THEORY OF COMPUTING, Volume 22 (2111), 2026, pp. 1–25
www.theoryofcomputing.org

A New Framework for Matrix Discrepancy: Partial Coloring Bounds via Mirror Descent

Daniel Dadush Haotian Jiang Victor Reis

Received August 16, 2022; Revised September 30, 2023; Published December 5, 2026

Abstract. Motivated by the Matrix Spencer conjecture, we study the problem of finding signed sums of matrices with a small matrix norm. A well-known strategy to obtain these signs is to prove, given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$, a Gaussian measure lower bound of $2^{-O(n)}$ for a scaling of the discrepancy body $\{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le 1\}$. We show this is equivalent to covering its polar with $2^{O(n)}$ translates of the cube $\frac{1}{n}B_\infty^n$, and construct such a cover via mirror descent. As applications of our framework, we show:

Matrix Spencer for low-rank matrices. If the matrices satisfy $\|A_i\|_{\mathrm{op}} \le 1$ and $\mathsf{rank}(A_i) \le r$, we can efficiently find a coloring $x \in \{\pm 1\}^n$ with $\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}} \lesssim \sqrt{n \log(\min(rm/n, r))}$. This improves upon the naive $O(\sqrt{n \log r})$ bound for random coloring and proves the matrix Spencer conjecture when $rm \le n$.

A conference version of this paper appeared in the Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2022) [18].

ACM Classification: G.1.6

AMS Classification: 68W20

Key words and phrases: matrix, discrepancy, mirror descent, Gaussian measure, covering number, operator norm

© 2026 Daniel Dadush and Haotian Jiang and Victor Reis
Licensed under a Creative Commons Attribution License (CC-BY)
DOI: 10.4086/toc.2026.v022a2111

Matrix Spencer for block diagonal matrices. For block diagonal matrices with $\|A_i\|_{\mathrm{op}} \le 1$ and block size h, we can efficiently find a coloring $x \in \{\pm 1\}^n$ with $\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}} \lesssim \sqrt{n \log(hm/n)}$. This bound was previously shown in [Levy, Ramadas and Rothvoss, IPCO 2017] under the assumption $h \le \sqrt{n}$, which we remove. Using our proof, we reduce the matrix Spencer conjecture to the existence of a $O(\log(m/n))$ quantum relative entropy net on the spectraplex.

Matrix discrepancy for Schatten norms. We generalize our discrepancy bound for matrix Spencer to Schatten norms $2 \le p \le q$. Given $\|A_i\|_{S_p} \le 1$ and $\mathsf{rank}(A_i) \le r$, we can efficiently find a partial coloring $x \in [-1,1]^n$ with $|\{i : |x_i| = 1\}| \ge n/2$ and $\|\sum_{i=1}^n x_i A_i\|_{S_q} \lesssim \sqrt{n \min(p, \log(rk))} \cdot k^{1/p-1/q}$, where $k := \min(1, m/n)$.

Our partial coloring bound is tight when $m = \Theta(\sqrt{n})$. We also provide tight lower bounds of $\Omega(\sqrt{n})$ for rank-1 matrix Spencer when $m = n$, and $\Omega(\sqrt{\min(m,n)})$ for $S_2 \to S_\infty$ discrepancy, precluding a matrix version of the Komlós conjecture.

1 Introduction

Discrepancy minimization has been a well-studied area of research both in mathematics and computer science [16, 28]. We start with a classical setting: given vectors $a_1, \dots, a_n \in \mathbb{R}^m$ each satisfying $\|a_i\|_\infty \le 1$, the goal is to find a coloring $x \in \{\pm 1\}^n$ that minimizes the discrepancy, defined as $\|\sum_{i=1}^n x_i a_i\|_\infty$. A seminal result of Spencer [39] improves upon the $O(\sqrt{n \log m})$ bound of a random coloring via Chernoff and union bound:

Theorem 1.1 (Spencer [39])**.** *Let $m \ge n$. Given vectors $a_1, \dots, a_n \in \mathbb{R}^m$ with $\|a_i\|_\infty \le 1$, there exists $x \in \{\pm 1\}^n$ such that $\|\sum_{i=1}^n x_i a_i\|_\infty \lesssim \sqrt{n \log(2m/n)}$.*

In particular, when $m = n$, Theorem 1.1 states that the discrepancy is at most $O(\sqrt{n})$, as opposed to the $O(\sqrt{n \log n})$ bound for a random coloring. Spencer's theorem is known to be tight up to constants for all $m \ge n$ [16, 28].

The partial coloring method. All known proofs of Spencer's theorem are essentially based on the *partial coloring* method, one of the most important and widely applied techniques in discrepancy theory. The method states that to obtain the type of discrepancy bound in Theorem 1.1, it suffices to prove the same bound for a partial coloring $x \in [-1,1]^n$ with at least $\Omega(n)$ coordinates in $\{\pm 1\}$. This process is then iterated over the set of coordinates $\{i : |x_i| < 1\}$ to obtain a full coloring. For Spencer-type problems, the discrepancy of the full coloring is at most a constant factor off from the discrepancy of the partial coloring (see Theorem 3.2).

The partial coloring method was developed in the early 80s by Beck and refined by Spencer using the entropy method [13, 39]. A convex geometry view of partial coloring was developed independently by Gluskin [21]. While these original arguments used the pigeonhole principle and were non-algorithmic, a breakthrough result of Bansal [6], followed by a rich line of work [26, 36, 25, 20, 34], gave various algorithmic versions. These recent developments also led to new results in approximation algorithms and differential privacy [35, 32, 7, 12].

Matrix Spencer setting. A natural generalization of Spencer's setting to matrices is the following. Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$, each satisfying $\|A_i\|_{\mathrm{op}} \le 1$, the goal is to find a coloring $x \in \{\pm 1\}^n$ that minimizes $\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}}$. In particular, Spencer's setting corresponds to the case where all matrices A_i are diagonal.

In the matrix Spencer setting, the non-commutative Khintchine inequality of Lust-Piquard and Pisier [27, 33] shows that a random coloring $x \in \{\pm 1\}^n$ has expected discrepancy $\mathbb{E}[\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}}] \lesssim \sqrt{n \log r}$, where each matrix A_i has rank at most $r \le m$. It is conjectured that the discrepancy bound in Theorem 1.1 can be generalized as follows:

Conjecture 1.2 (Matrix Spencer conjecture [30, 42])**.** *Let $m \ge \sqrt{n}$. Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{\mathrm{op}} \le 1$, there exists $x \in \{\pm 1\}^n$ such that*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{\mathrm{op}} \lesssim \sqrt{n \cdot \max(1, \log(m/n))}.$$

In particular, when $\sqrt{n} \le d \le n$, the conjectured discrepancy bound is $O(\sqrt{n})$. Despite significant effort, Theorem 1.2 has remained largely open, with partial progress for block diagonal matrices [25]. A subsequent work showed that Theorem 1.2 holds for matrices of rank at most $n/\log^3(n)$, and that in general a bound of $O(n^{1/4} d^{1/4} \log d)$ holds [11].

Matrix discrepancy for Schatten norms. More generally, let[1] $2 \le p \le q \le \infty$, we consider the following matrix discrepancy setting for Schatten norms. Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$, each satisfying $\|A_i\|_{S_p} \le 1$, where $\|\cdot\|_{S_p}$ denotes the Schatten-p norm. The goal is to find a coloring $x \in \{\pm 1\}^n$ to minimize $\|\sum_{i=1}^n x_i A_i\|_{S_q}$, the $S_p \to S_q$ discrepancy. In particular, the matrix Spencer setting corresponds to the case where $p = q = \infty$.

The diagonal case of $S_p \to S_q$ discrepancy, i.e. $\ell_p \to \ell_q$ discrepancy for vectors, is well studied (see [19, 34] and the references therein). In fact, the well-known Komlós conjecture asserts that the $\ell_2 \to \ell_\infty$ discrepancy can be upper bounded by a universal constant. For general $\ell_p \to \ell_q$ discrepancy, Reis and Rothvoss [34] proves an optimal partial coloring bound of $O(\sqrt{\min(p, \log(m/n))} \cdot n^{1/2-1/p+1/q})$, assuming $m \ge n$ and $2 \le p \le q \le \infty$. It is a natural question whether these bounds generalize to $S_p \to S_q$ discrepancy.

The challenge in using partial coloring method for matrix discrepancy. Central to the partial coloring method is to show that the discrepancy body $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le t\}$, i.e. the set of fractional colorings with discrepancy at most t under norm $\|\cdot\|$, is "large" in some sense. A natural notion of largeness, due to Gluskin [21], is that the body D has Gaussian measure at least $2^{-O(n)}$. This measure of largeness has been adopted (sometimes implicitly) in essentially all work on constructive partial coloring [6, 26, 36, 25, 20, 34].

For the setting in Theorem 1.1, the discrepancy body D is a polytope defined by the intersection of strips of the form $|\langle r_i, x\rangle| \le t$, where $r_i \in \mathbb{R}^n$ are the rows of the $m \times n$ matix whose columns are $a_1, \dots, a_n$. Therefore, Šidák's lemma [38] can be readily used to give a Gaussian measure lower bound of the form $\gamma_n(D) \ge \prod_{i=1}^m \gamma_n(\{x \in \mathbb{R}^n : |\langle r_i, x\rangle| \le t\})$.

[1] We make the assumption that $p \le q$ to avoid a polynomial dependence on m in the discrepancy bound. If $q < p$, then even a single matrix (i.e. $n = 1$) can have discrepancy $m^{1/q-1/p}$.

In the setting of matrix discrepancy, however, the discrepancy body D has an infinite number of facets. This prevents the use of Gaussian correlation inequalities to lower bound $\gamma_n(D)$. To get around this barrier and use the partial coloring method for matrix discrepancy, one needs a different approach for proving Gaussian measure lower bounds.

1.1 Our results

We lower bound the Gaussian measure of the discrepancy body D via covering numbers for its polar D° with respect to the ℓ_∞-ball (see Section 3.1). We then prove the desired covering number estimates using mirror descent, the powerful convex optimization primitive of Nemirovski and Yudin [31] (see Sections 3.2 to 3.4). Our method yields the following applications.

Matrix Spencer for low-rank matrices. Our first result is the following improvement over the $O(\sqrt{n \log r})$ bound for random coloring in the matrix Spencer setting.

