This is in sharp contrast to the diagonal case, where an $O(\sqrt{r \log n})$ bound for rank r matrices holds [Ban98], and a $O(\sqrt{r})$ bound was conjectured [BF81].

Further, when matrices A_i have dimension $d = \omega(n)$ but $\mathsf{rank}(A_i) \leq n/\log^3 n$, the bound in Theorem 1.3 is $O(\sqrt{n})$ and is stronger than the $\omega(\sqrt{n})$ bound suggested by Conjecture 1.2.

A new ingredient in our proof is a recent strengthening of the non-commutative Khintchine inequality for Gaussian random matrices of the form $\sum_i g_i A_i$ where $g_1, \ldots, g_n$ are independent standard Gaussian random variables due to Bandeira, Boedihardjo, and van Handel [BBvH21]. The central idea is to pick a suitable projection of the random matrix to a subspace so that the bound of [BBvH21] matches $O(\sqrt{n})$. We defer the details to the full proof.

The same proof strategy also implies an improvement over the random coloring bound of $O(\sqrt{n \log d})$ for all $d = o(n)$ by using the result of [Tro18] together with [BBvH21][1].

Implications for Quantum Random Access Codes. [HRS22] identified a beautiful connection between the matrix Spencer conjecture and quantum random access codes that achieve advantage $C/\sqrt{n}$ for a big enough constant. They use this connection in their proof of the conjecture for matrices of rank $O(\sqrt{n})$.

Consider the following two-party communication problem: Alice is given a vector $x \in \{\pm 1\}^n$ and Bob an index $i \in [n]$. We are interested in the one-way quantum communication complexity (from Alice to Bob) of computing x_i. That is, Alice gets to send a quantum message to Bob and Bob must use this message to compute a guess for x_i. For a protocol Π, let $\mathsf{adv}_\Pi(x, i) = \max(0, \mathbb{P}[\Pi(x, i) = x_i] - 1/2)$ be the advantage over random guessing that Alice and Bob have. Note that the randomness is over that of the protocol.

The seminal works of [ANTSV02] showed that for any protocol Π with $\mathbb{E}_{x,i}[\mathsf{adv}_\Pi(x, i)] = \Omega(1)$, Alice must send $\Omega(n)$ qubits to Bob[2]. In [HRS22], the following elegant connection between Conjecture 1.2 and the above communication problem is made: The conjecture is true if and only if there is some constant C such that any protocol Π with $\min_x \mathbb{E}_i[\mathsf{adv}_\Pi(x, i)] > C/\sqrt{n}$ must send at least $\log_2 n - O(1)$ qubits from Alice to Bob.

---

[1]Specifically, using the bound $\mathbb{E}[\|X\|_{\mathsf{op}}] = O((\log d)^{1/4}\sigma(X) + (\log d)^{1/2}\sqrt{v(X)\sigma(X)})$ (this follows by combining Corollary 3.6 in [Tro18] with Proposition 4.6 in [BBvH21]), instead of the bound given by Theorem 2.1, in the argument in Section 3 gives the discrepancy bound $o(\sqrt{n \log d})$ for all $d = o(n)$.

[2]On a related note, if one is not interested in the exact constant, one can obtain an $\Omega(n)$ bound easily from the matrix Chernoff bound in (1) (without using any quantum information theory).

As our main result, Theorem 1.3, proves the conjecture for matrices of dimension $n/\log^3 n$, this combined with Claim 1.6 in [HRS22] immediately imply the following corollary:

Corollary 1.4 (QRAC Lower Bound). *There exists a universal constant $C > 0$ such that the following holds. Any quantum one-way protocol Π as above with $\min_x \mathbb{E}_i[\mathsf{adv}_\Pi(x,i)] > C/\sqrt{n}$ requires at least $\log_2 n - 3 \log_2 \log_2 n - O(1)$ qubits of communication from Alice to Bob.*

Note that the leading constant of 1 in front of $\log_2 n$ is right for the first time and is the best possible (for sufficiently large constant $C > 0$). Previously, the results of [HRS22, DJR22] imply a lower bound of $(1/2) \log_2 n - O(1)$ on the quantum one-way communication complexity.

