On the number of Hadamard matrices via anti-concentration

Asaf Ferber * Vishesh Jain† Yufei Zhao ‡

Abstract

Many problems in combinatorial linear algebra require upper bounds on the number of solutions to an underdetermined system of linear equations $Ax = b$, where the coordinates of the vector x are restricted to take values in some small subset (e.g. $\{\pm 1\}$) of the underlying field. The classical ways of bounding this quantity are to use either a rank bound observation due to Odlyzko or a vector anti-concentration inequality due to Halász. The former gives a stronger conclusion except when the number of equations is significantly smaller than the number of variables; even in such situations, the hypotheses of Halász's inequality are quite hard to verify in practice. In this paper, using a novel approach to the anti-concentration problem for vector sums, we obtain new Halász-type inequalities which beat the Odlyzko bound even in settings where the number of equations is comparable to the number of variables. In addition to being stronger, our inequalities have hypotheses which are considerably easier to verify. We present two applications of our inequalities to combinatorial (random) matrix theory: (i) we obtain the first non-trivial upper bound on the number of $n \times n$ Hadamard matrices, and (ii) we improve a recent bound of Deneanu and Vu on the probability of normality of a random $\{\pm 1\}$ matrix.

1 Introduction

1.1 The number of Hadamard matrices

A square matrix H of order n whose entries are $\{\pm 1\}$ is called a *Hadamard matrix of order* n if its rows are pairwise orthogonal i.e. if $HH^T = nI_n$. They are named after Jacques Hadamard, who studied them in connection with his maximal determinant problem. Specifically, Hadamard asked for the maximum value of the determinant of any $n \times n$ square matrix all of whose entries are bounded in absolute value by 1. He proved [8] that the value of the determinant of such matrices cannot exceed $n^{n/2}$. Moreover, he showed that Hadamard matrices are the only ones that can attain this bound. Since their introduction, Hadamard matrices have been the focus of considerable attention from many different communities – coding theory, design theory, statistical inference, and signal processing to name a few. We refer the reader to the surveys [10, 19] and the books [1, 11] for a comprehensive account of Hadamard matrices and their numerous applications.

Hadamard matrices of order 1 and 2 are trivial to construct, and it is quite easy to see, by considering the first few rows, that every other Hadamard matrix (if exists) must be of order $4m$ for some $m \in \mathbb{N}$. Whereas Hadamard matrices of infinitely many orders have been constructed, the question of whether one of order $4m$ exists for every $m \in \mathbb{N}$ is the most important open question on this topic, and remains wide open.

*Massachusetts Institute of Technology. Department of Mathematics. Email: ferbera@mit.edu. Research is partially supported by NSF DMS-6935855.

†Massachusetts Institute of Technology. Department of Mathematics. Email: visheshj@mit.edu. Research is partially supported by NSF CCF 1665252 and DMS-1737944 and ONR N00014-17-1-2598.

‡Massachusetts Institute of Technology. Department of Mathematics. Email: yufeiz@mit.edu. Research is partially supported by NSF DMS-1362326 and DMS-1764176.

Conjecture 1.1 (The Hadamard conjecture, [15])**.** *There exists a Hadamard matrix of order $4m$ for every $m \in \mathbb{N}$.*

In this paper, we study the question of *how many* Hadamard matrices of order $n = 4m$ could possibly exist for a given $m \in \mathbb{N}$. Let us denote this number by $H(n)$. Note that if a single Hadamard matrix of order n exists, then we immediately get at least $(n!)^2$ distinct Hadamard matrices by permuting all the rows and columns. Thus, if the Hadamard conjecture is true, then $H(n) = 2^{\Omega(n \log n)}$ for every $n = 4m, m \in \mathbb{N}$. On the other hand, the bound $H(n) \leq 2^{\binom{n+1}{2}}$ is quite easy to obtain, as we will discuss in the next subsection.

This bound also appeared in the work of de Launey and Levin [3] on the enumeration of partial Hadamard matrices (i.e. $k \times 4m$ matrices whose rows are pairwise orthogonal, in the limit as $m \to \infty$) using Fourier analytic techniques; notably, while they were able to get a very precise answer to this problem (up to an overall $(1+o(1))$ multiplicative factor), their techniques still did not help them to obtain anything better than the essentially trivial bound for the case of square Hadamard matrices. As our first main result, we give the only known non-trivial upper bound on the number of square Hadamard matrices.

Theorem 1.2. *There exists an absolute constant $c_H > 0$ such that $H(n) \leq 2^{\frac{(1-c_H)n^2}{2}}$ for all sufficiently large n that is a multiple of 4.*

Remark 1.3. In our proof of the above theorem, we have focused on the simplicity and clarity of presentation and have made no attempt to optimize this constant, since our proof cannot give a value of c_H larger than (say) $\frac{1}{2}$ whereas we believe that the correct value of c_H should be close (as a function of n) to 1.

Conjecture 1.4. *For any $n = 4m, m \in \mathbb{N}$, $H(n) = 2^{O(n \log n)}$.*

We believe that proving a bound of the form $H(n) = 2^{o(n^2)}$ will already be very interesting, and will likely require new ideas.

1.1.1 The approach

We now discuss the proof of the trivial upper bound $H(n) \leq 2^{\binom{n+1}{2}}$. The starting point is the following classical (and almost trivial to prove) observation due to Odlyzko.

Lemma 1.5 (Odlyzko, [14])**.** *Let W be a d-dimensional subspace of $\mathbb{R}^n$. Then, $|W \cap \{\pm 1\}^n| \leq 2^d$.*

Sketch. As W is a d-dimensional space, it depends only on d coordinates. Therefore, it spans at most 2^d vectors with entries from $\{\pm 1\}$. □

The bound $H(n) \leq 2^{\binom{n+1}{2}}$ is now immediate. Indeed, we construct the matrices row by row, and note that by the orthogonality of the rows, the first k rows span a subspace of dimension k to which the remaining rows are orthogonal. In particular, once the first k rows have been selected, the $(k+1)^{st}$ row lies in a specified subspace of dimension $n-k$ (the orthogonal complement of the vector space spanned by the first k (linearly independent) rows), and hence, by Lemma 1.5, is one of at most 2^{n-k} vectors. It follows that $H(n) \leq \prod_{i=0}^{n-1} 2^{n-i} = 2^{\binom{n+1}{2}}$.

The weak point in the above proof is the following – while Odlyzko's bound is tight in general, we should expect it to be far from the truth in the average case. Indeed, working with vectors in $\{0,1\}^n$ for the moment, note that a subspace of dimension k spanned by vectors in $\{0,1\}^n$ has exactly 2^{n-k} vectors in $\{0,1\}^n$ orthogonal to it *viewed as elements of* $\mathbb{F}_2^n$. However, typically, the inner products

will take on many values in $2\mathbb{Z} \setminus \{0\}$ so that many of these vectors will not be orthogonal viewed as elements of $\mathbb{R}^n$.

The study of the difference between the Odlyzko bound and how many $\{\pm 1\}^n$ vectors a subspace actually contains has been very fruitful in discrete random matrix theory, particularly for the outstanding problem of determining the probability of singularity of random $\{\pm 1\}$ matrices. Following Kahn, Komlós and Szemerédi [13], Tao and Vu [20] isolated the following notion.

Definition 1.6 (Combinatorial dimension)**.** The *combinatorial dimension* of a subspace W in $\mathbb{R}^n$, denoted by $d_\pm(W)$, is defined to be smallest real number such that

$$|W \cap \{\pm 1\}^n| \leq 2^{d_\pm(W)}.$$

Thus, Odlyzko's lemma says that for any subspace W, its combinatorial dimension is no more than its dimension. However, improving on another result of Odlyzko [14], Kahn, Komlós and Szemerédi showed that this bound is very loose for *typical subspaces spanned by* $\{\pm 1\}^n$ *vectors*:

Theorem 1.7 (Kahn-Komlós-Szemerédi, [13])**.** *There exists a constant $C > 0$ such that if $r \leq n - C$, and if $v_1, \ldots, v_r$ are chosen independently and uniformly from $\{\pm 1\}^n$, then*

$$\Pr\left[d_\pm(\mathrm{span}\{v_1, \ldots, v_r\}) > \log_2(2r)\right] = (1 + o(1))4\binom{r}{3}\left(\frac{3}{4}\right)^n.$$

In other words, they showed that a typical r-dimensional subspace spanned by r vectors in $\{\pm 1\}^n$ contains the minimum possible number of $\{\pm 1\}^n$ vectors i.e. only the $2r$ vectors consisting of the vectors spanning the subspace and their negatives.

Compared to the setting of Kahn, Komlós and Szemerédi, our setting has two major differences:

(i) We are interested not in the combinatorial dimension of subspaces spanned by $\{\pm 1\}^n$ vectors but of their *orthogonal complements*.

(ii) The $\{\pm 1\}^n$ vectors spanning a subspace in our case are *highly dependent* due to the mutual orthogonality constraint – indeed, as the proof of the trivial upper bound at the start of the subsection shows, the probability that the rows of a random $k \times n$ $\{\pm 1\}$ matrix are mutually orthogonal is $2^{-\Omega(k^2)}$; this rules out the strategy of conditioning on the rows being orthogonal when $k = \Omega(\sqrt{n})$, even if one were to prove a variant of the result of Kahn, Komlós and Szemerédi to deal with orthogonal complements.

Briefly, our approach to dealing with these obstacles is the following. For $k < n$, let $H_{k,n}$ denote a $k \times n$ matrix with all its entries in $\{\pm 1\}$ and all of whose rows are orthogonal. We will show that there exist absolute constants $0 < c_1 < c_2 < 1$ such that if $k \in [c_1 n, c_2 n]$ and if n is sufficiently large, then $H_{k,n}$ must have a certain desirable linear algebraic property; this is the only way in which we use the orthogonality of the rows of $H_{k,n}$, and takes care of (ii). Next, to deal with (i), we will show that for any $k \times n$ matrix A which has this linear algebraic structure, the number of solutions x in $\{\pm 1\}^n$ to $Ax = 0$ is at most $2^{n-(1+C)k}$, where $C > 0$ is a constant depending only on c_1 and c_2. Using these improved bounds with the same strategy as for the trivial proof, we see that for n sufficiently large,

$$\begin{aligned} H(n) &\leq \prod_{i=0}^{n-1} 2^{n-i} \prod_{i=c_1 n}^{c_2 n} 2^{-Ck} \\ &\leq 2^{\binom{n+1}{2}} 2^{-\frac{C(c_2^2 - c_1^2)n^2}{2}}, \end{aligned}$$

which gives the desired improvement. We discuss this in more detail in the next subsection.