Theorem 1.3 (Matrix Spencer for low-rank matrices)**.** *Let $m \geq \sqrt{n}$. Given symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathrm{op}} \leq 1$ and $\mathsf{rank}(A_i) \leq r$ for all $i \in [n]$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ such that*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{\mathrm{op}} \lesssim \sqrt{n \cdot \max(1, \log(r \cdot \min(1, m/n)))}.$$

When the input matrices have rank $r \lesssim n/m$, the discrepancy bound in Theorem 1.3 is $O(\sqrt{n})$ and this proves Theorem 1.2 for low rank matrices in the regime where $m \leq n$.[2]

Matrix Spencer for block diagonal matrices. Our second application is the following improved matrix Spencer bound for block diagonal matrices.

Theorem 1.4 (Matrix Spencer for block diagonal matrices)**.** *Let $m \geq \sqrt{n}$ and $h \leq m$. Given block diagonal symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathrm{op}} \leq 1$ and block size $h \times h$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{\mathrm{op}} \lesssim \sqrt{n \cdot \max(1, \log(hm/n))}.$$

In particular, Theorem 1.4 proves Theorem 1.2 whenever $h \lesssim n/m$. This bound was previously proved in [25] under the assumption $h \leq \sqrt{n}$,[3] which we remove here.

We also obtain the following reduction of Theorem 1.2 to the construction of a better quantum relative entropy net for the spectraplex $\mathcal{S}_m := \{X \in \mathbb{R}^{m \times m} : X \succeq 0, \mathsf{tr}(X) = 1\}$.

[2] In this case, the conjecture already follows from Corollary A.2, which gives a bound of $\sqrt{r \cdot m} = \sqrt{n} \cdot \sqrt{rm/n}$ for matrices of rank at most r, but Theorem 1.3 scales exponentially better in terms of rm/n.

[3] While their Theorem 3 does not make this assumption explicit, their analysis requires $\lfloor \frac{n}{16h^2} \rfloor$ to be a positive integer and that $m \geq n$. In particular, this last assumption is used in the last inequality of their proof, namely that $\log(h^2 m/n) = O(\log(hm/n))$.

Corollary 1.5 (Better entropy net implies matrix Spencer). *Let $m \geq \sqrt{n}$. If we can find $T \subseteq \mathcal{S}_m$ with $|T| \leq 2^{O(n)}$ such that for each $X \in \mathcal{S}_m$ there exists $Y \in T$ with $S(X\|Y) \lesssim \max(1, \log(m/n))$, where $S(X\|Y)$ is the quantum relative entropy between X and Y, then the matrix Spencer conjecture is true.*

In particular, in the proof of Theorem 1.4, we construct a $O(\max(1, \log(hm/n)))$-relative entropy net for the set of block diagonal matrices on $\mathcal{S}_m$ with block size $h \times h$ (see Section 3.4). Our construction of such relative entropy nets might be of independent interest.

Matrix discrepancy for Schatten norms. Theorem 1.3 is a special case of the following general matrix discrepancy bound for Schatten norms.

Theorem 1.6 (Matrix discrepancy for Schatten norms). *Let $m \geq \sqrt{n}$ and $2 \leq p \leq q \leq \infty$. Given symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$ and $\operatorname{rank}(A_i) \leq r$ for all $i \in [n]$, one can efficiently find $x \in [-1, 1]^n$ so that $|\{i : |x_i| = 1\}| \geq n/2$ and*

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{S_q} \lesssim \sqrt{n \cdot \min(p, \max(1, \log(rk)))} \cdot k^{1/p - 1/q},$$

where we denote $k := \min(1, m/n)$. Moreover, we can find a full coloring $x \in \{\pm 1\}^n$ at the expense of a factor of $(1/2 + 1/q - 1/p)^{-1}$.

Our partial coloring result in Theorem 1.6 is tight when either $m = \Theta(\sqrt{n})$ (for which we give an alternative proof using Banaszczyk's result [4] in Appendix A), or when $r = 1$ and $m \geq n$. We provide matching lower bounds for both cases in Sections 6.1 and 6.2. In particular, our lower bound examples imply a tight $\Omega(\sqrt{n})$ lower bound for rank-1 matrix Spencer when $m = n$, which is a corollary of Lemma 6.2:

Corollary 1.7 (Rank-1 matrix Spencer lower bound). *There exist rank-1 symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{n \times n}$ with $\|A_i\|_{\mathrm{op}} \leq 1$ such that any $x \in \{\pm 1\}^n$ has $\| \sum_{i=1}^n x_i A_i \|_{\mathrm{op}} \gtrsim \sqrt{n}$.*

Another immediate consequence of our lower bounds is an optimal $\Omega(\sqrt{\min(m, n)})$ lower bound for $S_2 \to S_\infty$ discrepancy. This is in stark contrast to the well-known Komlós conjecture for vectors, which asserts that the $\ell_2 \to \ell_\infty$ discrepancy is $O(1)$. Theorem 1.8 states that such a conjecture is far from being true for matrices.

Corollary 1.8 (Lower bound for matrix Komlós). *For any m and n, there exist symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_F \leq 1$ such that any $x \in \{\pm 1\}^n$ has $\| \sum_{i=1}^n x_i A_i \|_{\mathrm{op}} \gtrsim \sqrt{\min(m, n)}$.*

Finally, we propose the following generalization of Theorem 1.2:

Conjecture 1.9 ($S_p \to S_q$ Matrix discrepancy). *Let $m \geq \sqrt{n}$ and $2 \leq p \leq q \leq \infty$. Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, there exists $x \in \{\pm 1\}^n$ such that*

$$\left\| \sum_{i=1}^{n} x_i A_i \right\|_{S_q} \lesssim \sqrt{n \cdot \min(p, \max(1, \log(m/n)))} \cdot \min(1, m/n)^{1/p - 1/q}.$$

When $m = n$, the right hand side is $O(\sqrt{n})$, and for diagonal matrices the conjecture is known to be true for any $2 \leq p \leq q$. When $p = q$, the conjecture is also known to be true for diagonal matrices for all m and n [34].

1.2 Overview of our approach

We give a brief overview of our partial coloring framework in this subsection, and leave a more detailed discussion to Section 3.

Partial coloring via covering numbers. Let $K := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le 1\}$ be the unit discrepancy body[4] and t be the target discrepancy bound. A recent refinement by Reis and Rothvoss [34] of Gluskin's convex geometry approach [21] shows that whenever $\gamma_n(tK) \ge 2^{-O(n)}$ for any constant in the exponent, one can efficiently find a partial coloring $x \in O(t) \cdot K \cap [-1,1]^n$ with at least $n/2$ coordinates in $\{-1,1\}$ (see Theorem 3.1). For settings where the target discrepancy bound is $n^{\Omega(1)}$, we may iterate the partial coloring to find a full coloring with the same discrepancy bound up to constants (see Theorem 3.2).

Our new approach for proving a Gaussian measure lower bound $\gamma_n(tK) \ge 2^{-O(n)}$ is via the covering numbers (Theorem 2.2) of K or K° with respect to the Euclidean ball B_2^n or the ℓ_∞ ball B_∞^n. In particular, since $\gamma_n(\sqrt{n}B_2^n)$ has constant Gaussian measure, as long as $\mathcal{N}(\sqrt{n}B_2^n, tK) \le 2^{O(n)}$, we get $\gamma_n(tK) \ge 2^{-O(n)}$. Using the duality of covering numbers and connections with volume, we obtain several equivalent conditions for $\gamma_n(tK) \ge 2^{-O(n)}$ in terms of covering (Theorem 3.3). The condition that we will work with is $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \le 2^{O(n)}$, where $K^\circ = \{(\langle A_1, U\rangle, \dots, \langle A_n, U\rangle) : \|U\|_* \le 1\}$ is the polar discrepancy body.

Covering via mirror descent. We prove the covering number bound $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \le 2^{O(n)}$ using mirror descent, a powerful convex optimization primitive of Nemirovski and Yudin [31] (see Section 3.2 for an overview). In particular, denote the linear map $\mathcal{A}(U) := (\langle A_1, U\rangle, \dots, \langle A_n, U\rangle)$. We shall assume that each $\|A_i\| \le 1$. This is true for the matrix Spencer setting with $\|\cdot\|$ being the operator norm. In the more general setting of matrix discrepancy for Schatten norms, we have $\|A_i\|_{S_p} \le 1$ while the norm for measuring discrepancy is $\|\cdot\|_{S_q}$. One can get around this issue by leveraging known covering number estimates between Schatten classes (Theorem 2.6).

For any matrix $\|U\|_* \le 1$, consider minimizing the function $f_U(X) := \|\mathcal{A}(X - U)\|_\infty$ over the dual unit ball $B_* := \{U : \|U\|_* \le 1\}$. The function has minimum value $f_U(U) = 0$ and since it has subgradients in $\{\pm A_1, \dots, \pm A_n\}$ with $\|A_i\| \le 1$, the function $f_U(X)$ is 1-Lipschitz with respect to the dual norm $\|\cdot\|_*$. So as long as there exists a 1-strongly convex mirror map Φ on B_*, we can minimize $f_U(X)$ by starting from some matrix $U_0 = U_0(U) \in B_*$ and running mirror descent for n steps. Denoting by U_s the matrix in the s-th step, standard guarantees for mirror descent (Theorem 3.5) yield

$$\min_{s\in[n]} f_U(U_s) = \min_{s\in[n]} f_U(U_s) - f_U(U) \le \sqrt{\frac{2D_\Phi(U, U_0)}{n}}, \tag{1.1}$$

where $D_\Phi(U, U_0) = \Phi(U) - \Phi(U_0) - \langle \nabla\Phi(U_0), U - U_0\rangle$ is the Bregman divergence. We let T be the set of all matrices encountered when running mirror descent for all possible $U \in B_*$, i.e. $T := \{U_s : s \in [n], U \in B_*\}$, and $T_0 := \{U_0 : U \in B_*\}$ be the set of all starting matrices. The net $\mathcal{A}(T)$ will be our covering for K°.

[4] To avoid confusion when talking about discrepancy bodies, K denotes the unit discrepancy body, and D denotes a scaling of K by the target discrepancy bound.

To see that this indeed gives a good covering, we denote $D_\Phi^{\max} := \sup_{U \in B_*} D_\Phi(U\|U_0)$. By the definition of the function f_U, we have from (1.1) that

$$\min_{s\in[n]} \|\mathcal{A}(U) - \mathcal{A}(U_s)\|_\infty \le \sqrt{\frac{2D_\Phi(U,U_0)}{n}} \le \sqrt{\frac{2D_\Phi^{\max}}{n}},$$

and so the dual body admits the covering $K^\circ \subseteq \mathcal{A}(T) + \sqrt{2D_\Phi^{\max}/n} \cdot B_\infty^n$. Thus as long as our target discrepancy bound $t \le \sqrt{2nD_\Phi^{\max}}$, we have $\mathcal{N}(K^\circ, \frac{t}{n}B_\infty^n) \le |T|$, which we need to show to be at most $2^{O(n)}$.