Further, a modification of the example in [DJR22] shows that there exists a protocol Π such that for all $x \in \{\pm 1\}^n, i \in [n]$, $\mathsf{adv}_\Pi(x,i) > c/\sqrt{n}$ for some constant $c > 0$ and involves at most $(1/2) \log_2 n + O(1)$ qubits of communication. Combined with our lower bound, this shows a somewhat sharp transition in the communication required for protocols as in Corollary 1.4: for some constants $0 < c < C$, achieving an advantage of $C/\sqrt{n}$ requires $\log_2 n - O(\log \log n)$ qubits, whereas one can achieve $c/\sqrt{n}$ advantage with $(1/2) \log_2 n + O(1)$ qubits. Interestingly, the transition is a quantum phenomenon and is absent for classical randomized communication; a tight bound of $\log_2 n + \Theta(\alpha^2)$ bits of communication is known for achieving advantage $\alpha/\sqrt{n}$ for all $\alpha > 0$.

1.2 Further Related Works

Discrepancy Theory. Discrepancy theory is widely studied and has applications to many other mathematics and computer science areas. We refer readers to the excellent books [Cha00, Mat09, CST+14] for a more comprehensive account of the rich history of discrepancy. Recent developments in discrepancy have led to several applications in approximation algorithms, differential privacy, fair allocation, experimental design, and more [MN12, Rot13, NTZ13, BCKL14, Ban19, JKS19, HSSZ19, BJSS20, BRS22].

Matrix Discrepancy and Non-Commutativity Random Matrix Theory. Many natural problems in the study of spectra of matrices can be viewed as questions about matrix discrepancy, e.g., graph sparsification [BSS12, RR20a], the Kadison-Singer problem [MSS15] and its generalization [KLS20], and the design of quantum random access codes [ANTSV02, HRS22].

Matrix discrepancy is also closely related to non-commutative random matrix theory, where the typical value of $\|\sum_i x_i A_i\|_{\mathsf{op}}$ for a random coloring x has received significant attention. The bound of $\mathbb{E}[\|\sum_i x_i A_i\|_{\mathsf{op}}] \leq O(\sqrt{n \log m})$ by matrix Chernoff [AW02] or matrix Khintchine [LP86, LPP91, Pis03] that is generally tight for commutative matrices, can be often improved in the non-commutative case (e.g. [Ver18, Tro18, BBvH21] and the references therein). We refer readers to the book [Tao12, Vu14] for a more comprehensive account of random matrix theory.

2 Preliminaries

We first recall some basic facts about matrices and describe the notations. For a square matrix $A \in \mathbb{R}^{m \times m}$ with entries a_{ij}, its trace $\mathsf{Tr}(A) = \sum_i a_{ii}$ and Frobenius norm $\|A\|_F = \sqrt{\mathsf{Tr}(A^T A)} = (\sum_{ij} a_{ij}^2)^{1/2}$. If A is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, then we have $\mathsf{Tr}(A) = \sum_i \lambda_i$, $\|A\|_F =$

$(\sum_i \lambda_i^2)^{1/2}$ and its operator norm $\|A\|_{\mathsf{op}} = \max_{\|x\|_2=1} \|Ax\|_2 = \max_i |\lambda_i|$. A symmetric matrix A is positive semidefinite (PSD) if all its eigenvalues $\lambda_i \geq 0$.

For a linear subspace $H \subseteq \mathbb{R}^n$, let $H^\perp$ denote its orthogonal complement. For any matrix $A \in \mathbb{R}^{d \times d}$, we let $\mathsf{vec}(A) \in \mathbb{R}^{d^2}$ be the vector formed by the d^2 entries of A in a fixed order. For a subspace H and convex set $K \subseteq H$, denote $\gamma_H(K)$ the Gaussian measure of K restricted to H, i.e. the probability that a standard Gaussian vector on H lies in K.

2.1 Matrix Concentration

Let $X \in \mathbb{R}^{d \times d}$ be a symmetric multi-variate Gaussian random matrix (i.e., the entries of X are jointly Gaussian). Equivalently, we can assume that X is of the form $X = \sum_{i=1}^n g_i A_i$ where g_i are independent standard Gaussians and $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ are symmetric matrices[3]. Note that this representation of X is not unique, and by the rotational invariance of the Gaussians, one also has $X = \sum_j g_j B_j$ where $B_j = \sum_i (v^j)_i A_i$ for any $n \times n$ orthogonal matrix with columns v^j.