1.2 Improved Halász-type inequalities

As mentioned above, our goal is to study the number of $\{\pm 1\}^n$ solutions to an underdetermined system of linear equations $Ax = 0$ possessing some additional structure. This question was studied by Halász, who proved the following:

Theorem 1.8 (Halász, [9]). *Let $a_1, \dots, a_n$ be a collection of vectors in $\mathbb{R}^d$. Suppose there exists a constant $\delta > 0$ such that for any unit vector $e \in \mathbb{R}^d$, one can select at least δn vectors a_k with $|\langle a_k, e\rangle| \geq 1$. Then,*

$$\sup_{u \in \mathbb{R}^d} \Pr\left[\left\|\sum_{i=1}^n \epsilon_i a_i - u\right\|_2 < 1\right] \leq c(\delta, d)\left(\frac{1}{\sqrt{n}}\right)^d,$$

where $\epsilon_1, \dots, \epsilon_n$ are independent Rademacher random variables i.e. they take the values ± 1 with probability $1/2$ each.

The constant $c(\delta, d)$, which is crucial for our applications, was left implicit by Halász. However, explicit estimates on this constant may be obtained, as was done by Howard and Oskolkov [12].

Theorem 1.9 ([12]). *Let $a_1, \dots, a_n$ be a collection of vectors in $\mathbb{R}^d$. Suppose that there exists some $m \in \mathbb{N}$ such that for every unit vector $e \in \mathbb{R}^d$, one can select at least m vectors $a_{i_1}, \dots, a_{i_m}$ with $|\langle a_{i_j}, e\rangle| \geq 1/(2\sqrt{d})$ for all $j \in [m]$. Then,*

$$\sup_{u \in \mathbb{R}^d} \Pr\left[\left\|\sum_{i=1}^n \epsilon_i a_i - u\right\|_\infty < \frac{1}{2}\right] \leq C(d)\left(\frac{1}{\sqrt{m}}\right)^d,$$

where $C(d) = \left(\frac{\pi^{3/2} d}{\sqrt{2}}\right)^d$.

Remark 1.10. When $a_1, \dots, a_n$ and u belong to $\mathbb{Z}^d$, as will be the case in our applications, the event '$\|\sum_{i=1}^n \epsilon_i a_i - u\|_\infty < 1/2$' is equivalent to the event '$\sum_{i=1}^n \epsilon_i a_i = u$'. In this case, it was noted by Tao and Vu (Exercise 7.2.3 in [21]) that the condition $|\langle a_{i_j}, e\rangle| \geq 1/(2\sqrt{d})$ may be relaxed to $|\langle a_{i_j}, e\rangle| > 0$. However, as stated, their proof still gives a constant $C(d) = \Theta(d)^d$ due to a 'duplication' step, which we will show is unnecessary.

There are two drawbacks to using the results mentioned above for the kinds of applications we have in mind. Firstly, a constant of the form $C(d) = \Theta(d)^d$ does not give any non-trivial information when $d = \Omega(n)$, whereas as discussed in the proof outline, we require an improvement over the Odlyzko bound for $d = \Theta(n)$. Secondly, the hypotheses of these theorems, which involve two quantifiers ('for all' followed by 'there exists'), are quite stringent and not easy to verify; in fact, we were unable to find any direct applications of Theorem 1.8 in the literature.

Our key structural observation is that *a 'pseudorandom' rectangular matrix contains many disjoint submatrices of large rank*. This motivates replacing the double quantifier hypothesis by a (weaker) hypothesis involving just one existential quantifier which, as we will see, is readily verified to hold in pseudorandom situations. Moreover, while our hypothesis is weaker, we are able to obtain conclusions with asymptotically much better constants, since our structural setting allows us to efficiently leverage the existing rich literature on anti-concentration of sums of independent random variables and anti-concentration of linear images of high dimensional distributions. In particular, we are able to give short and transparent proofs of our inequalities for very general classes of distributions; in contrast, Theorems 1.8 and 1.9 hold only for (vector)-weighted sums of independent Rademacher variables, and their proofs involve explicit trigonometric manipulations. We discuss this in more detail in Sections 2.2 and 3.1.

Our first inequality is a strengthening of Theorem 1.9 in the setting of Remark 1.10, both in terms of the hypothesis and the conclusion. A more general statement appears in Theorem 3.1.

Theorem 1.11. *Let $a_1, \dots, a_n$ be a collection vectors in $\mathbb{R}^d$ which can be partitioned as $\mathcal{A}_1, \dots, \mathcal{A}_\ell$ with ℓ even such that $\dim_{\mathbb{R}^d}(\mathrm{span}\{a : a \in \mathcal{A}_i\}) =: r_i$. Then,*

$$\sup_{u \in \mathbb{R}^d} \Pr\left[\sum_{i=1}^n \epsilon_i a_i = u\right] \leq \left(2^{-\ell}\binom{\ell}{\ell/2}\right)^{\frac{r_1+\cdots+r_\ell}{\ell}} \leq \left(\sqrt{\frac{2}{\pi\ell}}\left(1+O\left(\frac{1}{\ell}\right)\right)\right)^{\frac{r_1+\cdots+r_\ell}{\ell}}.$$

Remark 1.12. This inequality is tight, as can be easily seen by taking (assuming n is divisible by d) a_i to be $e_{i \bmod d}$, where $e_1, \dots, e_d$ denotes the standard basis of $\mathbb{R}^d$, in which case we can take $\ell = n/d$ and $r_1 = \cdots = r_\ell = d$.

To see how Theorem 1.11 strengthens Theorem 1.9, note that the assumptions of Theorem 1.9 guarantee that there exist $\ell := \lfloor m/d \rfloor$ disjoint subsets $\mathcal{A}_1, \dots, \mathcal{A}_\ell$ such that $r_1 = \cdots = r_\ell = d$. Such a collection of disjoint subsets can be obtained greedily by repeating the following construction ℓ times: let $v_1 \in \{a_1, \dots, a_n\}$ be any nonzero vector that has not already been chosen in a previous iteration. Having chosen $v_1, \dots, v_s$ for $s < d$, let $u_s \in (\mathrm{span}\{v_1, \dots, v_s\})^\perp$, and let v_{s+1} be any vector satisfying $|\langle v_{s+1}, u_s\rangle| > 0$ which has not already been chosen in a previous iteration – such a vector is guaranteed to exist since there are at least m choices of v_{s+1} by assumption, of which at most $(\ell-1)d < m$ could have been chosen in a previous iteration. It follows that under the assumptions of Theorem 1.9, when $a_1, \dots, a_n \in \mathbb{Z}^d$, we have:

$$\sup_{u \in \mathbb{R}^d} \Pr\left[\left\|\sum_{i=1}^n \epsilon_i a_i - u\right\|_\infty < \frac{1}{2}\right] \leq C'(d)\left(\frac{1}{\sqrt{m}}\right)^d,$$

where $C'(d) \leq \left(\sqrt{\frac{2d}{3}}\right)^d$. In particular, we now have a non-trivial bound all the way up to $d = \Theta(m)$, as opposed to just $d = O(\sqrt{m})$ as before.

Our second inequality is a 'small-ball probability' version of Theorem 1.11. In order to state it, we need the following definition.

Definition 1.13. The *stable rank* of A, denoted by $r_s(A)$, is defined as

$$r_s(A) := \left\lfloor \frac{\|A\|_{\mathrm{HS}}^2}{\|A\|^2} \right\rfloor,$$

where $\|A\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm of A, and $\|A\|$ denotes the operator norm of A.

Remark 1.14. Recall that $\|A\| = s_1(A)$ and $\|A\|_{\mathrm{HS}}^2 = \sum_{i=1}^{\mathrm{rank}(A)} s_i(A)^2$, where $s_1(A), s_2(A), \dots$ denote the singular values of A arranged in non-increasing order. Hence,

$$r_s(A) := \left\lfloor \frac{\sum_{i=1}^{\mathrm{rank}(A)} s_i(A)^2}{s_1(A)^2} \right\rfloor;$$

in particular, for any non-zero matrix A, $1 \leq r_s(A) \leq \mathrm{rank}(A)$, with the right inequality being an equality if and only if A is an orthogonal projection up to an isometry.

We can now state our inequality. A more general version appears in Theorem 3.2.

Theorem 1.15. *Let $a_1, \dots, a_n$ be a collection vectors in $\mathbb{R}^d$. For some $\ell \in 2\mathbb{N}$, let $\mathcal{A}_1, \dots, \mathcal{A}_\ell$ be a partition of the set $\{a_1, \dots, a_n\}$, and for each $i \in [\ell]$, let A_i denote the $d \times |\mathcal{A}_i|$ dimensional matrix whose columns are given by the elements of $\mathcal{A}_i$. Then, for every $M \geq 1$ and $\varepsilon \in (0,1)$,*

$$\Pr\left[\left\|\sum_{i=1}^n \epsilon_i a_i - u\right\|_2 \leq M\right] \leq 2^d \prod_{i=1}^{\ell}\left(\frac{CM}{\sqrt{\varepsilon\ell}\|A_i\|_{HS}}\right)^{\frac{\lceil(1-\varepsilon)r_s(A_i)\rceil}{\ell}},$$

where $r_s(A_i)$ denotes the stable rank of A_i and C is an absolute constant.

For illustration, consider a situation like above where the set of vectors $a_1, \ldots, a_n$ can be partitioned into m/d subsets, each of rank d. Assume further that each a_i has norm at least one, so that each of the m/d matrices has Hilbert-Schmidt norm at least $\sqrt{d}$. Then, if the stable rank of each of these matrices is at least δd for some $\delta > 0$, it follows that

$$\Pr\left[\left\|\sum_{i=1}^{n} \epsilon_i a_i - u\right\|_2 \leq 1\right] \leq K^d,$$

where $K \leq 2\,(C/\sqrt{m})^{\delta/2}$, which is a big improvement over the bound coming from Theorem 1.2 provided that δ is not too small and d is large.

1.3 Counting $\{\pm 1\}$-valued normal matrices

Recall that a matrix M is *normal* if it commutes with its adjoint, i.e., $MM^* = M^*M$ (for real matrices this is the same as $MM^T = M^TM$). Recently, Deneanu and Vu [4] studied the number of $n \times n$ $\{\pm 1\}$-valued normal matrices. Since real symmetric matrices are normal, there are at least $2^{\binom{n+1}{2}}$ $\{\pm 1\}$-valued normal matrices. They conjectured that this lower bound is essentially sharp.

Conjecture 1.16 (Deneanu-Vu, [4])**.** *There are $2^{(0.5+o(1))n^2}$ $n \times n$ $\{\pm 1\}$-valued normal matrices.*

As a first non-trivial step towards this conjecture, they showed the following.