The key observation we make here is that for our choices of the mirror maps in Sections 4 and 5, U_s only depends[5] on the sum of the subgradients, but not on their order. Since there are only $2n$ choices of subgradients $\{\pm A_i\}_{i\in[n]}$ and we run mirror descent for n steps, a counting argument reveals that there are at most $2^{O(n)}$ possible sums of gradients (Theorem 3.6). So long as the starting matrices satisfy $|T_0| \le 2^{O(n)}$, we have $|T| \le |T_0| \cdot 2^{O(n)} \le 2^{O(n)}$.

A view of mirror descent as refining the net. In the diagonal case, i.e. Spencer's setting, we can directly build the net T by repeatedly sampling the ith diagonal coordinate $e_i e_i^\top$ proportional to its weight in the target matrix. Since the set of diagonal matrices on the Schatten-1 ball has only $2m$ vertices $\{\pm e_i e_i^\top\}_{i\in[m]}$, the approximate Carathéodory theorem (see [40], Theorem 0.0.2) implies that the image of the net $\mathcal{A}(T)$ already gives a good covering for K°, and mirror descent is not necessary in this case.

However, this argument fails beyond diagonal matrices, as the number of vertices becomes infinite. In these more general cases, we use mirror descent to boost a coarse net T_0 to a finer net T which has a better covering guarantee in the image space, at the expense of increasing the size of the net by a factor of $2^{O(n)}$.

Relative entropy nets for the Spectraplex. For our application in Section 5 to low-rank matrices, it suffices to take $T_0 = \{0\}$. For the application in Section 4 to block diagonal matrix Spencer, we run mirror descent on the spectraplex $\mathcal{S}_m := \{X \in \mathbb{R}^{m\times m} : X \succeq 0, \mathrm{tr}(X) = 1\}$ and carefully construct a set $|T_0| \le 2^{O(n)}$ with small $D_\Phi^{\max}$. Since $D_\Phi(X\|Y)$ is the quantum relative entropy between X and Y in the spectraplex setup, we refer to such T_0 as a (quantum) relative entropy net (Theorem 3.7).

We use an operator norm net for the Schatten-1 ball from [22] to construct a relative entropy net with error $O(\log(m^2/n))$ for the spectraplex $\mathcal{S}_m$ (Theorem 3.8). When restricted to block diagonal matrices with block size $h \times h$, we use a hybrid of this argument and the earlier approximate Caratheodory argument to find a refined relative entropy net with error $O(\log(hm/n))$ (Theorem 3.9). Taking T_0 to be this net in our mirror descent framework gives Theorem 1.4. This also allows us to reduce the matrix Spencer conjecture to the existence of a better relative entropy net with error $O(\log(m/n))$ for the spectraplex (Theorem 1.5).

[5] In general, mirror descent projects back onto the feasible set according to the Bregman divergence in each iteration, and therefore might not satisfy this property.

1.3 Further related work

Banaszczyk's approach. While the partial coloring method has been extensively applied in discrepancy and obtains the optimal bound for many problems, for several applications where the target discrepancy bound is $n^{o(1)}$ (e.g. the Komlós problem or Tusnady's problem), partial coloring is potentially sub-optimal by a logarithmic factor. In breakthrough work, Banaszczyk [4] obtained an improvement over the partial coloring method for these applications using deep techniques from convex geometry. While Banaszczyk's original proof is non-constructive, a fascinating recent line of work has obtained algorithmic versions of Banaszczyk's result [17, 8, 10, 25, 9].

Matrix Spencer conjecture and non-commutative random matrix theory. The typical value of $\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}}$ for a random coloring has attracted significant attention in random matrix theory. For commutative matrices, the bound $\mathbb{E}[\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}}] \lesssim \sqrt{n \log m}$ by matrix Khintchine [27, 33] or matrix Chernoff bound [1] is in general tight. It is also known to be tight for Toeplitz matrices [29]. For matrices with certain non-commutative structures (e.g. random Gaussian matrices), improved bounds of $O(\sqrt{n})$ are known (see [40, 5]). In the context of Theorem 1.2, these results imply that a random coloring already achieves the conjectured bound when the input matrices have certain non-commutative structures. On the other hand, by Theorem 1.1, Theorem 1.2 is known when all the matrices commute.

Concurrent and independent work. In concurrent and independent work, Hopkins, Raghavendra and Shetty [23] proved a bound of $\sqrt{n \log\left(\mathrm{tr}(\sum_{i=1}^n A_i^2)/n^{1.5}\right)}$ for matrix Spencer using quantum communication. For matrices of rank at most r, this gives a bound of $\sqrt{n \cdot \max(1, \log(r^2/n))}$. For $r \le \sqrt{n}$, this simplifies to $O(\sqrt{n})$, whereas Theorem 1.3 gives $O(\sqrt{n \log r})$. When $r = m$, the discrepancy bound in [23] coincides with Theorem 1.3, both giving a bound of $\sqrt{n \cdot \log(m/\sqrt{n})}$ which is optimal when $m = \sqrt{n}$. However, our approach is completely different and can also be used to show matrix discrepancy bounds for block diagonal matrices and general Schatten norms. We believe both approaches are interesting and may lead to further progress in resolving the matrix Spencer conjecture.

2 Preliminaries

Norms and convex bodies. A convex body is a compact convex set with non-empty interior. We say a convex set K is symmetric if $x \in K$ implies $-x \in K$. We use $\|\cdot\|_p$ to denote the ℓ_p-norm and $\|\cdot\|_{S_p}$ to denote the Schatten-p norm. In particular, the operator norm $\|\cdot\|_{\mathrm{op}} = \|\cdot\|_{S_\infty}$ and the Frobenius norm $\|\cdot\|_F = \|\cdot\|_{S_2}$. We use B_p^n to denote the unit ℓ_p-ball in $\mathbb{R}^n$ and $B_{S_p}^n := \{A \in \mathbb{R}^{n\times n} : \|A\|_{S_p} \le 1\}$ to denote the unit Schatten-p ball in $\mathbb{R}^{n\times n}$, with $B_{\mathrm{op}}^n := B_{S_\infty}^n$. Let $\mathbb{R}_+^n$ denote the set of non-negative vectors in $\mathbb{R}^n$ and denote the simplex $\Delta_n := \{x \in \mathbb{R}_+^n : \|x\|_1 = 1\}$. Let $\mathbb{S}_+^n$ (resp. $\mathbb{S}_{++}^n$) denote the set of positive semidefinite (resp. positive definite) $n \times n$ matrices, and define the spectraplex $\mathcal{S}_n := \{X \in \mathbb{S}_+^n : \mathrm{tr}(X) = 1\}$. For a norm $\|\cdot\|$ in $\mathbb{R}^n$, we define the dual

norm as $\|x\|_* := \sup\{\langle y, x\rangle : y \in \mathbb{R}^n, \|y\| \le 1\}$. Dual norms are similarly defined for matrix norms with the trace inner product $\langle A, B\rangle := \mathsf{tr}(A^\top B)$.

Convex functions. A convex function $f : \mathcal{X} \to \mathbb{R}$ is said to be L-Lipschitz with respect to a norm $\|\cdot\|$ if $\|g\|_* \le L$ for all subgradients $g \in \partial f(x)$. We say that f is α-strongly convex with respect to a norm $\|\cdot\|$ if $f(y) \ge f(x) + g^\top(y-x) + \frac{\alpha}{2}\|x-y\|^2$, for all $x, y \in \mathcal{X}$ and all subgradients $g \in \partial f(x)$.

Polar. Given a convex set $K \subseteq \mathbb{R}^n$ with $0 \in K$, we define the polar of K to be $K^\circ := \{y \in \mathbb{R}^n : \sup_{x\in K}\langle x, y\rangle \le 1\}$. It is immediate from the definition that for any constant $t > 0$, $(tK)^\circ = \frac{1}{t}K^\circ$. When K is closed, the polarity theorem states that $(K^\circ)^\circ = K$.

Lemma 2.1 (Polar of discrepancy set)**.** *Given matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ and a norm $\|\cdot\|$ in $\mathbb{R}^{m\times m}$, we define the unit discrepancy set as $K := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le 1\}$. Then $K' := \{(\langle A_1, U\rangle, \dots, \langle A_n, U\rangle) : \|U\|_* \le 1\}$ is the polar body $K' = K^\circ$.*

Proof. By the definition of polar body, we may write

$$\begin{aligned}(K')^\circ &= \left\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i\langle A_i, U\rangle \le 1,\ \forall\, U \text{ s.t. } \|U\|_* \le 1\right\}\\ &= \left\{x \in \mathbb{R}^n : \left\langle \sum_{i=1}^n x_i A_i, U\right\rangle \le 1,\ \forall\, U \text{ s.t. } \|U\|_* \le 1\right\}\\ &= K,\end{aligned}$$

by the definition of dual norm. It then follows from the polarity theorem that $K' = K^\circ$. □

Covering numbers. We start with the definition of covering numbers.

Definition 2.2 (Covering numbers)**.** For two convex bodies $K, T \subseteq \mathbb{R}^n$, we define the covering number $\mathcal{N}(K, T)$ as the minimum number N such that there exist centers $x_1, \dots, x_N \in \mathbb{R}^n$ with $K \subseteq \cup_{i=1}^N (x_i + T)$, i.e. K can be covered by N translates of T.

We need the following few standard facts about covering numbers (see [2]).

Lemma 2.3 (Volume bounds for covering numbers)**.** *Given convex bodies $K, T \subseteq \mathbb{R}^n$. If T is symmetric, we have $\frac{\mathsf{vol}_n(K)}{\mathsf{vol}_n(T)} \le \mathcal{N}(K, T) \le 2^n \cdot \frac{\mathsf{vol}_n(K+\frac{T}{2})}{\mathsf{vol}_n(T)}$.*

Lemma 2.4 (Symmetrization)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body, then $\mathcal{N}(K - K, K) \le 2^{O(n)}$.*

Theorem 2.5 (Duality of covering numbers, [24])**.** *Given symmetric convex bodies $K, T \subseteq \mathbb{R}^n$, we have*

$$2^{-\Theta(n)} \cdot \mathcal{N}(T^\circ, K^\circ) \le \mathcal{N}(K, T) \le 2^{\Theta(n)} \cdot \mathcal{N}(T^\circ, K^\circ).$$

We will also need the following upper bound on the covering numbers of Schatten balls[6].