Let $\sigma(X)^2 = \|\mathbb{E}[X^2]\|_{\mathsf{op}} = \|\sum_i A_i^2\|_{\mathsf{op}}$. The fundamental matrix-Chernoff inequality or non-commutative Khintchine inequality implies, among other things, that for X as above, we have

$$\mathbb{E}[\|X\|_{\mathsf{op}}] = O(\sigma(X) \cdot \sqrt{\log d}).$$

Note that this bound is tight in general, for instance, if X is a suitable diagonal matrix. Much attention has been given to finding special cases where the $\sqrt{\log d}$ factor in the bound above can be improved. Of particular note is the work of Tropp [Tro18] where he introduced a specific matrix alignment parameter to capture the non-commutativity of the matrices A_i.

Recently, Bandeira, Boedihardjo, and van Handel made substantial progress in this direction in [BBvH21]. In particular, they related the matrix alignment parameter of Tropp to the following more natural parameter. Let

$$\mathsf{Cov}(X) := \mathbb{E}[\mathsf{vec}(X)\mathsf{vec}(X)^\top] = \mathbb{E}\Big[\sum_{i=1}^n \mathsf{vec}(A_i)\mathsf{vec}(A_i)^\top\Big] \tag{2}$$

be the $d^2 \times d^2$ covariance matrix of its d^2 scalar entries and define

$$v(x)^2 := \|\mathsf{Cov}(X)\|_{\mathsf{op}}.$$

Bandeira, Boedihardjo, and van Handel [BBvH21] showed the following refinement of the non-commutative Khintchine inequality of Lust-Piquard and Pisier [LP86, LPP91, Pis03].

Theorem 2.1 ([BBvH21], Theorem 1.2)**.** *Given symmetric matrices $A_1, \cdots, A_n \in \mathbb{R}^{d \times d}$, let $X = \sum_{i=1}^n g_i A_i$ where g_i are i.i.d. standard Gaussians. Then*

$$\mathbb{E}[\|X\|_{\mathsf{op}}] \leq C \cdot (\sigma(X) + (\log^{3/4} d)\sigma(X)^{1/2}v(X)^{1/2}),$$

where C is some universal constant. In particular, $\mathbb{E}[\|X\|_{\mathsf{op}}] = O(\sigma(X) + (\log^{3/2} d)v(X))$.

We remark that the bound in [BBvH21] is substantially more potent and, in particular, gives the optimum constant for the $\sigma(X)$ term and even control over the full spectrum of X. However, the weaker version above suffices for our purposes.

---

[3]It will be useful to think of the matrix X by itself as a random matrix, and only use the specific representation $\sum_i g_i A_i$ when needed.

2.2 Partial Colorings in Convex Sets

The seminal work of Gluskin [Glu89] introduced the idea of finding partial colorings via techniques from convex geometry. At the core is the idea that any symmetric convex set $K \subseteq \mathbb{R}^n$ with sufficiently large Gaussian volume must contain a vector from $\{-1, 0, 1\}^n$ with $\Omega(n)$ nonzero coordinates (i.e., a *good partial coloring*). In particular, Giannopoulos [Gia97] showed that if $\gamma(K) \geq e^{-\delta n}$ for a sufficiently small constant δ, then K must contain a good partial coloring. Rothvoss [Rot13] gave an algorithmic version of Giannopoulos's result and extended it to subspaces with dimension close to n. This extension will be useful for our purposes.

Lemma 2.2 ([Rot17], Lemma 9). *Let $\varepsilon \leq 1/60000$ and $\delta := \frac{3}{2}\varepsilon \log_2(1/\varepsilon)$. Given a subspace $H \subseteq \mathbb{R}^n$ of dimension at least $(1 - \delta)n$, a symmetric convex set $K \subseteq H$ with $\gamma_H(K) \geq e^{-\delta n}$ and a point $x_0 \in (-1, 1)^n$. There exists a polynomial time algorithm to find a point $x \in (x_0 + K) \cap [-1, 1]^n$ so that $|\{i : x_i \in \{\pm 1\}\}| \geq \varepsilon n/2$.*

3 Proof of the Main Result

Following the standard approach, it suffices to find a partial coloring with $O(n^{1/2})$ discrepancy. We show this below in Section 3.1, and then show how Theorem 1.3 follows from it in Section 3.2.