Theorem 1.17 (Deneanu-Vu, [4])**.** *The number of $n \times n$ $\{\pm 1\}$-valued normal matrices is at most $2^{(c_{DV}+o(1))n^2}$ for some constant $c_{DV} < 0.698$.*

The problem of counting normal matrices also boils down to the problem of counting the number of solutions to some underdetermined system of linear equations, and using our framework, it is very easy to obtain an upper bound on the number of such matrices of the form $2^{(1-\alpha)n^2}$, for some $\alpha > 0$. Unfortunately, it does not seem that one can get $1 - \alpha < c_{DV}$ using this simple method. However, the proof of Theorem 1.17 in [4] itself uses the Odlyzko bound at a certain stage; therefore, by using their strategy as a black-box, with the application of the Odlyzko bound at this stage replaced by our better bound, we obtain:

Theorem 1.18. *There exists some $\delta > 0$ such that the number of $n \times n$ $\{\pm 1\}$-valued normal matrices is at most $2^{(c_{DV}-\delta+o(1))n^2}$, where c_{DV} denotes the constant in [4].*

2 Tools

2.1 The Fourier transform

For $p \in [1, \infty)$, let $\mathcal{L}^p(\mathbb{R}^d)$ denote the set of functions $f\colon \mathbb{R}^d \to \mathbb{C}$ such that $\int_{\mathbb{R}^d} |f(x)|^p dx < \infty$. For $f \in \mathcal{L}^1(\mathbb{R}^d)$, the Fourier transform of f – denoted by $\widehat{f}$ – is a function from $\mathbb{R}^d$ to $\mathbb{C}$ given by:

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i \langle x, \xi \rangle} dx,$$

where $\langle x, \xi \rangle := x_1\xi_1 + \cdots + x_d\xi_d$ denotes the standard inner product on $\mathbb{R}^d$. For the reader's convenience, as well as to establish notation, we summarize the following basic properties of the Fourier transform which may be found in any standard textbook on analysis (see, e.g., [18]).

- (Parseval's formula) Let $f, g \in \mathcal{L}^1(\mathbb{R}^d) \cap \mathcal{L}^2(\mathbb{R}^d)$. Then, the Fourier transforms $\widehat{f}, \widehat{g}$ are also in $\mathcal{L}^2(\mathbb{R}^d)$. Moreover,
$$\int_{\mathbb{R}^d} f(x)\overline{g(x)}dx = \int_{\mathbb{R}^d} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}d\xi.$$
- (Convolution formula) For $f, g \in \mathcal{L}^1(\mathbb{R}^d)$, let $f * g \colon \mathbb{R}^d \to \mathbb{C}$ denote the convolution of f and g i.e.
$$f * g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)dy.$$
Then, $f * g \in \mathcal{L}^1(\mathbb{R}^d)$, and for any $\xi \in \mathbb{R}^d$
$$\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi).$$
- (Fourier inversion) Let $f \in \mathcal{L}^1(\mathbb{R}^d)$ be such that $\widehat{f}$ is also in $\mathcal{L}^1(\mathbb{R}^d)$. Then, for any $x \in \mathbb{R}^d$
$$f(x) = \int_{\mathbb{R}^d} \widehat{f}(\xi)e^{2\pi i\langle x,\xi\rangle}d\xi.$$
- (Fourier transform of autocorrelation) Let $f \in \mathcal{L}^1(\mathbb{R}^d)$ be real-valued, and let h denote the autocorrelation of f i.e.
$$h(x) := \int_{\mathbb{R}^d} f(y)f(x+y)dy.$$
Then, for all $\xi \in \mathbb{R}^d$,
$$\widehat{h}(\xi) = \left|\widehat{f}(\xi)\right|^2.$$

The notion of Fourier transform extends more generally to finite Borel measures on $\mathbb{R}^d$. For such a measure μ, the Fourier transform is a function from $\mathbb{R}^d$ to $\mathbb{C}$ given by:
$$\widehat{\mu}(\xi) := \int_{\mathbb{R}^d} e^{-2\pi i\langle x,\xi\rangle}d\mu(x).$$
To see the connection with the Fourier transform for functions in $\mathcal{L}^1(\mathbb{R}^d)$, note that if the measure μ is absolutely continuous with respect to the Lebesgue measure λ, then the density (more precisely, the Radon-Nikodym derivative) $f_\mu := d\mu/d\lambda$ is in $\mathcal{L}^1(\mathbb{R}^d)$, and we have $\widehat{\mu}(\xi) = \widehat{f_\mu}(\xi)$.

The only finite Borel measures we will deal with are those which arise as distributions of random vectors valued in $\mathbb{R}^d$. For a d-dimensional random vector X, let μ_X denote its distribution. Then, we have (see, e.g., [5]):

- (Fourier transform of independent random variables) Let $X_1, \ldots, X_\ell$ be independent d-dimensional random vectors, and let $S_\ell := X_1 + \cdots + X_\ell$ denote their sum. Then, for all $\xi \in \mathbb{R}^d$,
$$\widehat{\mu_{S_\ell}}(\xi) = \prod_{i=1}^{\ell} \widehat{\mu_{X_i}}(\xi).$$
- (Inversion at atoms) Let X be a d-dimensional random vector. For any $x \in \mathbb{R}^d$,
$$\mu_X(\{x\}) = \lim_{T_1,\ldots,T_d \to \infty} \frac{1}{\mathrm{vol}(B[T_1,\ldots,T_d])} \int_{B[T_1,\ldots,T_d]} e^{2\pi i\langle x,t\rangle}\widehat{\mu_X}(t)dt,$$
where $B[T_1, \ldots, T_d]$ denotes the box $[-T_1, T_1] \times \cdots \times [-T_d, T_d]$.
- (Fourier transform of origin-symmetric random vectors) Let X be a d-dimensional, origin-symmetric random vector i.e. $\mu_X(x) = \mu_X(-x)$ for all $x \in \mathbb{R}^d$. Then, $\widehat{\mu_X}$ is a real-valued function.

2.2 Anti-concentration

Definition 2.1. For a random vector X valued in $\mathbb{R}^d$, its (Euclidean) *Lévy concentration function* $\mathcal{L}(X, \cdot)$ is a function from $\mathbb{R}^{\geq 0}$ to $\mathbb{R}$ defined by:

$$\mathcal{L}(X, \delta) := \sup_{u \in \mathbb{R}^d} \Pr[\|X - u\|_2 \leq \delta].$$

Anti-concentration inequalities seek to upper bound the Lévy concentration function for various values of δ. In the discrete setting, a particularly important case is $\delta = 0$, which corresponds to the size of the largest atom in the distribution of the random variable X. The proofs of our Halász-type inequalities will exploit two very general anti-concentration phenomena.

The first principle states that sums of independent random variables do not concentrate much more than sums of suitable independent Gaussians. In particular, for the weighted sum of independent Rademacher variables, Erdős gave a beautiful combinatorial proof to show (improving on a previous bound of Littlewood and Offord) the following.

Theorem 2.2 (Erdős, [6])**.** *Let $a = (a_1, \ldots, a_n)$ be a vector in $\mathbb{R}^n$ all of whose entries are nonzero. Let S_a denote the random sum $\epsilon_1 a_1 + \cdots + \epsilon_n a_n$, where the ϵ_i's are independent Rademacher random variables. Then,*

$$\sup_{c \in \mathbb{R}} \Pr[S_a = c] \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} \sim \sqrt{\frac{2}{\pi n}}.$$

Up to a constant, this was subsequently generalized by Rogozin to handle the Lévy concentration function of sums of general independent random variables.

Theorem 2.3 (Rogozin, [16])**.** *There exists a universal constant $C > 0$ such that for any independent random variables $X_1, \ldots, X_n$, and any $r > 0$, we have*

$$\mathcal{L}(S_n, \delta) \leq \frac{C}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(X_i, \delta))}},$$

where $S_n := X_1 + \cdots + X_n$.

The second anti-concentration principle concerns random vectors of the form AX, where A is a fixed $m \times n$ matrix, and $X = (X_1, \ldots, X_n)$ is a random vector with independent coordinates. It states roughly that if the X_i's are anti-concentrated on the line, and if A has large rank in a suitable sense, then the random vector AX is anti-concentrated in space [17].

As a first illustration of this principle, we present the following lemma, which may be viewed as a 'tensorization' of the Erdős-Littlewood-Offord inequality.

Lemma 2.4. *Let A be an $m \times n$ matrix (where $m \leq n$) of rank r, and let X be a random vector distributed uniformly on $\{\pm 1\}^n$. Then for any $\ell \in \mathbb{N}$,*

$$\sup_{u \in \mathbb{R}^m} \Pr[AX^{(1)} + \cdots + AX^{(\ell)} = u] \leq \left(2^{-\ell} \binom{\ell}{\lfloor \ell/2 \rfloor}\right)^r,$$

where $X^{(1)}, \ldots, X^{(\ell)}$ are i.i.d. copies of X.

Proof. By relabeling the coordinates if needed, we may write A as a block matrix $\begin{pmatrix} E & F \\ G & H \end{pmatrix}$ where E is an $r \times r$ invertible matrix, F is an $r \times (n-r)$ matrix, G is a $(m-r) \times r$ matrix, and H is a

$(m-r)\times(n-r)$ matrix. Let B denote the invertible $m\times m$ matrix $\begin{pmatrix} E^{-1} & 0 \\ 0 & I_{m-r}\end{pmatrix}$, and note that $BA = \begin{pmatrix} I_r & * \\ * & *\end{pmatrix}$. For a vector $v\in\mathbb{R}^s$ with $s\geq r$, let $Q_r(v)\in\mathbb{R}^r$ denote the vector consisting of the first r coordinates of v. Also, let $X_i^{(j)}$ denote the i^{th} coordinate of the random vector $X^{(j)}$, let $\mathcal{R}$ denote the collection of random variables $\{X_{r+1}^{(1)},\dots,X_n^{(1)},\dots X_{r+1}^{(\ell)},\dots,X_n^{(\ell)}\}$, and let $\mathcal{S}$ denote the collection of random variables $\{X_1^{(1)},\dots,X_r^{(1)},\dots,X_1^{(\ell)},\dots,X_r^{(\ell)}\}$. Then for any $u\in\mathbb{R}^m$, we have:

$$\begin{aligned}
\Pr\left[AX^{(1)}+\cdots+AX^{(\ell)}=u\right] &= \Pr\left[BAX^{(1)}+\cdots+BAX^{(\ell)}=Bu\right]\\
&\leq \Pr\left[Q_r(BAX^{(1)})+\cdots+Q_r(BAX^{(\ell)})=Q_r(Bu)\right]\\
&= \mathbb{E}_{\mathcal{R}}\left[\Pr_{\mathcal{S}}\left[Q_r(BAX^{(1)})+\cdots+Q_r(BAX^{(\ell)})=Q_r(Bu)|\mathcal{R}\right]\right]\\
&= \mathbb{E}_{\mathcal{R}}\left[\Pr_{\mathcal{S}}\left[Q_r(X^{(1)})+\cdots+Q_r(X^{(\ell)})=f(\mathcal{R})|\mathcal{R}\right]\right]\\
&= \mathbb{E}_{\mathcal{R}}\left[\prod_{i=1}^{r}\Pr\left[X_i^{(1)}+\cdots+X_i^{(\ell)}=f_i(\mathcal{R})|\mathcal{R}\right]\right]\\
&\leq \mathbb{E}_{\mathcal{R}}\left[\prod_{i=1}^{r}2^{-\ell}\binom{\ell}{\ell/2}\right]\\
&= \left(2^{-\ell}\binom{\ell}{\ell/2}\right)^r,
\end{aligned}$$

where the third line follows from the law of total probability; the fourth line follows from the explicit form of BA mentioned above; the fifth line follows from the independence of the coordinates of $X^{(j)}$; and the sixth line follows from the Erdős-Littlewood-Offord inequality (Theorem 2.2). Taking the supremum over $u\in\mathbb{R}^m$ completes the proof. □

Remark 2.5. By using Rogozin's inequality (Theorem 2.3) instead of the Erdős-Littlewood-Offord inequality, we may generalize the lemma to handle any random vector $X=(X_1,\dots,X_n)$ with independent coordinates X_i, provided we replace the conclusion by

$$\sup_{u\in\mathbb{R}^m}\Pr[AX^{(1)}+\cdots+AX^{(\ell)}=u]\leq\left(\frac{C}{\ell}\right)^{r/2}\times\max_{I\subseteq[n],|I|=r}\prod_{i\in I}\frac{1}{\sqrt{1-\mathcal{L}(X_i,0)}},$$

where C is a universal constant.