Theorem 2.6 ([22], Theorem 1.1)**.** *Let $m, n \in \mathbb{N}$ and $1 \le p \le q \le \infty$. Then we have*

$$\mathcal{N}\left(B_{S_p}^m, \min\left(1, \frac{m}{n}\right)^{1/p-1/q} B_{S_q}^m\right) \le 2^{O(n)}.$$

Gaussian measure. We use $\gamma_n(\cdot)$ to denote the standard Gaussian measure on $\mathbb{R}^n$. Gaussian measure is log-concave, i.e. $\gamma_n(\lambda A + (1-\lambda)B) \ge \gamma_n(A)^\lambda \gamma_n(B)^{1-\lambda}$ for any compact subsets $A, B \subseteq \mathbb{R}^n$. In particular, by taking $A = -x + K$ and $B = x + K$ for any $x \in \mathbb{R}^n$ and symmetric convex body K, and $\lambda = 1/2$, we have the following lemma.

Lemma 2.7 (Translation decreases Gaussian measure)**.** *Given any symmetric convex body $K \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we have $\gamma_n(K) \ge \gamma_n(x + K)$.*

We also use the following powerful Gaussian correlation inequality.

Theorem 2.8 (Gaussian correlation inequality, [37])**.** *Given any symmetric convex sets $K, T \subseteq \mathbb{R}^n$, we have $\gamma_n(K \cap T) \ge \gamma_n(K) \cdot \gamma_n(T)$.*

3 Our framework for partial coloring

3.1 Partial coloring via covering numbers

Given symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$, a norm $\|\cdot\|$ on $\mathbb{R}^{m \times m}$ for measuring the discrepancy, and a target discrepancy bound t, let $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\| \le t\}$ be the associated discrepancy body. The following partial coloring lemma from [34] states that one can efficiently find a partial coloring with discrepancy $O(t)$ as long as $\gamma_n(D) \ge 2^{-O(n)}$.

Theorem 3.1 ([34], special case of Theorem 6)**.** *For any constant $\alpha > 0$, there is a constant $c := c(\alpha) > 0$ and a randomized polynomial time algorithm that for a symmetric convex set $D \subseteq \mathbb{R}^n$ with $\gamma_n(D) \ge 2^{-\alpha n}$ and a shift $y \in (-1, 1)^n$, finds $x \in (c \cdot D) \cap [-1, 1]^n$ so that $x + y \in [-1, 1]^n$ and $|\{i \in [n] : |(x + y)_i| = 1\}| \ge n/2$.*

We have the following corollary for full colorings. Here $K_S := K \cap \{x \in \mathbb{R}^n : x_i = 0, \forall i \notin S\}$.

Corollary 3.2. *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set. Given an increasing function $f : [n] \to \mathbb{R}_{>0}$ with $\gamma_S(f(|S|) \cdot K_S) \ge 2^{-O(|S|)}$ for every $S \subseteq [n]$, there exists a randomized polynomial time algorithm to find a full coloring $x \in \{\pm 1\}^n$ so that $x \in \lambda K$, where $\lambda \lesssim \sum_{i=0}^{\lfloor \log n \rfloor} f(n/2^i)$. In particular, when $f(n) \lesssim n^\beta$ for some $\beta \le 1$, we have $\lambda \lesssim \frac{1}{\beta} n^\beta$.*

[6] We note that [22] claims the bound only up to a constant depending on p and q, but their argument readily gives a universal constant in the regime of $p, q \ge 1$.

Proof. Indeed, repeated iterations of Theorem 3.1 with $y_0 := 0$ and subsequent shifts y_{i+1} being the coordinates not reaching $\{-1,1\}$ find $x := x_0 + \cdots + x_T \in \{\pm 1\}^n$ for $T := \lfloor \log n \rfloor$ with $x_t \in O(f(n/2^t)) \cdot K$. When $f(n) \lesssim n^\beta$, the summation is upper bounded by

$$\sum_{i=0}^{\infty} (n/2^i)^\beta = (1-2^{-\beta})^{-1} \cdot n^\beta \lesssim \frac{1}{\beta} \cdot n^\beta,$$

and this proves the statement. □

We show that a $2^{-O(n)}$ Gaussian measure lower bound is equivalent to a $2^{O(n)}$ upper bound for certain covering numbers.

Lemma 3.3. *The following conditions are equivalent for a symmetric convex body $D \subseteq \mathbb{R}^n$:*

1. $\gamma_n(D) \geq 2^{-O(n)}$,
2. $\mathcal{N}(\sqrt{n}B_2^n, D) \leq 2^{O(n)}$,
3. $\mathcal{N}(nB_1^n, D) \leq 2^{O(n)}$,
4. $\mathcal{N}(D^\circ, \frac{1}{\sqrt{n}}B_2^n) \leq 2^{O(n)}$,
5. $\mathcal{N}(D^\circ, \frac{1}{n}B_\infty^n) \leq 2^{O(n)}$.

Proof. We start by proving that condition (1) implies (2). Suppose $\gamma_n(D) \geq 2^{-O(n)}$, then Theorem 2.8 implies $\gamma_n(D') \geq 2^{-O(n)}$, where we define $D' := D \cap \sqrt{n}B_2^n$. We thus also have $\mathsf{vol}_n(D') \geq \gamma_n(D') \geq 2^{-O(n)}$. Then by Theorem 2.3, we have

$$\mathcal{N}(\sqrt{n}B_2^n, D) \leq \mathcal{N}(\sqrt{n}B_2^n, D') \leq 2^n \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_2^n + D')}{\mathsf{vol}_n(D')} \leq 2^n \cdot \frac{\mathsf{vol}_n(2\sqrt{n}B_2^n)}{\mathsf{vol}_n(D')} \leq 2^{O(n)}.$$

We next show that condition (2) implies (1). Since $\gamma_n(\sqrt{n}B_2^n) = \Omega(1)$, we have $\gamma_n(x + D) \geq 2^{-O(n)}$ for some $x \in \mathbb{R}^n$. Theorem 2.7 then gives $\gamma_n(D) \geq \gamma_n(x + D) \geq 2^{-O(n)}$.

The implication (3) $\Rightarrow$ (2) immediately follows from $\sqrt{n}B_2^n \subseteq nB_1^n$. To prove the reverse implication (2) $\Rightarrow$ (3), we use Theorem 2.3 to obtain

$$\mathcal{N}(\sqrt{n}B_1^n, B_2^n) \leq 2^n \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_1^n + B_2^n)}{\mathsf{vol}_n(B_2^n)} \leq 2^{O(n)} \cdot \frac{\mathsf{vol}_n(\sqrt{n}B_1^n)}{\mathsf{vol}_n(B_2^n)} \leq 2^{O(n)}.$$

It thus follows that $\mathcal{N}(nB_1^n, D) \leq \mathcal{N}(nB_1^n, \sqrt{n}B_2^n) \cdot \mathcal{N}(\sqrt{n}B_2^n, D) \leq 2^{O(n)}$.

The last two equivalences follow from the duality of covering numbers in Theorem 2.5. □

For our mirror descent framework, we use the following corollary:

Corollary 3.4. *Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$, let $K_{q+}^\circ := \{(\langle A_1, U\rangle, \ldots, \langle A_n, U\rangle) : U \in B_{S_q}^m, U \succeq 0\}$. If we have $\mathcal{N}(K_{q+}^\circ, \frac{t}{n}B_\infty^n) \leq 2^{O(n)}$, then we can efficiently find a partial coloring $x \in [-1,1]^n$ with $|\{i : |x_i| = 1\}| \geq n/2$ and $\|\sum_{i=1}^n x_i A_i\|_{S_q} \lesssim t$.*

Proof. Recall that $D := \{x \in \mathbb{R}^n : \|\sum_{i=1}^n x_i A_i\|_{S_q} \le t\}$ denotes the discrepancy body. Since $tD^\circ \subseteq K_{q+}^\circ - K_{q+}^\circ$, by Theorem 2.4 we have $\mathcal{N}(D^\circ, \frac{1}{n}B_\infty^n) = 2^{O(n)}$. The equivalence (1) $\Leftrightarrow$ (5) in Theorem 3.3 implies $\gamma_n(D) \ge 2^{-O(n)}$, and Theorem 3.1 gives the corollary. □

3.2 Mirror descent: An overview

The mirror descent method was introduced by Nemirovski and Yudin [31]. Here, we follow the presentation in [14]. Let $\mathcal{D}$ be an open subset of $\mathbb{R}^m$ and $\mathcal{X}$ a subset of its closure. We fix a convex function $f : \mathcal{X} \to \mathbb{R}$ assumed to be L-Lipschitz with respect to a norm $\|\cdot\|$, and a differentiable function $\Phi : \mathcal{D} \to \mathbb{R}$ that is ρ-strongly convex with respect to $\|\cdot\|$ and has a surjective gradient $\nabla\Phi : \mathcal{D} \to \mathbb{R}^m$. The mirror descent algorithm, given a starting point $x_0 \in \mathcal{X} \cap \mathcal{D}$, consists of the iterations

$$\nabla\Phi(y_{t+1}) := \nabla\Phi(x_t) - \eta g_t,$$
$$x_{t+1} := \operatorname{argmin}_{x \in \mathcal{X}\cap\mathcal{D}} D_\Phi(x, y_{t+1}),$$

where $g_t \in \partial f(x_t)$ and $D_\Phi(x, y) := \Phi(x) - \Phi(y) - \nabla\Phi(y)^\top(x - y)$ is the Bregman divergence. Note that $y_t \in \mathcal{D}$ and $x_t \in \mathcal{X} \cap \mathcal{D}$ for all $t \ge 0$. We use the following convergence guarantee:

Theorem 3.5 ([14], Theorem 4.2)**.** *Let f be L-Lipschitz and Φ be ρ-strongly convex with respect to $\|\cdot\|$, and $D_\Phi^{\max} \ge D_\Phi(x^*, x_0)$ be any upper bound. Then the mirror descent algorithm with $\eta := \frac{1}{L}\sqrt{\frac{2\rho D_\Phi^{\max}}{T}}$ satisfies*

$$\min_{s\in[T]} f(x_s) - f(x^*) \le L\sqrt{\frac{2D_\Phi^{\max}}{\rho T}}.$$

The Spectraplex setup. Here we take $\mathcal{X} := \mathcal{S}_m = \{X \in \mathbb{S}_+^m : \operatorname{tr}(X) = 1\}$. The mirror map is $\Phi(X) = \operatorname{tr}(X \log X)$, defined on $\mathcal{D} = \mathbb{S}_{++}^m$, which is $\frac{1}{2}$-strongly convex with respect to the Schatten-1 norm by the quantum Pinsker inequality [15]. Then the convergence bound in Theorem 3.5 becomes $2L\sqrt{\frac{S(X^*\|X_0)}{T}}$, where $S(X\|Y) := \operatorname{tr}(X(\log X - \log Y))$ is the quantum relative entropy between matrices $X, Y \in \mathcal{S}_m$. The projection step corresponds to a trace normalization, so given a starting point $X_0 \in \mathcal{S}_m \cap \mathbb{S}_{++}^m$, we may write in closed form

$$X_{t+1} = \frac{\exp(\log X_0 - \eta\sum_{i=0}^t g_i)}{\operatorname{tr}(\exp(\log X_0 - \eta\sum_{i=0}^t g_i))}, \tag{3.1}$$

for subgradients $g_i \in \partial f(X_i)$.