3.1 Main Partial Coloring Lemma

Lemma 3.1 (Main Partial Coloring Lemma). *There exist constants $c, c' > 0$ such that the following holds. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ that satisfy $\|\sum_{i=1}^n A_i^2\|_{\mathsf{op}} \leq \sigma^2$ and $\sum_{i=1}^n \|A_i\|_F^2 \leq nf^2$ and a point $x_0 \in (-1, 1)^n$, there exists a point $x \in [-1, 1]^n$ such that*

$$\Big\| \sum_{i=1}^n (x_i - x_{0,i}) A_i \Big\|_{\mathsf{op}} \leq c(\sigma + (\log^{3/4} d)\sqrt{\sigma f}\,),$$

and $|\{i : x_i \in \{\pm 1\}\}| > c'n$. Moreover, such a point can be found in polynomial time.

The partial coloring upper bound could be changed to the clearer bound of $O(\sigma + (\log d)^{3/2} f)$ without too much loss; but the above is better for our recursion. In particular, note that if $\sigma \leq \sqrt{n}$ and $f^2 \leq n/\log^3 d$ (which will be true when $\|A_i\|_{\mathsf{op}} \leq 1$, and $\mathsf{rank}(A_i) \leq n/\log^3 d$), we get a partial coloring with a spectral norm bound of $O(\sqrt{n})$.

The idea behind the proof is as follows. Let $X = \sum_{i=1}^n g_i A_i$ where g_i are i.i.d standard Gaussian random variables. Consider the convex body

$$K = \Big\{ x \in \mathbb{R}^n : \Big\| \sum_{i=1}^n x_i A_i \Big\|_{\mathsf{op}} \leq c\sigma \Big\} \subseteq \mathbb{R}^n$$

for some suitably large constant $c > 0$. If K had Gaussian measure $\gamma(K) \geq \exp(-\Omega(n))$, then we could directly use Rothvoss's partial coloring result [Rot17]. As $\sigma(X) \leq \sigma$, one may hope that the improved concentration bound in Theorem 2.1 can be used to show such a lower bound on $\gamma(K)$. However, it is unclear how to do this directly, as we do not have any control on $v(X)$ and it might even be larger than $\sigma(X)$. So our key idea is to work with a suitable slice of the body K.

5

A key observation is that even if $v(X)$ itself is large, the number of large eigenvalues of $\mathsf{Cov}(X)$ must be small as $\mathsf{Tr}(\mathsf{Cov}(X)) = \sum_{i=1}^{n} \|A_i\|_F^2 \leq nf^2$. In particular, if we set $\Delta^2 \geq f^2/\delta$, then the number of *bad* eigenvectors of $\mathsf{Cov}(X)$ with eigenvalue greater than Δ^2 is at most δn. The main idea is to restrict the g_i's to lie in a subspace $H \subseteq \mathbb{R}^n$ so that if $y \in H$ is drawn from the standard Gaussian distribution on H, the resulting matrix $Y = \sum_i y_i A_i$ is perpendicular to each of the *bad* eigenvectors of $\mathsf{Cov}(X)$. This ensures that $v(Y) \leq \Delta$ and by Theorem 2.1, $\mathbb{E}[\|Y\|_{\mathsf{op}}] = O(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f})$. Further, as the number of such bad eigenvectors of $\mathsf{Cov}(X)$ is small, we can ensure that H has dimension at least $(1-\delta)n$. We can now apply Lemma 2.2 to get the desired partial coloring. We now give the details.