For the Lévy concentration function for general δ, a version of Lemma 2.4 was proved by Rudelson and Vershynin in [17].

Theorem 2.6 (Rudelson-Vershynin, [17])**.** *Consider a random vector $X=(X_1,\dots,X_d)$ where X_i are real-valued independent random variables. Let $\delta,\rho\geq 0$ be such that for all $i\in[d]$,*

$$\mathcal{L}(X_i,\delta)\leq\rho.$$

Then, for every $m\times n$ matrix A, every $M\geq 1$ and every $\varepsilon\in(0,1)$, we have

$$\mathcal{L}\left(AX,M\delta\|A\|_{HS}\right)\leq\left(C_\varepsilon M\rho\right)^{\lceil(1-\varepsilon)r_s(A)\rceil},$$

where $C_\varepsilon=C/\sqrt{\varepsilon}$ for some absolute constant $C>0$.

More general statements of a similar nature may be found in [17].

2.3 The replication trick

In this section, we present the 'replication trick', which allows us to reduce considerations about anti-concentration of sums of independent random vectors to considerations about anti-concentration of sums of independent *identically distributed* random vectors. This will be useful since the 'correct' analog of Rogozin's inequality for general random vectors with independently coordinates is not available; to our knowledge, the best result in this direction is due to Esseen [7], who proved an inequality of this form for such random vectors satisfying additional symmetry conditions, which will not be available in our applications. The statement/proof of the 'atomic' version of the replication trick (Proposition 2.7) is similar in spirit to Corollaries 7.12 and 7.13 in [21] with an important difference: we have no need for the lossy 'domination' and 'duplication' steps in [21]; instead, we ensure the non-negativity of the Fourier transform at various places by using the previously stated simple fact that the Fourier transform of the distribution of an origin-symmetric random vector is real valued, and restricting ourselves to even powers thereof.

Proposition 2.7. *Let $X_1, \dots, X_n$ be independent random vectors valued in $\mathbb{R}^d$. For each $i \in [n]$, let $\tilde{X}_i := X_i - X_i'$, where X_i' is an independent copy of X_i. Let $S_n := X_1 + \cdots + X_n$, and for any $i \in [n]$, $m \in \mathbb{N}$, let $\tilde{S}_{i,m} := \tilde{X}_i^{(1)} + \dots \tilde{X}_i^{(m)}$, where $\tilde{X}_i^{(1)}, \dots, \tilde{X}_i^{(m)}$ are independent copies of $\tilde{X}_i$. Then for any $v \in \mathbb{R}^d$,*

$$\Pr\left[S_n = v\right] \leq \prod_{i=1}^{n} \Pr\left[\tilde{S}_{i,a_i/2} = 0\right]^{\frac{1}{a_i}}$$

for any $a_1, \dots, a_n \in 4 \cdot \mathbb{N}$ such that $a_1^{-1} + \cdots + a_n^{-1} = 1$.

Here, $4 \cdot \mathbb{N}$ denotes the subset of natural numbers given by $\{4m \colon m \in \mathbb{N}\}$.

Proof. As before, we let μ_X denote the distribution of the d-dimensional random vector X. We have:

$$\begin{aligned}
\mu_{S_n}(v) &= \lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \int_{B[T_1,\dots,T_d]} e^{-i\langle t, v\rangle} \widehat{\mu_{S_n}}(t) dt \\
&= \lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \int_{B[T_1,\dots,T_d]} e^{-i\langle t, v\rangle} \prod_{i=1}^{n} \widehat{\mu_{X_i}}(t) dt \\
&\leq \lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \prod_{i=1}^{n} \left(\int_{B[T_1,\dots,T_d]} |\widehat{\mu_{X_i}}(t)|^{a_i} \, dt \right)^{\frac{1}{a_i}} \\
&= \lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \prod_{i=1}^{n} \left(\int_{B[T_1,\dots,T_d]} \left(\widehat{\mu_{\tilde{X}_i}}(t)\right)^{\frac{a_i}{2}} \, dt \right)^{\frac{1}{a_i}} \\
&= \lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \prod_{i=1}^{n} \left(\int_{B[T_1,\dots,T_d]} \widehat{\mu_{\tilde{S}_{i,a_i/2}}}(t) dt \right)^{\frac{1}{a_i}} \\
&= \prod_{i=1}^{n} \left(\lim_{T_1,\dots,T_d \to \infty} \frac{1}{\mathrm{vol} B[T_1,\dots,T_d]} \int_{B[T_1,\dots,T_d]} \widehat{\mu_{\tilde{S}_{i,a_i/2}}}(t) dt \right)^{\frac{1}{a_i}} \\
&= \prod_{i=1}^{n} \left(\mu_{\tilde{S}_{i,a_i/2}}(0) \right)^{\frac{1}{a_i}},
\end{aligned}$$

where the first line follows from the Fourier inversion formula at atoms; the second line follows from the independence of $X_1, \dots, X_n$; the third line follows from Hölder's inequality; the fourth line

follows from the fact that $\widehat{\mu_{\tilde{X}_i}}(t) = |\widehat{\mu_{X_i}}(t)|^2$ (since the distribution of $\tilde{X}_i$ is the autocorrelation of the distribution of X_i); the fifth line follows from the independence of $\tilde{X}_i^{(1)} \ldots, \tilde{X}_i^{(a_i/2)}$; and the last line follows again from the Fourier inversion formula at atoms. □

Remark 2.8. The same proof shows that when $X_1, \ldots, X_n$ are independent *origin symmetric* random vectors, then for any $v \in \mathbb{R}^d$

$$\Pr[S_n = v] \leq \prod_{i=1}^{n} \Pr[S_{i,a_i} = 0]^{\frac{1}{a_i}}$$

for any $a_1, \ldots, a_n \in 2\mathbb{N}$ such that $a_1^{-1} + \cdots + a_n^{-1} = 1$, where S_{i,a_i} denotes the sum of a_i independent copies of X_i.

The next proposition is a version of Proposition 2.7 for the Lévy concentration function. Essentially the same proof can also be used to prove variants for norms other than the Euclidean norm.

Proposition 2.9. *Let $X_1, \ldots, X_n$ be independent random vectors valued in $\mathbb{R}^d$. For each $i \in [n]$, let $\tilde{X}_i := X_i - X_i'$, where X_i' is an independent copy of X_i. Let $S_n := X_1 + \cdots + X_n$, and for any $i \in [n]$, $m \in \mathbb{N}$, let $\tilde{S}_{i,m} := \tilde{X}_i^{(1)} + \ldots \tilde{X}_i^{(m)}$, where $\tilde{X}_i^{(1)}, \ldots, \tilde{X}_i^{(m)}$ are independent copies of $\tilde{X}_i$. Then for any $\delta > 0$,*

$$\mathcal{L}(S_n, \delta) \leq 2^d \prod_{i=1}^{n} \mathcal{L}(\tilde{S}_{i,a_i/2}, 4\delta)^{1/a_i}$$

for any $a_1, \ldots, a_n \in 4\mathbb{N}$ such that $a_1^{-1} + \cdots + a_n^{-1} = 1$.

Proof. Let $\mathbf{1}_{B_\delta(0)}$ denote the indicator function of the ball of radius δ centered at the origin. We will make use of the readily verified elementary inequality

$$\mathrm{vol}(B_\delta(0))\mathbf{1}_{B_\delta(0)}(x) \leq \mathbf{1}_{B_{2\delta}(0)} * \mathbf{1}_{B_{2\delta}(0)}(x) \leq \mathrm{vol}(B_{2\delta}(0))\mathbf{1}_{B_{4\delta}(0)}(x). \tag{1}$$

By adding to each X_i an independent random vector with distribution given by a 'bump function' with arbitrarily small support around the origin, we may assume that the distributions of all the random vectors under consideration are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and thus have densities. For such a random vector Y, we will denote its density with respect to the d-dimensional Lebesgue measure by f_Y. Then, for any $v \in \mathbb{R}^d$, we have:

$$\begin{aligned}
\Pr[\|S_n - v\|_2 \leq \delta] &= \int_{x \in \mathbb{R}^d} \mathbf{1}_{B_\delta(0)}(x) f_{S_n}(x+v) dx \\
&\leq \mathrm{vol}(B_\delta(0))^{-1} \int_{x \in \mathbb{R}^d} \left(\mathbf{1}_{B(2\delta)} * \mathbf{1}_{B(2\delta)}\right)(x) f_{S_n}(x+v) dx \\
&= \mathrm{vol}(B_\delta(0))^{-1} \int_{\xi \in \mathbb{R}^d} e^{2\pi i \langle \xi, v \rangle} \left(\mathbf{1}_{B(2\delta)} * \mathbf{1}_{B(2\delta)}\right)^\wedge(\xi) \widehat{f_{S_n}}(\xi) d\xi \\
&= \mathrm{vol}(B_\delta(0))^{-1} \int_{\xi \in \mathbb{R}^d} e^{2\pi i \langle \xi, v \rangle} \left(\widehat{\mathbf{1}_{B(2\delta)}}(\xi)\right)^2 \prod_{i=1}^{n} \widehat{f_{X_i}}(\xi) d\xi \\
&= \mathrm{vol}(B_\delta(0))^{-1} \int_{\xi \in \mathbb{R}^d} e^{2\pi i \langle \xi, v \rangle} \prod_{i=1}^{n} \left(\left(\widehat{\mathbf{1}_{B(2\delta)}}(\xi)\right)^{\frac{2}{a_i}} \widehat{f_{X_i}}(\xi)\right) d\xi \\
&\leq \mathrm{vol}(B_\delta(0))^{-1} \prod_{i=1}^{n} \left(\int_{\xi \in \mathbb{R}^d} \left(\widehat{\mathbf{1}_{B(2\delta)}}(\xi)\right)^2 \left|\widehat{f_{X_i}}(\xi)\right|^{a_i} d\xi\right)^{\frac{1}{a_i}}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{vol}(B_\delta(0))^{-1} \prod_{i=1}^n \left(\int_{\xi\in\mathbb{R}^d} \left(\widehat{\mathbf{1}_{B(2\delta)}}(\xi)\right)^2 \left(\widehat{f_{\tilde{X}_i}}(\xi)\right)^{\frac{a_i}{2}} d\xi \right)^{\frac{1}{a_i}} \\
&= \mathrm{vol}(B_\delta(0))^{-1} \prod_{i=1}^n \left(\int_{\xi\in\mathbb{R}^d} \left(\mathbf{1}_{B(2\delta)} * \mathbf{1}_{B(2\delta)}\right)^\wedge(\xi) \widehat{f_{\tilde{S}_{i,a_i/2}}}(\xi) d\xi \right)^{\frac{1}{a_i}} \\
&= \mathrm{vol}(B_\delta(0))^{-1} \prod_{i=1}^n \left(\int_{x\in\mathbb{R}^d} \left(\mathbf{1}_{B(2\delta)} * \mathbf{1}_{B(2\delta)}\right)(x) f_{\tilde{S}_{i,a_i/2}}(x) dx \right)^{\frac{1}{a_i}} \\
&\le \mathrm{vol}(B_\delta(0))^{-1} \mathrm{vol}(B_{2\delta}(0)) \prod_{i=1}^n \left(\int_{x\in\mathbb{R}^d} \mathbf{1}_{B(4\delta)}(x) f_{\tilde{S}_{i,a_i/2}}(x) dx \right)^{\frac{1}{a_i}} \\
&= 2^d \prod_{i=1}^n \left(\Pr\left[\|\tilde{S}_{i,a_i/2}\|_2 \le 4\delta \right] \right)^{\frac{1}{a_i}},
\end{aligned}
$$

where the second line follows from (1); the third line follows from Parseval's formula; the fourth line follows from the convolution formula and the independence of $X_1, \dots, X_n$; the sixth line follows from Hölder's inequality, along with the fact that $\widehat{\mathbf{1}_{B(2\delta)}}(\xi)$ is real valued for all $\xi \in \mathbb{R}^d$; the seventh line follows from the fact that $|\widehat{f_{X_i}}(\xi)|^2 = \widehat{f_{\tilde{X}_i}}(\xi)$ for all $\xi \in \mathbb{R}^d$; the ninth line follows again from Parseval's formula; and the tenth line follows from (1). Taking the supremum over all $v \in \mathbb{R}^d$ gives the desired conclusion. □

Remark 2.10. As in Remark 2.8, if $X_1, \dots, X_n$ are origin-symmetric, then the same conclusion holds with $\tilde{S}_{i,a_i/2}$ replaced by S_{i,a_i}, for any $a_1, \dots, a_n \in 2\mathbb{N}$ with $a_1^{-1} + \cdots + a_n^{-1} = 1$.

3 Proofs

3.1 Proofs of Halász-type inequalities

By combining the tools from Sections 2.2 and 2.3, we can now prove our Halász-type inequalities. All of them follow the same general outline. We begin by proving Theorem 1.11.

Proof of Theorem 1.11. Let $\mathcal{A}_1, \dots, \mathcal{A}_\ell$ be the partition of $\{a_1, \dots, a_n\}$ as in the statement of the theorem. For each $i \in [\ell]$, let A_i denote $d \times |\mathcal{A}_i|$ dimensional matrix whose columns are given by the elements of $\mathcal{A}_i$. With this notation, we can rewrite the random vector $\sum_{i=1}^n \epsilon_i a_i$ as $\sum_{j=1}^\ell A_j Y_j$, where Y_j is uniformly distributed on $\{\pm 1\}^{|\mathcal{A}_j|}$ and $Y_1, \dots, Y_\ell$ are independent.

Since the random vectors $X_1 := A_1 Y_1, \dots, X_n := A_n Y_n$ are origin-symmetric, and since $\ell \in 2\mathbb{N}$, it follows from Proposition 2.7 and Remark 2.8 that for any $u \in \mathbb{R}^d$,

$$
\begin{aligned}
\Pr\left[\sum_{i=1}^n \epsilon_i a_i = u \right] &= \Pr\left[\sum_{j=1}^\ell X_j = u \right] \\
&\le \prod_{i=1}^\ell \Pr\left[X_j^{(1)} + \cdots + X_j^{(\ell)} = 0 \right]^{\frac{1}{\ell}},
\end{aligned}
$$

where $X_j^{(1)}, \dots, X_j^{(\ell)}$ are i.i.d. copies of X_j. Further, since $\mathrm{rank}(A_j) = r_j$ by assumption, it follows from Lemma 2.4 that

$$
\Pr\left[X_j^{(1)} + \cdots + X_j^{(\ell)} = 0 \right] = \Pr\left[A_j Y_j^{(1)} + \cdots + A_j Y_j^{(\ell)} = 0 \right]
$$

$$\leq \left(2^{-\ell}\binom{\ell}{\ell/2}\right)^{r_j}.$$

Substituting this bound in the previous inequality completes the proof. □

By using Remark 2.5 instead of Lemma 2.4, we can use the same proof to obtain the following more general statement.

Theorem 3.1. *Let $a_1, \ldots, a_n$ be a collection vectors in $\mathbb{R}^d$ which can be partitioned as $\mathcal{A}_1, \ldots, \mathcal{A}_\ell$ such that $\dim_{\mathbb{R}^d}(\mathrm{span}\{a : a \in \mathcal{A}_i\}) =: r_i$. Let $x_1, \ldots, x_n$ be independent random variables, and for each $i \in [n]$, let $\tilde{x}_i := x_i - x_i'$, where x_i' is an independent copy of x_i. Then,*

$$\sup_{u \in \mathbb{R}^d} \Pr\left[\sum_{i=1}^n x_i a_i = u\right] \leq 2^d \inf_{(b_1,\ldots,b_\ell)\in\mathcal{B}} \left(\frac{C}{\ell\lambda}\right)^{\sum_{i=1}^\ell \frac{r_i}{2b_i}},$$

where $\lambda := \min_{i\in[n]}(1 - \mathcal{L}_{\tilde{x}_i}(0))$ and $\mathcal{B} = \{(b_1, \ldots, b_\ell) \in (4\mathbb{N})^\ell : b_1^{-1} + \cdots + b_\ell^{-1} = 1\}$.

We now state and prove the general small-ball version of our anti-concentration inequality.

Theorem 3.2. *Let $a_1, \ldots, a_n$ be a collection vectors in $\mathbb{R}^d$. Let $\mathcal{A}_1, \ldots, \mathcal{A}_\ell$ be a partition of the set $\{a_1, \ldots, a_n\}$, and for each $i \in [\ell]$, let A_i denote the $d \times |\mathcal{A}_i|$ dimensional matrix whose columns are given by the elements of $\mathcal{A}_i$. Let $x_1, \ldots, x_n$ be independent random variables, and for each $i \in [n]$, let $\tilde{x}_i := x_i - x_i'$, where x_i' is an independent copy of x_i. Let $\delta, \lambda \geq 0$ be such that $\min_{i\in[n]}(1 - \mathcal{L}(\tilde{x}_i, \delta)) = \lambda$. Then, for every $M \geq 1$ and $\varepsilon \in (0, 1)$,*

$$\mathcal{L}\left(\sum_{i=1}^n x_i a_i, M\delta\right) \leq \inf_{(b_1,\ldots,b_\ell)\in\mathcal{B}} \prod_{i=1}^\ell \left(\frac{CM}{\sqrt{\varepsilon b_i \lambda}\|A_i\|_{HS}}\right)^{\frac{\lceil(1-\varepsilon)r_s(A_i)\rceil}{b_i}},$$

where $r_s(A_i)$ denotes the stable rank of A_i, C is an absolute constant, and $\mathcal{B} = \{(b_1, \ldots, b_\ell) \in (4\mathbb{N})^\ell : b_1^{-1} + \cdots + b_\ell^{-1} = 1\}$.

Proof. As before, we begin by rewriting the random vector $\sum_{i=1}^n x_i a_i$ as $\sum_{i=1}^\ell A_i Y_i$. From Proposition 2.9, it follows that for any $(b_1, \ldots, b_\ell) \in \mathcal{B}$,

$$\begin{aligned}\mathcal{L}\left(\sum_{i=1}^n x_i a_i, M\delta\right) &= \mathcal{L}\left(\sum_{i=1}^\ell A_i Y_i, M\delta\right) \\ &\leq 2^d \prod_{i=1}^\ell \mathcal{L}\left(A_i\left(\tilde{Y}_i^{(1)} + \cdots + \tilde{Y}_i^{(b_i/2)}\right), 4M\delta\right)^{\frac{1}{b_i}}.\end{aligned}$$

Next, since $1 - \mathcal{L}(\tilde{x}_i, \delta) \geq \lambda$ for all $i \in [n]$, it follows from Theorem 2.3 that

$$\mathcal{L}\left(\tilde{x}_i^1 + \cdots + \tilde{x}_i^{(b_i/2)}, \delta\right) \leq \frac{C}{\sqrt{b_i\lambda}},$$

where C is an absolute constant. In particular, all of the (independent) coordinates of the random vector $\tilde{Y}_i^{(1)} + \cdots + \tilde{Y}_i^{(b_i/2)}$ have δ-Lévy concentration function bounded by $C/\sqrt{b_i\lambda}$. Hence, it follows from Theorem 2.6 that

$$\mathcal{L}\left(A_i\left(\tilde{Y}_i^{(1)} + \cdots + \tilde{Y}_i^{(b_i/2)}\right), 4M\delta\right) \leq \left(\frac{CM}{\sqrt{\varepsilon b_i\lambda}\|A_i\|_{\mathrm{HS}}}\right)^{\lceil(1-\varepsilon)r_s(A_i)\rceil},$$

where C is an absolute constant. Substituting this in the first inequality completes the proof. □

Remark 3.3. When the x_i's are origin symmetric random variables, we may use Remark 2.10 instead of Proposition 2.9 to obtain a similar conclusion – with the infimum now over the larger set $\mathcal{B}' = \{(b_1, \ldots, b_\ell) \in (2\mathbb{N})^\ell : b_1^{-1} + \cdots + b_\ell^{-1} = 1\}$ – under the assumption that $\min_{i\in[n]}(1 - \mathcal{L}(x_i, \delta)) = \lambda$. In particular, if ℓ is even, then taking $b_1 = \cdots = b_\ell = \ell$ gives Theorem 1.15.

3.2 Proof of Theorem 1.2

As in Section 1.1.1, let $H_{k,n}$ denote a $k \times n$ matrix with all its entries in $\{\pm 1\}$ and all of whose rows are orthogonal. For convenience of notation, we isolate the following notion.