The Schatten norm setup. Here we take $\mathcal{X} = \mathcal{D} = \mathbb{R}^{m\times m}$, so that $X_t = Y_t$ for all t. The mirror map is $\Phi(X) := \frac{1}{2(p-1)}\|X\|_p^2$, which is known to be 1-strongly convex for all $p \in (1, 2]$ [3]. Thus given a starting point $X_0 \in \mathbb{R}^{m\times m}$, we may write in closed form

$$X_{t+1} = \nabla\Phi^{-1}\Big(\nabla\Phi(X_0) - \eta\sum_{i=0}^t g_i\Big), \tag{3.2}$$

for subgradients $g_i \in \partial f(X_i)$.

3.3 Covering via mirror descent

Given symmetric matrices $A_1, \dots, A_n$ with $\|A_i\| \le 1$ for all $i \in [n]$, where the dual norm $\|\cdot\|_*$ is either the Schatten-1 norm or the Schatten-p norm for some $p \in (1,2]$, we apply mirror descent on functions of the form $f_U(X) := \max_{i\in[n]} |\langle A_i, X - U\rangle|$ to cover the polar discrepancy body

$$K^\circ := \{\mathcal{A}(U) : \|U\|_* \le 1\}, \text{ where } \mathcal{A}(U) := (\langle A_1, U\rangle, \dots, \langle A_n, U\rangle).$$

Note that $f_U(X) = \|\mathcal{A}(X) - \mathcal{A}(U)\|_\infty$ and that f is 1-Lipschitz with respect to $\|\cdot\|_*$. The key property of such functions is that we may always choose subgradients from the set of $2n$ matrices $\{\pm A_i : i \in [n]\}$, which allows us to upper bound the number of different matrices encountered during the mirror descent process.

Lemma 3.6. *Let $\|\cdot\|_*$ be either $\|\cdot\|_{S_1}$ as in the Spectraplex Setup, or $\|\cdot\|_{S_p}$ with $p \in (1,2]$ as in the Schatten Norm Setup, and $\mathcal{X}, \mathcal{D}$ be defined accordingly. Let $T_0 \subseteq \mathcal{X} \cap \mathcal{D}$ be a set with size $|T_0| \le 2^{O(n)}$ and $K^\circ \supseteq K' = \mathcal{A}(T')$ the convex body to be covered, where $T' \subseteq \mathcal{X} \cap \mathcal{D}$. If for every $U \in T'$ there exists a starting point $U_0 := U_0(U) \in T_0$ with $D_\Phi(U, U_0) \le D_\Phi^{\max}$, then we can bound*

$$\mathcal{N}\left(K', \sqrt{\frac{D_\Phi^{\max}}{n}} B_\infty^n\right) \le 2^{O(n)}.$$

Proof. The key observation is that in either setup of mirror descent, the point X_t in (3.1) or (3.2) depends only on the starting point U_0 and on the sum of gradients $g_0, \dots, g_{t-1}$, but not on their order. Moreover, we can always choose from the set of $2n$ gradients $\{\pm A_i : i \in [n]\}$ at each step. Thus applying mirror descent to the function f_U for all possible U with the same starting point U_0, the total number $N(U_0)$ of points visited in $T := n$ iterations satisfies

$$N(U_0) \le \sum_{t=0}^{n} \binom{t+2n-1}{2n-1} \le (n+1) \cdot \binom{3n}{n} \le 2^{O(n)}.$$

Since $|T_0| \le 2^{O(n)}$, we obtain a set of $2^{O(n)}$ points $\mathcal{U}$ such that for every $Y = \mathcal{A}(U) \in K'$, there exists some $\tilde{U} \in \mathcal{U}$ so that $\|\mathcal{A}(\tilde{U}) - \mathcal{A}(U)\|_\infty = f_U(\tilde{U}) = f_U(\tilde{U}) - f_U(U) \lesssim \sqrt{D_\Phi^{\max}/n}$. □

In the Schatten Norm Setup, we shall pick $K' = K^\circ$ and $T_0 = \{0\}$, i.e. U_0 is always 0. For the Spectraplex Setup, we carefully choose a set of starting points $|T_0| \le 2^{O(n)}$ which has small $D_\Phi^{\max}$ with respect to $K' = \{\mathcal{A}(U) : U \in \mathcal{S}_m\}$. Since $D_\Phi(X\|Y)$ is the quantum relative entropy between X and Y in the Spectraplex Setup, we shall refer to the set of starting points T_0 as a (quantum) relative entropy net for $\mathcal{S}_m$.

Definition 3.7 (Quantum relative entropy net)**.** Given subsets $T, \mathcal{M} \subseteq \mathcal{S}_m$, T is a relative entropy net of $\mathcal{M}$ with error ε if for any $X \in \mathcal{M}$, we can find $Y \in T$ such that $S(X\|Y) \le \varepsilon$.

3.4 Initialization for Spectraplex setup: Relative entropy net

We start with the following lemma which constructs a relative entropy net on $\mathcal{S}_m$ from an operator norm net.

Lemma 3.8 (Relative entropy net from operator norm net)**.** *Let $X, Y \in \mathcal{S}_m$ satisfies $\|X - Y\|_{\mathrm{op}} \le \varepsilon$ for some $\varepsilon \ge 1/m$. Then $S(X\|Y') \le \log(2m\varepsilon)$, where $Y' := \frac{1}{2}(Y + \frac{I_m}{m}) \in \mathcal{S}_m$.*

Proof. Recall that $\log(\cdot)$ is operator monotone and note that $X \preceq Y + \varepsilon I_m$. We then have

$$\begin{aligned} S(X\|Y') &= \mathrm{tr}(X \cdot (\log X - \log Y')) \\ &\le \mathrm{tr}(X \cdot (\log(Y + \varepsilon I_m) - \log Y')) \\ &\le \mathrm{tr}(X) \cdot \|\log(Y + \varepsilon I_m) - \log Y'\|_{\mathrm{op}} \\ &\le \log\left(2 \cdot \left\|\frac{Y + \varepsilon I_m}{Y + \frac{I_m}{m}}\right\|_{\mathrm{op}}\right) \le \log(2m\varepsilon), \end{aligned}$$

where the first inequality follows from the operator monotonicity of $\log(\cdot)$, the second follows from matrix Hölder, and the last follows because $\varepsilon \ge 1/m$ and $\|Y\|_{\mathrm{op}} \le 1$. □

Using the lemma above, we give the following construction for relative entropy nets on $\mathcal{S}_m$.

Theorem 3.9 (Entropy net for Spectraplex)**.** *Given positive integers h, m and n such that m/h is an integer, let $\mathcal{S}_m^h \subseteq \mathcal{S}_m$ be the set of $m \times m$ block diagonal matrices on the spectraplex with block size $h \times h$. Then we can find a relative entropy net T for $\mathcal{S}_m^h$ with error at most $\max(1, \log(2hm/n))$ and size $|T| \le 2^{O(n)}$.*

Proof. By merging blocks as needed, we may assume $hm \ge n$. By Theorem 3.8, it suffices to find an operator norm net T' with size $|T'| \le 2^{O(n)}$ and distance $\varepsilon = \frac{\max\{h, \log(m/hn)\}}{n}$. Let $\ell := m/h$ be the number of blocks, $X_1, \dots, X_\ell \in \mathbb{R}^{h \times h}$ denote the blocks of matrix $X \in \mathcal{S}_m^h$, and $N := 2/\varepsilon = 2n/\max\{h, \log(\ell/n)\}$ (we assume that N is an integer). Let $Z := \{z \in \mathbb{Z}_{\ge 0}^\ell : \sum_{i=1}^\ell z_i = N\}$, and for each $z \in Z$, we define

$$T_z := \{X \in \mathcal{S}_m^h : \mathrm{tr}(X_i) = z_i/N, \forall i \in [\ell]\}.$$

It follows from a standard rounding argument that for any matrix $X \in \mathcal{S}_m^h$, one can find a matrix $Y \in \cup_{z \in Z} T_z$ with $\|X - Y\|_{\mathrm{op}} \le 1/N = \varepsilon/2$.

We first show that $|Z| \le 2^{O(n)}$. When $\ell \le 2n$, we have

$$|Z| \le \binom{N + \ell}{\ell} \le \binom{N + 2n}{2n} \le \binom{\frac{2n}{h} + 2n}{2n} \le 2^{O(n)}.$$

When $\ell \ge 2n \ge N$, we can bound

$$|Z| \le \binom{N + \ell}{N} \le \binom{2\ell}{N} \le \binom{2\ell}{\frac{2n}{\log(\ell/n)}} \le \left(\frac{e\ell \log(\ell/n)}{n}\right)^{\frac{2n}{\log(\ell/n)}} \le 2^{O(n)}.$$

It therefore suffices to construct an $\varepsilon/2$-operator norm net for each T_z.