Proof of Lemma 3.1. Let constants $\varepsilon := 1/60000$ and $\delta := \frac{3}{2}\varepsilon \log_2(1/\varepsilon)$ be as in Lemma 2.2. We define $X = \sum_{i=1}^{n} g_i A_i$ where g_i are i.i.d. standard Gaussian random variables. Consider the PSD matrix $\mathsf{Cov}(X) \in \mathbb{R}^{d^2 \times d^2}$ defined in (2). Note that by assumption,

$$\mathsf{Tr}(\mathsf{Cov}(X)) = \sum_{i=1}^{n} \|\mathsf{vec}(A_i)\|_2^2 = \sum_{i=1}^{n} \|A_i\|_F^2 \leq \delta n \Delta^2,$$

where $\Delta^2 := f^2/\delta$. This implies that there can be at most $k := \delta n$ eigenvalues of $\mathsf{Cov}(X)$ exceeding Δ^2. Let $V_1, \cdots, V_k \in \mathbb{R}^{d \times d}$ be such that $\mathsf{vec}(V_j)$ is the eigenvector for the jth largest eigenvalue of $\mathsf{Cov}(X)$. Define the subspace

$$H := \Big\{ y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i \cdot \mathsf{Tr}(A_i V_j) = 0, \forall j \in [k] \Big\}.$$

Now we sample the standard Gaussian vector $g \in \mathbb{R}^n$ as follows: first sample a standard Gaussian vector $y \in H$, then sample an independent standard Gaussian vector $r \in H^\perp$, and finally let $g = y + r$. We define $Y := \sum_{i=1}^{n} y_i A_i$ and $R := \sum_{i=1}^{n} r_i A_i$, which implies $X = Y + R$. Since Y and R are independent and have zero mean, we immediately have that $\mathbb{E}[X^2] = \mathbb{E}[Y^2] + \mathbb{E}[R^2]$ and therefore $\sigma(Y) \leq \sigma(X) \leq \sigma$ by the assumption in the Lemma.

We next show that $v(Y) \leq \Delta$. As $\mathsf{Tr}(YV_j) = 0$ for any $j \in [k]$, we have $\mathsf{vec}(V_j)^\top \mathsf{Cov}(Y)\mathsf{vec}(V_j) = 0$. As $\mathsf{Cov}(Y)$ is a PSD matrix, $W := \mathsf{span}\{\mathsf{vec}(V_1), \cdots, \mathsf{vec}(V_k)\} \subseteq \mathbb{R}^{d^2}$ must be a subspace of the eigenspace corresponding to the eigenvalue 0 of the matrix $\mathsf{Cov}(Y)$. For any vector $v \in \mathbb{R}^{d^2}$ with $v \perp W$, we thus have that

$$v^\top \mathsf{Cov}(Y)v \leq v^\top \mathsf{Cov}(X)v \leq \Delta^2,$$

as $\mathsf{Cov}(X) = \mathsf{Cov}(Y) + \mathsf{Cov}(R)$, and the $(k+1)$th eigenvalue of $\mathsf{Cov}(X)$ is at most Δ^2. This proves that $\|\mathsf{Cov}(Y)\|_{\mathsf{op}} \leq \Delta^2$, or equivalently $v(Y) \leq \Delta$.

Now we want to apply Theorem 2.1 to $Y = \sum_{i=1}^{n} y_i A_i$. A crucial but elementary fact is that Theorem 2.1 holds for any (symmetric) matrix-valued random variable whose entries are jointly Gaussian and the final bound only depends on the overall distribution of the random matrix and not on the specific representation as a sum of independent matrices. Clearly, the matrix Y we have is a multi-variate Gaussian random variable so we can apply their result.

To be precise, we can justify its validity even though the vector y does not have independent coordinates as follows. Let $v^1, \ldots, v^k \in \mathbb{R}^n$ be an orthonormal basis for H. Then, we can write

6

$y = \sum_{j=1}^{k} h_j v^j$ where h_i are i.i.d standard Gaussian variables. We can now write

$$Y = \sum_{i=1}^{n} y_i A_i = \sum_{j=1}^{k} h_j \Big(\sum_{i=1}^{n} (v^j)_i A_i \Big) = \sum_{j=1}^{k} h_j B_j,$$

where we define $B_j = \sum_{i=1}^{n} (v^j)_i A_i$. Thus Y can be written in the form of a Gaussian matrix series in terms of the i.i.d. standard Gaussians h_i.