Definition 3.4. For any $r, \ell \in \mathbb{N}$, a matrix M is said to admit an (r, ℓ)-*rank partition* if there exists a decomposition of the columns of M into ℓ disjoint subsets, each of which corresponds to a submatrix of rank at least r.

Note that the existence of an (r, ℓ)-rank partition is a uniform version of the condition appearing in Theorem 1.11. The next proposition shows that any $H_{k,n}$ with k admits an (r, ℓ)-rank partition with r and ℓ sufficiently large.

Proposition 3.5. *Let $r, \ell \in \mathbb{N}$ such that $2 \le \ell, r \le k$ and $(e^2\ell)^k < (n/r)^{k-r}$. Then, $H_{k,n}$ admits an (r, ℓ)-rank partition.*

Proof. The proof proceeds in two steps – first, we show that $H_{k,n}$ contains many non-zero $k \times k$ minors, and second, we apply a simple greedy procedure to these non-zero minors to produce an (r, ℓ)-rank partition for the desired values of r and ℓ.

The first step follows easily from the classical Cauchy-Binet formula (see, e.g., [2]), which asserts that:

$$\det(H_{k,n}H_{k,n}^T) = \sum_{A\in\mathcal{M}_k} \det(A)^2,$$

where $\mathcal{M}_k$ denotes the set of all $k \times k$ submatrices of $H_{k,n}$. In our case, $H_{k,n}H_{k,n}^T = n\mathrm{Id}_k$, so that $\det(H_{k,n}H_{k,n}^T) = n^k$. Moreover, since each $A \in \mathcal{M}_k$ is a $k \times k$ $\{\pm 1\}$-valued matrix, $\det(A)^2 \le k^k$ (with equality attained if and only if A is itself a Hadamard matrix). Hence, it follows from the Cauchy-Binet formula that $H_{n,k}$ has at least $(n/k)^k$ non-zero minors.

Next, we use these non-zero minors to construct an (r, ℓ)-rank partition in ℓ steps as follows: In Step 1, choose r columns of an arbitrary non-zero minor – such a minor is guaranteed to exist by the discussion above. Let $\mathcal{C}_k$ denote the union of the columns chosen by the end of Step k, for any $1 \le k \le \ell - 1$. In Step $k + 1$, we choose r linearly independent columns which are disjoint from $\mathcal{C}_k$. Then, the ℓ collections of r columns chosen at different steps gives an (r, ℓ)-rank partition of $H_{k,n}$.

Therefore, to complete the proof, it only remains to show that for each $1 \le k \le \ell - 1$, there is a choice of r linearly independent columns which are disjoint from $\mathcal{C}_k$. Since $|\mathcal{C}_k| = rk$, this is in turn implied by the stronger statement that there is a choice of r linearly independent columns which are disjoint from any collection $\mathcal{C}$ of at most $r\ell$ columns. In order to see this, we note that the number of $k \times k$ submatrices of $H_{k,n}$ which have at least $k - r$ columns contained in $\mathcal{C}$ is at most:

$$\begin{aligned}
\sum_{s=0}^{r} \binom{r\ell}{k-s}\binom{n}{s} &\le \binom{r\ell}{k}\sum_{s=0}^{r}\binom{n}{s} \\
&\le \left(\frac{er\ell}{k}\right)^k \left(\frac{en}{r}\right)^r \\
&< \left(\frac{n}{k}\right)^k,
\end{aligned}$$

where the first inequality uses $2 \leq \ell$ and the final inequality follows by assumption. Since there are at least $(n/k)^k$ non-zero minors of $H_{k,n}$, it follows that there exists a $k \times k$ submatrix A_{k+1} of $H_{n,k}$ of full rank which shares at most $k - r$ columns with $\mathcal{C}_k$. In particular, A_{k+1} contains r linearly independent columns which are disjoint from $\mathcal{C}_k$, as desired. □

The previous proposition essentially completes the proof of Theorem 1.2. Indeed, recall from Section 1.1.1 that it suffices to show the following: there exist absolute constants $0 < c_1 < c_2 < 1$ and $C > 0$ such that for all $k \in [c_1 n, c_2 n]$, the number of solutions $x \in \{\pm 1\}^n$ to $H_{k,n}x = 0$ is at most $2^{-(1+C)k}$. The previous proposition shows that $H_{k,n}$ admits an (r, ℓ)-rank partition with $r = \lfloor k/2 \rfloor$ and $\ell = \lfloor \sqrt{n/ke^4} \rfloor$. Hence, from Theorem 1.11, it follows that for $k \in [1, n/15000]$, the number of solutions $x \in \{\pm 1\}^n$ to $H_{k,n}x = 0$ is at most $2^{n-(1+1/10)k}$, which completes the proof.

Remark 3.6. For our problem of providing an upper bound on the number of Hadamard matrices, we could have used the somewhat simpler Proposition 3.8 (instead of Proposition 3.5), which shows that there are very few $H_{k,n}$ which do not admit an (r, ℓ)-rank partition for sufficiently large r, ℓ. However, we used Proposition 3.5 to show that it is easy to find such a rank partition even for a given $k \times n$ system of linear equations A – indeed, the proof of Proposition 3.5 goes through as long as $\det(AA^T)$ is 'large' (which is indeed the case for random or 'pseudorandom' A), and all $k \times k$ minors of A are uniformly bounded (which is guaranteed in settings where A has restricted entries, as in our case).

3.3 Proof of Theorem 1.18

In this section, we show how to obtain a non-trivial upper bound on the number of $\{\pm 1\}$-valued normal matrices using our general framework. As mentioned in the introduction, this bound by itself is not stronger than the one obtained by Deneanu and Vu [4]; however, it can be used in their proof in a modular fashion to obtain an improvement over their bound, thereby proving Theorem 1.18. As the proof of Deneanu and Vu is quite technical, we defer the details of this second step to Appendix A.

Following Deneanu and Vu, we consider the following generalization of the notion of normality:

Definition 3.7. Let N be a fixed (but otherwise arbitrary) $n \times n$ matrix. An $n \times n$ matrix M is said to be N-*normal* if and only if

$$MM^T - M^T M = N.$$

For any $n \times n$ matrix N, we let $\mathcal{N}(N)$ denote the set of all $n \times n$, $\{\pm 1\}$-valued matrices which are N-normal. In particular, $\mathcal{N}(0)$ is the set of all $n \times n$, $\{\pm 1\}$-valued normal matrices. The notion of N-normality is crucial to the proof of Deneanu and Vu, which is based on an inductive argument – they show that the quantity $2^{(c_{DV}+o(1))n^2}$ in Theorem 1.17 is actually a *uniform* upper bound on the size of the set $\mathcal{N}(N)$ for any N. While this general notion of normality is not required to obtain *some* non-trivial upper bound on the number of normal matrices, either using our framework or theirs, we will state and prove the results of this section for N-normality, since this greater generality will be essential in Appendix A.

We begin by introducing some notation, and discussing how to profitably recast the problem of counting N-normal matrices as a problem of counting the number of solutions to an underdetermined system of linear equations. Given any matrix X, we let $r_i(X)$ and $c_i(X)$ denote its i^{th} row and column respectively. With this notation, note that for a given matrix M, being N-normal is equivalent to satisfying the following equation for all $i, j \in [n]$:

$$r_i(M)r_j^T(M) - c_i(M)^T c_j(M) = N_{ij}. \tag{2}$$

In particular, writing M in block form as:

$$M = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix},$$

where A_k is a $k \times k$ matrix, we see that (2) amounts to the following equations:

(*i*) For all $i, j \in [k]$:

$$r_i(A_k)r_j(A_k)^T + r_i(B_k)r_j(B_k)^T - c_i(A_k)^T c_j(A_k) - c_i(C_k)^T c_j(C_k) = N_{ij}.$$

(*ii*) For all $i \in [k], j \in [n-k]$:

$$r_i(A_k)r_j(C_k)^T + r_i(B_k)r_j(D_k)^T - c_i(A_k)^T c_j(B_k) - c_i(C_k)^T c_j(D_k) = N_{i,k+j}.$$

(*iii*) For all $i, j \in [n-k]$:

$$r_i(C_k)r_j(C_k)^T + r_i(D_k)r_j(D_k)^T - c_i(B_k)^T c_j(B_k) - c_i(D_k)^T c_j(D_k) = N_{k+i,k+j}.$$

We now rewrite this system of equations in a form that will be useful for our application. Following Deneanu and Vu, we will count the size of $\mathcal{N}(N)$ by constructing N-normal matrices in $n+1$ steps, and bounding the number of choices available at each step. The steps are as follows: in Step 0, we select n entries $d_1, \ldots, d_n$ to serve as diagonal entries of the matrix M; in Step k for $1 \leq k \leq n$, we select $2(n-k)$ entries so as to completely determine the k^{th} row and the k^{th} column of M – of course, these $2(n-k)$ entries cannot be chosen arbitrarily, and must satisfy some constraints coming from the choice of entries in Steps $0, \ldots, k-1$.

More precisely, let M_k denote the structure obtained at the end of Step k. Then,

$$M_k = \left[\begin{array}{c|ccc} A_k & \multicolumn{3}{c}{B_k} \\ \hline & d_{k+1} & * & * \\ C_k & * & \ddots & * \\ & * & * & d_n \end{array}\right], \tag{3}$$

where the $*$'s denote the parts of D_k which have not been determined by the end of Step k. Observe that the matrix A_k, together with the first column of B_k, the first row of C_k, and the diagonal element d_{k+1} forms the matrix A_{k+1}; in particular, the matrix A_{k+1} is already determined at the end of Step k. Moreover, both B_{k+1} and C_{k+1} are determined at the end of Step k up to their last row and last column respectively.

In Step $k+1$, we choose $r_{k+1}(B_{k+1})$ and $c_{k+1}(C_{k+1})$. In order to make this choice in a manner such that the resulting M_{k+1} admits even a single extension to an N-normal matrix, it is necessary that for all $i \in [k]$:

$$r_{k+1}(A_{k+1})r_i(A_{k+1})^T + r_{k+1}(B_{k+1})r_i^T(B_{k+1}) - c_{k+1}(A_{k+1})^T c_i(A_{k+1}) - c_{k+1}(C_{k+1})^T c_i(C_{k+1}) = N_{k+1,i}.$$

Since A_{k+1} is completely determined by the end of Step k, and since N is fixed, we can rewrite the above equation as: for all $i \in [k]$,

$$r_{k+1}(B_{k+1})r_i^T(B_{k+1}) - c_{k+1}(C_{k+1})^T c_i(C_{k+1}) = N'_{k+1,i}, \tag{4}$$

for some $N'_{k+1,i}$ which is uniquely determined at the end of Step k. Let N'_k be the k-dimensional column vector whose i^{th} entry is given by $N'_{k+1,i}$, let $T_k := [U\ V]$ be the $k \times 2(n-k-1)$ matrix

formed by taking U to be the matrix consisting of the first k rows of B_{k+1} and V^T to be the matrix consisting of the first k columns of C_{k+1}, and let x_k be the $2(n-k-1)$-dimensional column vector given by $x_k := \begin{bmatrix} r_{k+1}^T(B_{k+1}) \\ -c_{k+1}(C_{k+1}) \end{bmatrix}$. With this notation, (4) can be written as:

$$T_k x_k = N'_k. \tag{5}$$

The next proposition is the analogue of Proposition 3.5 in the present setting.