Fix an arbitrary $z \in Z$. Note that the ith block of the matrices in T_z comes from $\frac{z_i}{N} \cdot \mathcal{S}_h$. Pick $n_i := z_i h$, we have from Theorem 2.6 that

$$\mathcal{N}\Big(\frac{z_i}{N}\mathcal{S}_h, \frac{z_i}{N} \cdot \frac{h}{n_i} B_{\mathrm{op}}^h\Big) = \mathcal{N}\Big(\mathcal{S}_h, \frac{h}{n_i} B_{\mathrm{op}}^h\Big) \le 2^{O(n_i)}.$$

We denote this net as $\widetilde{T}_{z,i}$. It follows from the above that for any $X_i \in \frac{z_i}{N}\mathcal{S}_h$, there exists $Y_i \in \widetilde{T}_{z,i}$ with $\|X_i - Y_i\|_{\mathrm{op}} \le \frac{z_i}{N} \cdot \frac{h}{n_i} = \varepsilon/2$. Define $\widetilde{T}_z := \{\mathsf{diag}(Y_1, \ldots, Y_\ell) : Y_i \in \widetilde{T}_{z,i}\ \forall i \in [\ell]\}$. Then for any $X \in T_z$, there exists $Y \in \widetilde{T}_z$ such that $\|X - Y\|_{\mathrm{op}} \le \varepsilon/2$, and thus $\widetilde{T}_z$ is indeed an $\varepsilon/2$-operator norm net for T_z. Furthermore, the size of $\widetilde{T}_z$ can be upper bounded as

$$|\widetilde{T}_z| \le \prod_{i \in [\ell]} 2^{O(n_i)} = 2^{O(\sum_{i=1}^n z_i h)} = 2^{O(Nh)} \le 2^{O(n)},$$

since $N \le 2n/h$. This proves that $\widetilde{T} := \cup_{z \in Z} \widetilde{T}_z$ is an ε-operator norm net for $\mathcal{S}_m^h$ and has size at most $|\widetilde{T}| \le 2^{O(n)}$, where we recall that $\varepsilon = \frac{\max\{h, \log(m/hn)\}}{n}$. Finally, invoking Theorem 3.8, $\widetilde{T}$ can be transformed into a relative entropy net T with size $|T| \le 2^{O(n)}$ and error at most $\log(2m\varepsilon) \le \log(2hm/n)$. This finishes the proof of the theorem. □

4 Applications of the Spectraplex setup

In this section, we prove our matrix Spencer bound for block diagonal matrices in Theorem 1.4, which we restate below.

Theorem 1.4 (Matrix Spencer for block diagonal matrices)**.** *Let $m \ge \sqrt{n}$ and $h \le m$. Given block diagonal symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{\mathrm{op}} \le 1$ and block size $h \times h$, one can efficiently find a coloring $x \in \{\pm 1\}^n$ with*

$$\Big\| \sum_{i=1}^n x_i A_i \Big\|_{\mathrm{op}} \lesssim \sqrt{n \cdot \max(1, \log(hm/n))}.$$

Proof of Theorem 1.4. By Theorem 3.9, we can find a relative entropy net T_0 of $\mathcal{S}_m^h$ with error $D_\Phi^{\max} := \max(1, \log(2hm/n))$ and size $|T_0| \le 2^{O(n)}$. Then using Theorem 3.6 with the Spectraplex Setup for $K' := \mathcal{A}(\mathcal{S}_m^h)$ and T_0 being the relative entropy net, we obtain

$$\mathcal{N}\Big(K', \frac{t}{n} B_\infty^n\Big) \le 2^{O(n)},$$

where $t = \sqrt{n \max(1, \log(2hm/n))}$. Let $\mathbb{S}_m^h$ be the set of $m \times m$ symmetric block diagonal matrices with block size $h \times h$. Define convex body $K'' := \mathcal{A}(B_{S_1}^m \cap \mathbb{S}_m^h \cap \mathbb{S}_+^m)$. We first prove that $\mathcal{N}(K'', \frac{t}{n} B_\infty^n) \le 2^{O(n)}$. Since $\mathcal{N}(K', \frac{t}{n} B_\infty^n) \le 2^{O(n)}$ by Theorem 3.9, we also have

$\mathcal{N}(\frac{j}{n^2}K', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$ for each integer $j \in [n^2]$. We let H_j be the set of centers for the minimum covering of $\frac{j}{n^2}K'$ by translates of $\frac{t}{n}B_\infty^n$ and define $H = \cup_{j\in[n^2]}H_j$. Since $|H_j| \le 2^{O(n)}$, it follows that $|H| \le 2^{O(n)}$. For each $X \in B_{S_1}^m$ that satisfies $X \succeq 0$, we let $\frac{j}{n^2}$ be the multiple of $\frac{1}{n^2}$ that is closest to $\mathrm{tr}(X)$, and set $X' := \frac{j}{n^2\mathrm{tr}(X)} \cdot X$. Then we have

$$\|\mathcal{A}(X') - \mathcal{A}(X)\|_\infty \le \frac{1}{n^2} \cdot \|\mathcal{A}(X)\|_\infty \le \frac{t}{n}.$$

As $\mathrm{tr}(X') = \frac{j}{n^2}$, we can also find $Y \in H_j$ with $\|\mathcal{A}(X') - Y\|_\infty \le \frac{t}{n}$. Therefore, $\|\mathcal{A}(X) - Y\|_\infty \le \frac{2t}{n}$, and it follows that $K'' \subseteq H + \frac{2t}{n}B_\infty^n$. This implies $\mathcal{N}(K'', \frac{t}{n}B_\infty^n) \le 2^{O(n)}$.

Next note that the dual discrepancy body $K^\circ := \mathcal{A}(B_{S_1}^m) = \mathcal{A}(B_{S_1}^m \cap \mathbb{S}_m^h)$ since each $A_i \in \mathbb{S}_m^h$. We have $K^\circ = K'' - K''$, so using Theorem 2.4 we get $\mathcal{N}(K^\circ, K'') \le 2^{O(n)}$. Thus

$$\mathcal{N}\left(K^\circ, \frac{t}{n}B_\infty^n\right) \le \mathcal{N}(K^\circ, K'') \cdot \mathcal{N}\left(K'', \frac{t}{n}B_\infty^n\right) \le 2^{O(n)},$$

and $\gamma_n(tK) \ge 2^{-O(n)}$ by using Theorem 3.3. Theorem 3.2 then gives a full coloring $x \in \{\pm 1\}^n$ with discrepancy $\|\sum_{i=1}^n x_i A_i\|_{\mathrm{op}} \le O(t)$. This finishes the proof of the theorem. □

The analysis above also shows that if we can improve the bound in Theorem 3.9 to $O(\log(m/n))$ for any block size h, then the matrix Spencer conjecture is true.

Corollary 4.1 (Better entropy net implies matrix Spencer)**.** *Let $m \ge \sqrt{n}$. If we can find $T \subseteq \mathcal{S}_m$ with $|T| \le 2^{O(n)}$ such that for each $X \in \mathcal{S}_m$ there exists $Y \in T$ with $S(X\|Y) \lesssim \max(1, \log(m/n))$, where $S(X\|Y)$ is the quantum relative entropy between X and Y, then the matrix Spencer conjecture is true.*

5 Matrix discrepancy for Schatten norms

In this section, we prove the following generalization of Theorem 1.3 for arbitrary Schatten norms by using a different regularizer for mirror descent.

Theorem 1.6 (Matrix discrepancy for Schatten norms)**.** *Let $m \ge \sqrt{n}$ and $2 \le p \le q \le \infty$. Given symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_{S_p} \le 1$ and $\mathrm{rank}(A_i) \le r$ for all $i \in [n]$, one can efficiently find $x \in [-1,1]^n$ so that $|\{i : |x_i| = 1\}| \ge n/2$ and*

$$\left\|\sum_{i=1}^n x_i A_i\right\|_{S_q} \lesssim \sqrt{n \cdot \min(p, \max(1, \log(rk)))} \cdot k^{1/p-1/q},$$

where we denote $k := \min(1, m/n)$. Moreover, we can find a full coloring $x \in \{\pm 1\}^n$ at the expense of a factor of $(1/2 + 1/q - 1/p)^{-1}$.

We first use mirror descent to prove the following covering lemma.

Lemma 5.1. *Let $m \geq \sqrt{n}$, $2 \leq p \leq q < \infty$, $k := \min(1, m/n)$, $t := \sqrt{(p-1)n} \cdot k^{1/p-1/q}$ and $q^* := q/(q-1)$. Given symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, we have*

$$\mathcal{N}\Big(\mathcal{A}(B^m_{S_{q^*}}), \frac{t}{n} B^n_\infty\Big) \leq 2^{O(n)}.$$

Proof. Denote $p^* := p/(p-1)$. Theorem 2.6 implies $\mathcal{N}(\mathcal{A}(B^m_{S_{q^*}}), k^{1/q^*-1/p^*}\mathcal{A}(B^m_{S_{p^*}})) \leq 2^{O(n)}$, so it suffices to show

$$\mathcal{N}\Big(\mathcal{A}(B^m_{S_{p^*}}), \sqrt{\frac{p-1}{n}} B^n_\infty\Big) \leq 2^{O(n)}.$$

This is a direct consequence of Theorem 3.6 with norm $\|\cdot\|_{S_{p^*}}$, as the Bregman divergence is $D_\Phi(U, 0) = \Phi(U) \leq \frac{1}{2(p^*-1)} = \frac{p-1}{2}$ for $\|U\|_{S_{p^*}} \leq 1$. □

Theorem 5.1 together with Theorem 3.3 immediately gives the following weaker measure bound, which we then bootstrap to prove the stronger bound in Theorem 1.6.

Corollary 5.2. *Let $m \geq \sqrt{n}$, $2 \leq p \leq q < \infty$ and $k := \min(1, m/n)$. Given symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, define the convex body*

$$K := \Big\{x \in \mathbb{R}^n : \Big\|\sum_{i=1}^n x_i A_i\Big\|_{S_q} \leq 1\Big\}.$$

Then $\gamma_n(\sqrt{(p-1)n} \cdot k^{1/p-1/q} \cdot K) \geq 2^{-O(n)}$.

Proof of Theorem 1.6. Let $p_0 := \max(2, \log(2rk))$. For $p \leq p_0$ the result follows directly from Theorem 5.2, so we may assume $p \geq p_0$. Also note that we may assume $rk \geq 1$ since we can increase smaller values of r without changing the bound on the right side. Remark that $\|A_i\|_{S_{p_0}} \leq r^{1/p_0-1/p}\|A_i\|_{S_p} \leq r^{1/p_0-1/p}$ since the matrices have rank at most r. Theorem 5.2 then implies that the convex body

$$\sqrt{p_0 n} \cdot k^{1/p_0-1/q} \cdot r^{1/p_0-1/p} \cdot K$$

has Gaussian measure $2^{-O(n)}$. Since $\sqrt{p_0 n} \cdot k^{1/p_0-1/q} \cdot r^{1/p_0-1/p} \lesssim \sqrt{p_0 n} \cdot k^{1/p-1/q}$ by the choice of p_0, it follows that

$$\gamma_n(\sqrt{n \max(1, \log(rk))} \cdot k^{1/p-1/q} \cdot K) \geq 2^{-O(n)},$$

so that Theorem 3.1 and Theorem 3.2 yield the partial coloring and full coloring, respectively. The factor $(1/2 + 1/p - 1/q)^{-1}$ comes from the contribution of the exponent of n in the geometric sum, analogous to the second part of Theorem 3.2. □

6 Lower bound examples for matrix discrepancy

In this section, we give a few examples to illustrate the tightness of our results in Theorem 1.6 for various regimes of the dimension m and rank r of the input matrices.