Thus we can apply Theorem 2.1 to Y and obtain for universal constant $C > 0$,

$$\mathbb{E}[\|Y\|_{\mathsf{op}}] \leq C \cdot (\sigma(Y) + (\log^{3/4} d) \cdot \sqrt{\sigma(Y) v(Y)}) \leq c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}), \tag{3}$$

for a sufficiently big constant $c > C(1 + 1/\sqrt{\delta})$. Let us consider the convex body

$$K' := \Big\{ x \in H : \Big\| \sum_{i=1}^{n} x_i A_i \Big\|_{\mathsf{op}} \leq 2c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}) \Big\}.$$

By Markov's inequality and (3), it follows that $\gamma_H(K') \geq 1/2 \geq e^{-\delta n}$. Also note that $\dim(H) \geq (1-\delta)n$ since H is defined by δn constraints. It then follows from Lemma 2.2 that we can efficiently find a point $x \in (x_0 + K') \cap [-1,1]^n$ such that $|\{i : |x_i| = 1\}| \geq \varepsilon n/2 = \Omega(n)$. By the definition of K, the guarantee that $x \in x_0 + K'$ translates to

$$\Big\| \sum_{i=1}^{n} (x_i - x_{0,i}) A_i \Big\|_{\mathsf{op}} \leq 2c(\sigma + (\log^{3/4} d)\sqrt{\sigma \cdot f}).$$

This completes the proof of the lemma. As $(\log^{3/4} d)\sqrt{\sigma \cdot f} \leq (\sigma + f(\log d)^{3/2})/2$, this implies a partial coloring discrepancy bound of at most $O(\sigma + (\log d)^{3/2} f)$. $\qquad\square$

3.2 Proof of Main Theorem

We can now prove Theorem 1.3 (restated below) by recursively applying Lemma 3.1.

Theorem 1.3 (Matrix Spencer Up to Poly-logarithmic Rank). *Given $d \times d$ symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ each with $\|A_i\|_{\mathsf{op}} \leq 1$ and $\|A_i\|_F^2 \leq n/\log^3 n$, there exist signs $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^{n} x_i A_i\|_{\mathsf{op}} = O(\sqrt{n})$. Moreover, these signs can be computed efficiently.*

Proof of Theorem 1.3. Denote $f^2 := n/\log^3 n$. First, without loss of generality, we can assume that $d \leq n^2$. Indeed, suppose to the contrary that $d > n^2$. Define $M := \sum_{i=1}^{n} A_i^2$ and note that $\mathsf{Tr}(M) = \sum_{i=1}^{n} \|A_i\|_F^2 \leq nf^2$. By a change of basis, we may assume without loss of generality that M is diagonal and its diagonal entries are in descending order. Note that $M_{n^2, n^2} \leq \mathsf{Tr}(M)/n^2 \leq f^2/n$. Define $B_i \in \mathbb{R}^{(d-n^2) \times d}$ the matrix obtained by removing the first n^2 rows of A_i. We have for any coloring $x \in \{\pm 1\}^n$,

$$\Big\| \sum_{i=1}^{n} x_i B_i \Big\|_{\mathsf{op}}^2 = \Big\| \Big(\sum_{i=1}^{n} x_i B_i \Big)^\top \Big(\sum_{i=1}^{n} x_i B_i \Big) \Big\|_{\mathsf{op}} \leq n \cdot \Big\| \sum_{i=1}^{n} B_i^\top B_i \Big\|_{\mathsf{op}} \leq f^2,$$

7

where the inequality follows as $x_i x_j (B_i^T B_j + B_j^T B_i) \preceq (B_i^T B_i + B_j^T B_j)$ for all i, j. Now we let $L_i \in \mathbb{R}^{d \times d}$ be the matrix obtained by zeroing out the top left $n^2 \times n^2$ block of A_i. Since matrices A_i are symmetric, it follows that for any coloring $x \in \{\pm 1\}^n$,

$$\Big\| \sum_{i=1}^n x_i L_i \Big\|_{\mathsf{op}} \leq 2 \Big\| \sum_{i=1}^n x_i B_i \Big\|_{\mathsf{op}} \leq 2f.$$

This shows that we only need to keep the top left $n^2 \times n^2$ block of each matrix A_i without affecting the discrepancy by more than an additive term of $2f$. We thus assume henceforth that $d \leq n^2$.