Proposition 3.8. *Let $0 < \gamma < 1$ be fixed, and let M be a random $m \times n'$ $\{\pm 1\}$-valued random matrix. Let $\mathcal{E}_{\gamma,\ell}$ denote the event that M does not admit a $(\gamma m, \ell)$-rank partition. Then,*

$$\Pr[\mathcal{E}_{\gamma,\ell}] \leq 2^{-(1-\gamma)^2 mn' + (1-\gamma)^2(m^2\ell + m^2) + O(n')}.$$

The proof of this proposition is based on the following lemma, which follows easily from Odlyzko's lemma (Lemma 1.5).

Lemma 3.9. *Let $0 < \gamma < 1$ be fixed, and let M be a random $m \times m$ $\{\pm 1\}$-valued random matrix. Then,*

$$\Pr[\mathrm{rank}(M) \leq \gamma m] \leq 2^{-(1-\gamma)^2 m^2 + O(m)}.$$

Proof. For any integer $1 \leq s \leq m$, let $\mathcal{R}_s$ denote the event that $\mathrm{rank}(M) = s$. Since

$$\Pr[\mathrm{rank}(M) \leq \gamma m] = \Pr\left[\bigvee_{s=1}^{m} \mathcal{R}_s\right] \leq \sum_{s=1}^{\gamma m} \Pr[\mathcal{R}_s],$$

it suffices to show that $\Pr[\mathcal{R}_s] \leq 2^{-(1-\gamma)^2 m^2 + O(m)}$ for all $s \in [\gamma m]$. To see this, note by symmetry that

$$\Pr[\mathcal{R}_s] \leq \binom{m}{s} \Pr\left[\mathcal{R}_s \wedge \mathcal{I}_{[s]}\right],$$

where $\mathcal{I}_{[s]}$ is the event that the first s rows of M are linearly independent. Moreover, letting $r_{[s+1,n]}(M)$ denote the set $\{r_{s+1}(M), \ldots, r_m(M)\}$ of the last $m - s$ rows of M, and V_s denote the random vector space spanned by the first s rows of M, we have:

$$\begin{aligned}
\Pr\left[\mathcal{R}_s \wedge \mathcal{I}_{[s]}\right] &\leq \Pr\left[r_{[s+1,n]} \subseteq V_s\right] \\
&= \sum_{v_1,\ldots,v_s \in \{\pm 1\}^m} \Pr\left[r_{[s+1,n]}(M) \subseteq V_s | r_i(M) = v_i, 1 \leq i \leq s\right] \Pr\left[r_i(M) = v_i, 1 \leq i \leq s\right] \\
&= \sum_{v_1,\ldots,v_s \in \{\pm 1\}^m} \left(\prod_{j=s+1}^{m} \Pr\left[r_j(M) \subseteq V_s | r_i(M) = v_i, 1 \leq i \leq s\right]\right) \Pr\left[r_i(M) = v_i, 1 \leq i \leq s\right] \\
&\leq \sum_{v_1,\ldots,v_s \in \{\pm 1\}^m} 2^{(s-m)(m-s)} \Pr\left[r_i(M) = v_i, 1 \leq i \leq s\right] \\
&= 2^{-(m-s)^2},
\end{aligned}$$

where the second line follows from the law of total probability; the third line follows from the independence of the rows of the matrix M; and the fourth line follows from Odlyzko's lemma (Lemma 1.5) along with the fact that conditioning on the values of $r_1(M), \ldots, r_s(M)$ fixes V_s to be a subspace of dimension at most s.

Finally, since $m - s \geq (1-\gamma)m$ and $\binom{m}{s} \leq 2^m$, we get the desired conclusion. ∎

Proof of Proposition 3.8. For each $i \in [t]$, where $t = \lfloor n'/m \rfloor$, let A_i denote the $m \times m$ submatrix of M consisting of the columns $c_{(i-1)m+1}(M), \ldots, c_{im}(M)$. Then,

$$\Pr[\mathcal{E}_{\gamma,\ell}] \leq \Pr\left[|\{i \in [t] : \mathrm{rank}(A_i) \leq \gamma m\}| > t - \ell\right].$$

By Lemma 3.9, we have for each $i \in [t]$ that

$$\Pr[\mathrm{rank}(A_i) \leq \gamma m] \leq 2^{-(1-\gamma)^2 m^2 + O(m)}.$$

Therefore, since the entries of the different A_i's are independent, the probability of having more than $t - \ell$ indices $i \in [t]$ for which $\mathrm{rank}(A_i) \leq \gamma m$ is at most:

$$\begin{aligned}
\sum_{k=t-\ell+1}^{t} \binom{t}{k} \left(2^{-(1-\gamma)^2 m^2 + O(m)}\right)^k &\leq t 2^t 2^{-(1-\gamma)^2 m^2 (t-\ell) + O(tm)} \\
&\leq 2^{-(1-\gamma)^2 m^2 t + (1-\gamma)^2 m^2 \ell + O(tm)} \\
&\leq 2^{-(1-\gamma)^2 m n' + (1-\gamma)^2 (m^2 \ell + m^2) + O(n')},
\end{aligned}$$

which completes the proof. ∎

We need one final piece of notation. For $1 \leq k \leq n$, we define the set of k-partial matrices – denoted by $\mathcal{P}_k$ – to be $\{\pm 1, *\}$-valued matrices of the form (3). For any $n \times n$ $\{\pm 1\}$-valued matrix M, let M_k denote k-partial matrix obtained by restricting M. For any $1 \leq k \leq n$ and any $n \times n$ matrix N, we define:

$$\mathcal{S}_k(N) := \{P \in \mathcal{P}_k : P = M_k \text{ for some } M \text{ which is } N\text{-normal}\}.$$

In words, $\mathcal{S}_k(N)$ denotes all the possible k-partial matrices arising as restrictions of N-normal matrices. The following proposition is the main result of this section.

Proposition 3.10. *There exist absolute constants $\beta, \delta > 0$ such that for any $n \times n$ matrix N,*

$$|\mathcal{S}_{\beta n}(N)| \leq 2^{(2\beta - \beta^2) n^2 - \delta n^2 + o(n^2)}.$$

Given this proposition, it is immediate to obtain a non-trivial upper bound on the number of $\{\pm 1\}$-valued N-normal matrices. Indeed, any N-normal matrix must be an extension of a matrix in $\mathcal{S}_{\beta n}(N)$; on the other hand, any matrix in $\mathcal{S}_{\beta n}(N)$ can be extended to at most $2^{(1-\beta)^2 n^2}$ N-normal matrices (as $D_{\beta n}$ is an $n(1-\beta) \times n(1-\beta)$ $\{\pm 1\}$-valued matrix). Hence, the number of N-normal matrices is at most $2^{(2\beta - \beta^2) n^2 - \delta n^2 + (1-\beta)^2 n^2} = 2^{(1-\delta) n^2 + o(n^2)}$.

Proof. For any m-partial matrix P and for any $1 \leq k \leq m$, let $T_k(P)$ denote the $k \times 2(n-k-1)$ matrix obtained from P as in (5). We will estimate the size of $\mathcal{S}_{\beta n}(N)$ by considering the following two cases.

First, we bound the number of partial matrices P in $\mathcal{P}_{\beta n}$ such that for some $\beta n/2 \leq k \leq \beta n$, $T_k(P)$ does not admit a $(\gamma k, \ell_k)$-rank-partition, where $\ell_k = n'/2k, n' = 2(n-k-1)$, and $0 < \gamma < 1$ is some constant to be chosen later. For this, note that Proposition 3.8 shows that there are at most

$$2^{kn' - (1-\gamma)^2 kn' + (1-\gamma)^2 (k^2 \ell_k + k^2) + O(n')} = 2^{kn' - (1-\gamma)^2 kn'/4 + O(n')}$$

choices for such a $T_k(P)$, provided $k < n'/4$, which holds for (say) $\beta < 1/4$. Since the remaining unknown entries of P which are not in $T_k(P)$ are $\{\pm 1\}$-valued, this shows that the number of βn partial matrices satisfying this first case is bounded above by

$$2^{(2\beta - \beta^2) n^2 - n^2 (1-\gamma)^2 \beta (1-\beta)/4 + o(n^2)},$$

for all $\beta < 1/4$.

Second, we bound the number of partial matrices $P \in \mathcal{S}_{\beta n}(N)$ which have the additional property that $T_k(P)$ admits a $(\gamma k, \ell_k)$-rank-partition for all $\beta n/2 \le k \le \beta n$. In this case, Theorem 1.11 shows that for any $\beta n/2 \le k \le \beta n$, the number of $\{\pm 1\}$-valued solutions to (5) $T_k(P)x_k = N'$ is at most

$$2^{2(n-k-1)} \ell_k^{-\gamma k/2} \le 2^{2(n-k)-\frac{\gamma k}{2}\log_2 \frac{n}{2k}}, \tag{6}$$

where in the last inequality, we have used $2(n-k-1) \ge n$ for all $k \le \beta n$, which is certainly true for $\beta < 1/4$. In other words, for a fixed $T_k(P)$, there are at most $2^{2(n-k-1)-\frac{\gamma k}{2}\log_2 \frac{n}{2k}}$ ways to extend it to $T_{k+1}(P')$ for some $P' \in \mathcal{S}_{\beta n}(N)$. Hence, it follows that the number of matrices in $\mathcal{S}_{\beta n}(N)$ with this additional property (stated at the beginning of the paragraph) is at most:

$$2^{(2\beta-\beta^2)n^2 - \sum_{k=\beta n/2}^{\beta n} \frac{\gamma k}{2}\log \frac{n}{2k}} \le 2^{(2\beta-\beta^2)n^2 - n^2\gamma\beta^2 \log_2(1/2\beta)/8 + o(n^2)},$$

for $\beta < 1/4$. Combining these two cases completes the proof. □

Remark 3.11. In particular, if we take $\gamma = 3/4$, it follows that for β sufficiently small (say $\beta \le 2^{-10}$), we can take $\delta \ge \beta^2$.

4 Acknowledgements

V.J. would like to thank Ethan Yale Jaffe for his insightful comments. A.F. and Y.Z. would like to thank Gwen McKinley, Guy Moshkovitz, and Clara Shikhelman for helpful discussions at the initial stage of this project.

References

[1] SS Agaian. *Hadamard matrices and their applications*, volume 1168. Springer, 2006.

[2] Martin Aigner, Günter M Ziegler, Karl H Hofmann, and Paul Erdos. *Proofs from the Book*, volume 274. Springer, 2010.