6.1 Low-dimension regime of $m = \Theta(\sqrt{n})$

In the regime of $m = \Theta(\sqrt{n})$, we have $k = \min(1, m/n) = \Theta(1/\sqrt{n})$ and $r \le O(\sqrt{n})$ and our partial coloring bound in Theorem 1.6 is thus $O(n^{1/2+1/2q-1/2p})$. This bound is tight up to constants due to the following example[7].

Lemma 6.1 (Example: $m = \sqrt{n}$)**.** *Let $m = \sqrt{n}$ be a power of 2, and $2 \le p \le q \le \infty$. There exist matrices[8] $A_1, \dots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \le 1$ such that $\|\sum_{i=1}^n x_i A_i\|_{S_q} \gtrsim n^{1/2+1/2q-1/2p}$ for any partial coloring $x \in \{\pm 1\}^n$ with $|\{i : |x_i| = 1\}| \ge n/2$.*

Proof. The idea is to construct an orthogonal basis on $\mathbb{R}^{m \times m}$ with $\|A_i\|_F^2 = m$. Let $H \in \mathbb{R}^{m \times m}$ be the Walsh-Hadamard matrix, and $D_1, \dots, D_m$ be diagonal matrices with $(D_i)_{j,j} := H_{i,j}$. Let $P_1, \dots, P_m$ be disjoint permutation matrices, i.e. each P_i permutes the standard orthonormal basis $\{e_1, \dots, e_m\}$ and each pair P_i, P_j have disjoint non-zero entries. For instance, we may take $(P_i)_{j,k} := 1$ if $j - k \equiv i \mod m$ and 0 otherwise. We then define the n matrices $A_{i+mj} := D_i P_j$ for $i, j \in [m]$. Note that these matrices form an orthogonal basis of $\mathbb{R}^{m \times m}$, so for any partial coloring $x \in \{\pm 1\}^n$ with $|\{i : |x_i| = 1\}| \ge n/2$, we have

$$\left\| \sum_{i=1}^n x_i A_i \right\|_F^2 = \operatorname{tr}\left(\left(\sum_{i=1}^n x_i A_i \right)^\top \right) = m \cdot \sum_{i=1}^n x_i^2 \ge mn/2.$$

By Hölder's inequality, this implies that

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{S_q} \ge m^{1/q-1/2} \cdot \left\| \sum_{i=1}^n x_i A_i \right\|_F \gtrsim n^{1/2+1/2q}.$$

Also note that each matrix A_i has all singular values equal to 1, and therefore $\|A_i\|_{S_p} = m^{1/p} = n^{1/2p}$. Scaling the matrices A_i down by a factor of $n^{1/2p}$ proves the lemma. □

6.2 Rank-1 matrices and $m \ge n$

In the regime of $r = 1$ and $m \ge n$, we may assume wlog that $p = 2$. Then the discrepancy bound in Theorem 1.3 is $O(\sqrt{n})$. This bound is again tight up to a constant factor. A similar construction is found in [41].

[7] Thanks to Aleksandar Nikolov for suggesting this construction.
[8] These matrices can easily be made symmetric in $\mathbb{R}^{2m \times 2m}$.

Lemma 6.2 (Example: $r = 1$ and $m = n$)**.** *Let $2 \le q \le \infty$. There exist symmetric rank-1 matrices $A_1, \dots, A_n \in \mathbb{R}^{n\times n}$ with $\|A_i\|_F \le 1$ such that any partial coloring $x \in [-1,1]^n$ with $|\{i : |x_i| = 1\}| \ge n/2$ satisfies*

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{S_q} \ge \left\| \sum_{i=1}^n x_i A_i \right\|_{\mathrm{op}} \gtrsim \sqrt{n}.$$

Proof. For each $i \in [n-1]$, we define the rank-1 matrices $A_i := \frac{1}{2}(e_i + e_n)(e_i + e_n)^\top$ for $i \in [n]$, where $e_i \in \mathbb{R}^n$ is the unit vector with a single 1 in the ith coordinate and 0 elsewhere, and $A_n = 0$. Note that each $\|A_i\|_F = 1$ by definition. For any partial coloring $x \in [-1,1]^n$ with $|\{i : |x_i| = 1\}| \ge n/2$, we have

$$\sum_{i=1}^n x_i A_i = \frac{1}{2} \cdot \begin{pmatrix} x_1 & 0 & \cdots & 0 & x_1 \\ 0 & x_2 & \cdots & 0 & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x_{n-1} & x_{n-1} \\ x_1 & x_2 & \cdots & x_{n-1} & \sum_{i=1}^{n-1} x_i \end{pmatrix}.$$

It then follows that

$$\left\| \sum_{i=1}^n x_i A_i \right\|_{\mathrm{op}} \ge \left\| \sum_{i=1}^n x_i A_i e_n \right\|_2 \gtrsim \sqrt{n}.$$

This completes the proof of the lemma. □

As an immediate corollary of Theorem 6.2, we obtain an $\Omega(\sqrt{n})$ lower bound for matrix Spencer when $m = n$ and all matrices are rank-1.

Corollary 6.3 (Rank-1 matrix Spencer lower bound)**.** *There exist rank-1 symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{n\times n}$ with $\|A_i\|_{\mathrm{op}} \le 1$ such that any $x \in \{\pm 1\}^n$ has $\| \sum_{i=1}^n x_i A_i\|_{\mathrm{op}} \gtrsim \sqrt{n}$.*

Another immediate consequence of Theorem 6.2 is a lower bound of $\Omega(\sqrt{\min(m,n)})$ for Schatten-2 to operator norm discrepancy, which is the generalization of the Komlós problem to matrices. This shows that the Komlós conjecture, which states that the ℓ_2 to ℓ_∞ vector discrepancy is upper bounded by a universal constant, cannot be true for matrices.

Corollary 6.4 (Lower bound for matrix Komlós)**.** *For any m and n, there exist symmetric matrices $A_1, \dots, A_n \in \mathbb{R}^{m\times m}$ with $\|A_i\|_F \le 1$ such that any $x \in \{\pm 1\}^n$ has $\| \sum_{i=1}^n x_i A_i\|_{\mathrm{op}} \gtrsim \sqrt{\min(m,n)}$.*

A An application of Banaszczyk's theorem

We give an alternative simpler proof of the $O(m^{1+1/q-1/p})$ bound for S_p to S_q matrix discrepancy when $m = O(\sqrt{n})$ using the following theorem of Banaszczyk [4].

Theorem A.1 (Banaszczyk [4])**.** *Let $K \subseteq \mathbb{R}^m$ be a convex body with $\gamma_m(K) \geq 1/2$. Then for any vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ with $\|v_i\|_2 \leq 1$, there exists $x \in \{\pm 1\}^n$ such that $\sum_{i=1}^n x_i v_i \in 5K$.*

Applying Theorem A.1 to a suitable scaling of the operator norm ball immediately gives the following matrix discrepancy bound.

Corollary A.2. *Let $2 \leq p \leq q \leq \infty$. Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_{S_p} \leq 1$, there exists $x \in \{\pm 1\}^n$ such that $\| \sum_{i=1}^n x_i A_i \|_{S_q} \lesssim m^{1+1/q-1/p}$.*

Proof. Note that $\|A_i\|_{S_p} \leq 1$ implies $\|A_i\|_{S_2} \leq m^{1/2-1/p}$. It is well-known that $\gamma_m(4m^{1/2} \cdot B_{\mathrm{op}}^m) \geq 1/2$ (see Theorem 7.3.1 of [40]). Thus, Theorem A.1 yields some $x \in \{\pm 1\}^n$ such that $\sum_{i=1}^n x_i A_i \in O(m^{1-1/p}) \cdot B_{\mathrm{op}}^m$. It follows that $\| \sum_{i=1}^n x_i A_i \|_{S_q} \leq O(m^{1+1/q-1/p})$. □

Corollary A.3 (Matrix Komlós)**.** *Given matrices $A_1, \ldots, A_n \in \mathbb{R}^{m \times m}$ with $\|A_i\|_F \leq 1$, there exists $x \in \{\pm 1\}^n$ such that $\| \sum_{i=1}^n x_i A_i \|_{S_q} \lesssim \sqrt{\min(m, n)}$, matching the lower bound in Theorem 1.8.*

Proof. It suffices to take the best between a random coloring, which has discrepancy $O(\sqrt{n})$, and that of Theorem A.2. □

Acknowledgements

We thank Aleksandar Nikolov for insightful discussions and for suggesting the lower bound example when $m = \sqrt{n}$. We thank Thomas Rothvoss, Sivakanth Gopi and Mehtaab Sawhney for helpful discussions.

References

[1] RUDOLF AHLSWEDE AND ANDREAS WINTER: Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, 2002. [doi:10.1109/18.985947, arXiv:quant-ph/0012127] 8

[2] SHIRI ARTSTEIN-AVIDAN, APOSTOLOS GIANNOPOULOS, AND VITALI D. MILMAN: *Asymptotic geometric analysis, Part I*. Amer. Math. Soc., 2015. [doi:10.1090/surv/202] 9

[3] KEITH BALL, ERIC A. CARLEN, AND ELLIOTT H. LIEB: Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones Math.*, 115(1):463–482, 1994. [doi:10.1007/BF01231769] 12

[4] WOJCIECH BANASZCZYK: Balancing vectors and Gaussian measures of n-dimensional convex bodies. *Random Struct. Algor.*, 12(4):351–360, 1998. [doi:10.1002/(SICI)1098-2418(199807)12:4<351::AID-RSA3>3.0.CO;2-S] 5, 8, 19, 20

[5] AFONSO S. BANDEIRA, MARCH T. BOEDIHARDJO, AND RAMON VAN HANDEL: Matrix concentration inequalities and free probability. *Inventiones Math.*, 234(1):419–487, 2023. [doi:10.1007/s00222-023-01204-6, arXiv:2108.06312] 8

[6] NIKHIL BANSAL: Constructive algorithms for discrepancy minimization. In *Proc. 51st FOCS*, pp. 3–10. IEEE Comp. Soc., 2010. [doi:10.1109/FOCS.2010.7] 2, 3

[7] NIKHIL BANSAL, MOSES CHARIKAR, RAVISHANKAR KRISHNASWAMY, AND SHI LI: Better algorithms and hardness for broadcast scheduling via a discrepancy approach. In *Proc. 25th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'14)*, pp. 55–71. SIAM, 2014. [doi:10.1137/1.9781611973402.5] 2

[8] NIKHIL BANSAL, DANIEL DADUSH, AND SHASHWAT GARG: An algorithm for Komlós conjecture matching Banaszczyk's bound. *SIAM J. Comput.*, 48(2):534–553, 2019. Preliminary version in FOCS'16. [doi:10.1137/17M1126795, arXiv:1605.02882] 8