By assumption, the matrices A_i satisfy $\| \sum_{i=1}^n A_i^2 \|_{\mathsf{op}} \leq n$ and $\sum_{i=1}^n \|A_i\|_F^2 \leq nf^2$. Therefore, we can apply Lemma 3.1 with $x_0 = 0$ to obtain a partial coloring $x^{(1)} \in [-1,1]^n$ with $\| \sum_{i=1}^n x_i^{(1)} A_i \|_{\mathsf{op}} = O(\sqrt{n})$ and $|\{i : |x_i^{(1)}| = 1\}| = \Omega(n)$. Next we let $I_1 := \{i \in [n] : |x_i^{(1)}| < 1\}$, and recursively apply Lemma 3.1 to the set of matrices $\{A_i\}_{i \in I_1}$ with point $x^{(1)}|_{I_1}$. Continuing this process of recursively applying Lemma 3.1 to the set of coordinates i such that $|x_i| < 1$, the number of such coordinates decreases by a constant factor in each iteration.

Let $x^{(t)} \in [-1,1]^n$ be the resulting vector in the tth iteration and let n_t denote the number of coordinates in $x^{(t)}$ that are in $(-1,1)$. Then, we have $n_{t+1} < \lambda n_t$ for some constant $\lambda < 1$ and by using Lemma 3.1 with $\sigma \leq \sqrt{n_t}$, we get that the discrepancy increases additively by at most $c(\sqrt{n_t} + (\log^{3/4} d) \cdot f^{1/2} n_t^{1/4})$. Therefore, repeating it for $O(\log n)$ iterations, we get a full coloring with discrepancy at most

$$c \sum_t (\sqrt{n_t} + (\log^{3/4} d) \cdot f^{1/2} n_t^{1/4}) = O(\sqrt{n}) + O((\log^{3/4} n) \cdot f^{1/2} n^{1/4}),$$

where we have used that $d \leq n^2$ and n_t's form a geometrically decreasing series. The theorem now follows since we have chosen $f^2 = n/\log^3 n$. $\qquad \square$

Remark 3.2. *One can also use the version of Lemma 2.2 without defining the subspace as in the proof above, but this requires assuming $\|A_i\|_F^2 \leq n/\log^4 d$ in Theorem 1.3. In particular, by taking $\mathsf{vec}(A_i')$ to be the eigenvectors of the matrix $\mathsf{Cov}(X) = \sum_{i=1}^n \mathsf{vec}(A_i) \mathsf{vec}(A_i)^\top$ in descending order of eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, the random matrix $\sum_{i=1}^n g_i \lambda_i^{1/2} A_i'$ has the same distribution as $\sum_{i=1}^n g_i A_i$. One can then guarantee that $\gamma_n(K) \geq 2^{-O(n)}$ by considering the event that $g_1, \cdots, g_k$ are all $1/\mathsf{poly}(n)$ small for $k = \Theta(n/\log n)$, and applying Theorem 2.1 to control $\| \sum_{i=k+1}^n g_i \lambda_i^{1/2} A_i' \|_{\mathsf{op}}$.*

References

[ANTSV02] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM (JACM)*, 49(4):496–511, 2002.

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

[Ban98]   Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[Ban10]   Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Symposium on Foundations of Computer Science*, pages 3–10, 2010.

[Ban19]   Nikhil Bansal. On a generalization of iterated and randomized rounding. In *Symposium on Theory of Computing, STOC*, pages 1125–1135, 2019.

[BBvH21]   Afonso S Bandeira, March T Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *arXiv preprint arXiv:2108.06312*, 2021.

[BCKL14]   Nikhil Bansal, Moses Charikar, Ravishankar Krishnaswamy, and Shi Li. Better algorithms and hardness for broadcast scheduling via a discrepancy approach. In *Symposium on Discrete Algorithms*, pages 55–71, 2014.

[Bec81]   József Beck. Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981.

[BF81]   József Beck and Tibor Fiala. "integer-making" theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[BJM22]   Nikhil Bansal, Haotian Jiang, and Raghu Meka. An elementary proof of a more general matrix discrepancy bound. *Unpublished Manuscript*, 2022.

[BJSS20]   Nikhil Bansal, Haotian Jiang, Sahil Singla, and Makrand Sinha. Online vector balancing and geometric discrepancy. In *Symposium on Theory of Computing*, pages 1139–1152, 2020.