[9] NIKHIL BANSAL, DANIEL DADUSH, SHASHWAT GARG, AND SHACHAR LOVETT: The Gram–Schmidt walk: A cure for the Banaszczyk blues. *Theory of Computing*, 15(21):1–27, 2019. Preliminary version in STOC'18. [doi:10.4086/toc.2019.v015a021] 8

[10] NIKHIL BANSAL AND SHASHWAT GARG: Algorithmic discrepancy beyond partial coloring. In *Proc. 49th STOC*, pp. 914–926. ACM Press, 2017. [doi:10.1145/3055399.3055490] 8

[11] NIKHIL BANSAL, HAOTIAN JIANG, AND RAGHU MEKA: Resolving matrix Spencer conjecture up to poly-logarithmic rank. *SIAM J. Comput.*, pp. STOC23:57–67, 2024. Preliminary version in STOC'23. [doi:10.1137/23M1592201, arXiv:2208.11286] 3

[12] NIKHIL BANSAL AND VISWANATH NAGARAJAN: Approximation-friendly discrepancy rounding. In MARTIN LOEBL, JAROSLAV NEŠETŘIL, AND ROBIN THOMAS, editors, *A Journey Through Discrete Mathematics*, pp. 89–114. Springer, 2017. Preliminary version in IPCO'16. [doi:10.1007/978-3-319-44479-6_4, arXiv:1512.02254] 2

[13] JÓZSEF BECK: Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981. [doi:10.1007/BF02579452] 2

[14] SÉBASTIEN BUBECK: Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learning*, 8(3–4):231–357, 2015. [doi:10.1561/2200000050, arXiv:1405.4980] 12

[15] ERIC A. CARLEN: A remainder term for Hölder's inequality for matrices and quantum entropy inequalities. *Archiv der Mathematik*, 109(4):365–371, 2017. [doi:10.1007/s00013-017-1066-8, arXiv:1605.04581] 12

[16] BERNARD CHAZELLE: *The Discrepancy Method: Randomness and Complexity*. Cambridge Univ. Press, 2000. [doi:10.1017/CBO9780511626371] 2

[17] DANIEL DADUSH, SHASHWAT GARG, SHACHAR LOVETT, AND ALEKSANDAR NIKOLOV: Towards a constructive version of Banaszczyk's vector balancing theorem. *Theory of Computing*, 15(15):1–58, 2019. Preliminary version in RANDOM'16. [doi:10.4086/toc.2019.v015a015] 8

[18] DANIEL DADUSH, HAOTIAN JIANG, AND VICTOR REIS: A new framework for matrix discrepancy: Partial coloring bounds via mirror descent. In *Proc. 54th STOC*, pp. 649–658. ACM Press, 2022. [doi:10.1145/3519935.3519967] 1

[19] DANIEL DADUSH, ALEKSANDAR NIKOLOV, KUNAL TALWAR, AND NICOLE TOMCZAK-JAEGERMANN: Balancing vectors in any norm. In *Proc. 59th FOCS*, pp. 1–10. IEEE Comp. Soc., 2018. [doi:10.1109/FOCS.2018.00010] 3

[20] RONEN ELDAN AND MOHIT SINGH: Efficient algorithms for discrepancy minimization in convex sets. *Random Struct. Algor.*, 53(2):289–307, 2018. [doi:10.1002/rsa.20763, arXiv:1409.2913] 2, 3

[21] EFIM DAVYDOVICH GLUSKIN: Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. *Sbornik Math.*, 64(1):85–96, 1989. [doi:10.1070/SM1989v064n01ABEH003295] 2, 3, 6

[22] AICKE HINRICHS, JOSCHA PROCHNO, AND JAN VYBÍRAL: Entropy numbers of embeddings of Schatten classes. *J. Functional Anal.*, 273(10):3241–3261, 2017. [doi:10.1016/j.jfa.2017.08.008, arXiv:1612.08105] 7, 10

[23] SAMUEL B. HOPKINS, PRASAD RAGHAVENDRA, AND ABHISHEK SHETTY: Matrix discrepancy from quantum communication. In *Proc. 54th STOC*, pp. 637–648. ACM Press, 2022. [doi:10.1145/3519935.3519954] 8

[24] HERMANN KÖNIG AND VITALI D. MILMAN: On the covering numbers of convex bodies. In JORAM LINDENSTRAUSS AND VITALI D. MILMAN, editors, *Geometrical Aspects of Functional Analysis*, pp. 82–95. Springer, 1987. [doi:10.1007/BFb0078138] 9

[25] AVI LEVY, HARISHCHANDRA RAMADAS, AND THOMAS ROTHVOSS: Deterministic discrepancy minimization via the multiplicative weight update method. In *Proc. 19th Integer Prog. Combinat. Optim. (IPCO'17)*, volume 10328 of *LNCS*, pp. 380–391. Springer, 2017. [doi:10.1007/978-3-319-59250-3_31, arXiv:1611.08752] 2, 3, 4, 8

[26] SHACHAR LOVETT AND RAGHU MEKA: Constructive discrepancy minimization by walking on the edges. *SIAM J. Comput.*, 44(5):1573–1582, 2015. [doi:10.1137/130929400, arXiv:1203.5747] 2, 3

[27] FRANÇOISE LUST-PIQUARD AND GILLES PISIER: Non commutative Khintchine and Paley inequalities. *Arkiv för Mat.*, 29(1–2):241–260, 1991. [doi:10.1007/BF02384340] 3, 8

[28] JIŘÍ MATOUŠEK: *Geometric Discrepancy: An Illustrated Guide*. Volume 18 of *Algorithms and Combinatorics*. Springer, 1999. [doi:10.1007/978-3-642-03942-3] 2

[29] MARK W. MECKES: On the spectral norm of a random Toeplitz matrix. *Electr. Comm. Probability*, 12:315–325, 2007. [doi:10.1214/ECP.v12-1313] 8

[30] RAGHU MEKA: Discrepancy and beating the union bound, 2014. Windows on Theory. 3

[31] ARKADI SEMENOVICH NEMIROVSKI AND DAVID BORISOVICH YUDIN: *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983. Link at GaTech. 4, 6, 12

[32] ALEKSANDAR NIKOLOV, KUNAL TALWAR, AND LI ZHANG: The geometry of differential privacy: The small database and approximate cases. *SIAM J. Comput.*, 45(2):575–616, 2016. Preliminary version in STOC'13. [doi:10.1137/130938943] 2

[33] GILLES PISIER: *Introduction to Operator Space Theory*. Volume 294 of *LMS Lecture Notes*. Cambridge Univ. Press, 2003. [doi:10.1017/CBO9781107360235] 3, 8

[34] VICTOR REIS AND THOMAS ROTHVOSS: Vector balancing in Lebesgue spaces. *Random Struct. Algor.*, 62(3):667–688, 2023. [doi:10.1002/rsa.21113, arXiv:2007.05634] 2, 3, 5, 6, 10

[35] THOMAS ROTHVOSS: Approximating bin packing within $O(\log OPT \cdot \log\log OPT)$ bins. In *Proc. 54th FOCS*, pp. 20–29. IEEE Comp. Soc., 2013. [doi:10.1109/FOCS.2013.11] 2

[36] THOMAS ROTHVOSS: Constructive discrepancy minimization for convex sets. *SIAM J. Comput.*, 46(1):224–234, 2017. [doi:10.1137/141000282] 2, 3

[37] THOMAS ROYEN: A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East J. Theor. Stat.*, 48(2):139–145, 2014. Available at publisher's website. [arXiv:1408.1028] 10

[38] ZBYNĚK ŠIDÁK: Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statistical Assoc.*, 62(318):626–633, 1967. [doi:10.1080/01621459.1967.10482935] 3

[39] JOEL SPENCER: Six standard deviations suffice. *Trans. AMS*, 289(2):679–706, 1985. [doi:10.1090/S0002-9947-1985-0784009-0] 2

[40] ROMAN VERSHYNIN: *High-Dimensional Probability: An Introduction with Applications in Data Science*. Volume 47 of *Cambr. Ser. in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, 2018. [doi:10.1017/9781108231596] 7, 8, 20

[41] NIK WEAVER: The Kadison–Singer problem in discrepancy theory. *Discr. Math.*, 278(1–3):227–239, 2004. [doi:10.1016/S0012-365X(03)00253-X] 18

[42] ANASTASIOS ZOUZIAS: A matrix hyperbolic cosine algorithm and applications. In *Proc. 39th Internat. Colloq. on Automata, Languages, and Programming (ICALP'12)*, pp. 846–858. Springer, 2012. [doi:10.1007/978-3-642-31594-7_71, arXiv:1103.2793] 3

AUTHORS

Daniel Dadush
Researcher
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
dadush@cwi.nl
https://homepages.cwi.nl/~dadush/

Haotian Jiang
Postdoctoral researcher
Microsoft Research
Redmond, WA, USA
jhtdavid96@gmail.com
https://jhtdavid96.wixsite.com/jianghaotian

Victor Reis
Postdoctoral researcher
Institute for Advanced Study
Princeton, NJ, USA
voreis@ias.edu
https://homes.cs.washington.edu/~voreis/

ABOUT THE AUTHORS

DANIEL DADUSH is a tenured researcher at the Centrum Wiskunde & Informatica (CWI) in Amsterdam, Netherlands. He earned his Ph. D. in algorithms, combinatorics and optimization (ACO) from Georgia Tech in 2012, where his advisor was Santosh Vempala. Before joining CWI, he spent two years as a Simons postdoctoral fellow in the Computer Science Department at New York University. His research has focused on algorithms for lattice problems, integer programming, and high-dimensional convex geometry.

He lives and works in Amsterdam, and is glad that, thus far, the city has remained above water. He enjoys reading the New York Times, listening to NPR, and taking long bike rides along the canals when it is not raining.

HAOTIAN JIANG is a Postdoctoral Researcher in the Algorithms Group at Microsoft Research, Remond. He will be joining the Computer Science Department at the University of Chicago in Summer 2024. He completed his Ph. D. at the University of Washington where he worked in the theory group in the Paul G. Allen School of Computer Science & Engineering, advised by Yin Tat Lee. His research has focused on convex optimization and its applications, submodular function minimization, and discrepancy theory.

VICTOR REIS is a Postdoctoral Researcher in the Computer Science and Discrete Mathematics Group at the Institute for Advanced Study, Princeton. He completed his Ph. D. at the University of Washington where he worked in the theory group in the Paul G. Allen School of Computer Science & Engineering, advised by Thomas Rothvoss. His research has focused on discrepancy theory, graph sparsification and integer programming.