[BRS22]   Nikhil Bansal, Lars Rohwedder, and Ola Svensson. Flow time scheduling and prefix beck-fiala. *arXiv preprint arXiv:2202.02217*, 2022.

[BS13]   Nikhil Bansal and Joel Spencer. Deterministic discrepancy minimization. *Algorithmica*, 67(4):451–471, 2013.

[BSS12]   Joshua Batson, Daniel A Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012.

[Cha00]   Bernard Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, 2000.

[CST+14]   William Chen, Anand Srivastav, Giancarlo Travaglini, et al. *A panorama of discrepancy theory*, volume 2107. Springer, 2014.

[DJR22]   Daniel Dadush, Haotian Jiang, and Victor Reis. A new framework for matrix discrepancy: Partial coloring bounds via mirror descent. In *Symposium on Theory of Computing (STOC)*, 2022.

[ES18]   Ronen Eldan and Mohit Singh. Efficient algorithms for discrepancy minimization in convex sets. *Random Structures & Algorithms*, 53(2):289–307, 2018.

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

[Glu89]   Efim Davydovich Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85, 1989.

[HRS22]   Samuel B Hopkins, Prasad Raghavendra, and Abhishek Shetty. Matrix discrepancy from quantum communication. In *Symposium on Theory of Computing (STOC)*, 2022.

[HSSZ19]   Christopher Harshaw, Fredrik Sävje, Daniel Spielman, and Peng Zhang. Balancing covariates in randomized experiments using the gram-schmidt walk. *arXiv e-prints*, pages arXiv–1911, 2019.

[JKS19]   Haotian Jiang, Janardhan Kulkarni, and Sahil Singla. Online geometric discrepancy for stochastic arrivals with applications to envy minimization. *arXiv preprint arXiv:1910.01073*, 2019.

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

[LM15]   Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.

[LP86]   Françoise Lust-Piquard. Inégalités de khintchine dans $c_p(1 < p < \infty)$. *C. R. Math. Acad. Sci. Paris*, 303(7):289–292, 1986.

[LPP91]   Françoise Lust-Piquard and Gilles Pisier. Non commutative khintchine and paley inequalities. *Arkiv för matematik*, 29(1):241–260, 1991.

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

[Mat09]   Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 2009.

[Mek14]   Raghu Meka. Discrepancy and beating the union bound. 2014.

[MN12]   S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Symposium on Theory of computing*, pages 1285–1292, 2012.

[MSS15]   Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison—singer problem. *Annals of Mathematics*, pages 327–350, 2015.

[NTZ13]   Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Symposium on Theory of computing*, pages 351–360, 2013.

[Oli10]   Roberto Oliveira. Sums of random hermitian matrices and an inequality by rudelson. *Electron. Commun. Probab.*, 15:no. 19, 203–212, 2010.

[Pis03]   Gilles Pisier. *Introduction to operator space theory.* Cambridge University Press, 2003.

[Rot13]   Thomas Rothvoss. Approximating bin packing within o $(\log OPT * \log \log OPT)$ bins. In *Foundations of Computer Science*, pages 20–29, 2013.

[Rot17]   Thomas Rothvoss. Constructive discrepancy minimization for convex sets. *SIAM Journal on Computing*, 46(1):224–234, 2017.

[RR20a]   Victor Reis and Thomas Rothvoss. Linear size sparsifier and the geometry of the operator norm ball. In *Symposium on Discrete Algorithms (SODA)*, pages 2337–2348. SIAM, 2020.

[RR20b]   Victor Reis and Thomas Rothvoss. Vector balancing in lebesgue spaces. *arXiv preprint arXiv:2007.05634*, 2020.

[Spe85]   Joel Spencer. Six standard deviations suffice. *Transactions of the American mathematical society*, 289(2):679–706, 1985.

[Tao12]   Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

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

[Tro18]   Joel A. Tropp. Second-order matrix concentration inequalities. *Applied and Computational Harmonic Analysis*, 44(3):700–736, 2018.

[Ver18]   Roman Vershynin. *High-dimensional probability: An introduction with applications in data science.* Cambridge university press, 2018.

[Vu14]   Van H. Vu. *Modern aspects of random matrix theory*, volume 72. American Mathematical Society, 2014.

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