[3] Warwick De Launey and David A Levin. A Fourier-analytic approach to counting partial Hadamard matrices. *Cryptography and Communications*, 2(2):307–334, 2010.

[4] Andrei Deneanu and Van Vu. Random matrices: Probability of Normality. *arXiv preprint arXiv:1711.02842*, 2017.

[5] Rick Durrett. *Probability: theory and examples*. Cambridge University Press, 2010.

[6] Paul Erdös. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[7] CG Esseen. On the Kolmogorov-Rogozin inequality for the concentration function. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 5(3):210–216, 1966.

[8] Jacques Hadamard. Resolution d'une question relative aux determinants. *Bull. des sciences math.*, 2:240–246, 1893.

[9] G Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.

[10] A Hedayat, Walter Dennis Wallis, et al. Hadamard matrices and their applications. *The Annals of Statistics*, 6(6):1184–1238, 1978.

[11] Kathy J Horadam. *Hadamard matrices and their applications*. Princeton University Press, 2012.

[12] Ralph Howard. Estimates on the concentration function of sets in $\mathbb{R}^d$: Notes on lectures of Oskolkov.

[13] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random±1-matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, 1995.

[14] Andrew M Odlyzko. On subspaces spanned by random selections of ±1 vectors. *Journal of Combinatorial Theory, Series A*, 47(1):124–133, 1988.

[15] Raymond EAC Paley. On orthogonal matrices. *Journal of Mathematics and Physics*, 12(1-4):311–320, 1933.

[16] Boris Alekseevich Rogozin. On the increase of dispersion of sums of independent random variables. *Theory of Probability & Its Applications*, 6(1):97–99, 1961.

[17] Mark Rudelson and Roman Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617, 2014.

[18] Walter Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.

[19] Jennifer Seberry and Mieko Yamada. Hadamard matrices, sequences, and block designs. *Contemporary design theory: a collection of surveys*, pages 431–560, 1992.

[20] Terence Tao and Van Vu. On the singularity probability of random Bernoulli matrices. *Journal of the American Mathematical Society*, 20(3):603–628, 2007.

[21] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

A Completing the proof of Theorem 1.18

We now show how to combine the strategy of Deneanu and Vu with Section 3.3 in order to prove Theorem 1.18. We begin with a few definitions.

Definition A.1. Let S_n denote the symmetric group on n letters. For any $\sigma \in S_n$ and for any $n \times n$ matrix M, we define

$$M_\sigma := P_\sigma M P_\sigma^T,$$

where P_σ is the permutation matrix representing σ. In other words, M_σ is the matrix obtained from M by permuting the row and columns according to σ.

The previous definition motivates the following equivalence relation $\sim$ on the set of $n \times n$ matrices: given two $n \times n$ matrices M and M', we say that $M \sim M'$ if and only if there exists $\sigma \in S_n$ such that $M' = M_\sigma$. The next definition isolates a notion of normality which is invariant under this equivalence relation.

Definition A.2. Let N be a fixed $n \times n$ matrix. We say that an $n \times n$ matrix M is *N-normal-equivalent* if and only if there exists some $\sigma \in S_n$ such that $MM^T - M^T M = N_\sigma$.

By definition, it is clear that for any N and for any $M \sim M'$, M is N-normal-equivalent if and only if M' is N-normal equivalent. On the other hand, as we will see below, one can find a permutation ρ_M for any matrix M such that for the matrix $M' := M_{\rho_M}$, the ranks of many of the matrices $T_k(M'), 1 \leq k \leq n$ are large, where $T_k(M')$ denotes the matrix from (5). Therefore, by Odlyzko's lemma, we will be able to obtain good upper bounds on the probability of the random matrix $M' := M_{\rho_M}$ being C-normal, for any fixed C, which then translates to an upper bound on the probability of N-normality of M as follows: for any fixed N,

$$\begin{aligned}
\Pr\left[M \text{ is } N\text{-normal}\right] &\leq \Pr\left[M \text{ is } N\text{-normal-equivalent}\right] \\
&= \Pr\left[M_{\rho_M} \text{ is } N\text{-normal-equivalent}\right] \\
&\leq \sum_{\sigma \in S_n} \Pr\left[M_{\rho_M} \text{ is } N_\sigma\text{-normal}\right] \\
&\leq n! \sup_{\sigma \in S_n} \Pr\left[M_{\rho_M} \text{ is } N_\sigma\text{-normal}\right] \\
&\leq 2^{o(n^2)} \sup_{C \in \mathcal{M}_{n\times n}} \Pr\left[M_{\rho_M} \text{ is } C\text{-normal}\right],
\end{aligned}$$

where $\mathcal{M}_{n\times n}$ denotes the set of all $n \times n$ matrices, and we have used the fact that $n! = 2^{o(n^2)}$. Hence, it suffices to provide a good *uniform* upper bound on the probability that the random matrix M_{ρ_M} is N-normal for any fixed N.

To make the special property of the matrix M_{ρ_M} precise, we need the following functions, defined for all integers $1 \leq s \leq t \leq n$:

$$R_{s,t}(i) := \begin{cases} i & \text{if } 0 < i \leq s \\ s & \text{if } s < i \leq t \\ s + t - i & \text{if } t < i \leq 2n - s - t \\ 2n - 2i & \text{if } 2n - s - t < i \leq n. \end{cases}$$

The next proposition is one of the key ideas in the proof of Deneanu and Vu.

Proposition A.3 (Permutation Lemma, Lemma 3.5 in [4])**.** *Let M be any (fixed) $n \times n$ matrix. Then, there exist $s, t \in \mathbb{N}$ and $\rho_M \in S_n$ such that M_{ρ_M} satisfies:*

$$\operatorname{rank}(T_i(M_{\rho_M})) = R_{s,t}(i) \text{ for all } 1 \leq i \leq n.$$

For a fixed matrix N, let $\mathcal{N}_{s,t}(N)$ denote the set of $\{\pm 1\}$-valued $n \times n$ matrices M such that M is N-normal, and $\operatorname{rank}(T_i(M)) = R_{s,t}(i)$ for all $i \in [n]$. Then, it follows from the previous proposition that M_{ρ_M} is N-normal if and only if $M_{\rho_M} \in \bigcup_{1 \leq s \leq t \leq n} \mathcal{N}_{s,t}(N)$. This, in turn, can happen only if M itself is one of the at most $n! \sum_{s,t} |\mathcal{N}_{s,t}(N)|$ matrices obtained by permuting the rows and columns of $\mathcal{N}_{s,t}(N)$. Hence, it suffices to provide a good upper bound on $|\mathcal{N}_{s,t}(N)|$ *uniformly* in N, s and t.

Deneanu and Vu note (Observation 3.7 in [4]) that $\mathcal{N}_{s,t}$ is empty unless the following restrictions on s and t are met:

- $1 \leq s \leq 2n/3$, and
- $\frac{s}{2} < n - t < s$.

Then, letting

$$\beta := \sup\{c > 0 : |\mathcal{N}_{s,t}(N)| \leq 2^{-(c+o(1))n^2} \text{ for all } s, t, N\},$$

and for some small fixed (but otherwise arbitrary) $\epsilon > 0$, letting

$$\alpha := \beta - \epsilon,$$

they show (Lemmas 5.1 and 5.4 in [4]) the following:

$$|\mathcal{N}_{s,t}(N)| \leq 2^{n^2} \times \begin{cases} \min\left(2^{g_1(n,s,t)+o(n^2)}, 2^{f(\alpha,n,s,t)+o(n^2)}\right) & s \leq \frac{n}{2} \\ \min\left(2^{g_2(n,k,t)+o(n^2)}, 2^{f(\alpha,n,k,t)+o(n^2)}\right) & s \geq \frac{n}{2} \end{cases}, \tag{7}$$

where

$$\begin{aligned} f(\alpha,n,s,t) &:= (1-\alpha)t^2 - s^2/2 - n^2 + ns \\ g_1(n,s,t) &:= t^2 - 3s^2 + 2sn + st - 2nt \\ g_2(n,s,t) &:= n^2 + s^2 + t^2 + st - 2sn - 2nt. \end{aligned}$$

Finally, they analyze (7) to obtain their bound on β. For this, they note that since for fixed s, both g_1 and g_2 are decreasing functions of t while f is an increasing function of t, the worst restrictions on β (i.e. those requiring β to be small) can only be obtained in one of the following six cases:

1. $t = n - s$ and $s \leq n/2$, which places the restriction $\beta \leq 0.425$;
2. $t = n - s/2$ and $s \leq n/2$, which places the restriction $\beta \leq 0.307$;
3. $t = n - s/2$ and $s \geq n/2$, which places the restriction $\beta \leq 0.3125$;
4. $t = s$ and $s \geq n/2$, which places the restriction $\beta \leq 0.323$;
5. $f(\alpha,n,s,t) = g_2(n,s,t)$ and $s \geq n/2$, which places the restriction $\beta \leq 0.307$; and finally,
6. $f(\alpha,n,s,t) = g_1(n,s,t)$ and $s \leq n/2$, which places the worst restriction $\beta \leq 0.302$.

Hence, any improvement in Case 6 translates to an overall improvement in their bound. Moreover, note that for $1 \leq s \leq \frac{n}{10}$, Case 6 only leads to the restriction $\beta \leq 0.7$. Therefore, it suffices to improve Case 6 for $\frac{n}{10} \leq s \leq \frac{n}{2}$. We will do this using Proposition 3.10.

We start by showing how to deduce the upper bound $g_1(n,s,t)$, as in [4]. For any $0 \leq k \leq n$, we define

$$\mathcal{S}_{k,(s,t)}(N) := \{P \in \mathcal{P}_k : P = M_k \text{ for some } M \in \mathcal{N}_{s,t}(N)\},$$

where recall that $\mathcal{P}_k$ denotes the set of k-partial matrices, and M_k denotes the k-partial matrix associated to M. By definition, the number of ways to extend any k-partial matrix in $\mathcal{S}_{k,(s,t)}(N)$ to a $(k+1)$-partial matrix in $\mathcal{S}_{k+1,(s,t)}(N)$ is at most the number of $\{\pm 1\}$-valued solutions to : $T_k x_k = N'_k$, which is at most $2^{\max\{2(n-k-1)-\mathrm{rank}(T_k),0\}}$ by Odlyzko's lemma (Lemma 1.5). Hence, it follows that the total number of matrices in $\mathcal{N}_{s,t}(N)$ is at most

$$\begin{aligned} |\mathcal{N}_{s,t}(N)| &\leq 2^{o(n^2)} \prod_{k=0}^{n} 2^{\max\{2(n-k-1)-\mathrm{rank}(T_k),0\}} \\ &= 2^{o(n^2)} \prod_{k=0}^{n} 2^{\max\{2(n-k)-R_{s,t}(i),0\}} \\ &= 2^{g_1(n,s,t)+o(n^2)}, \end{aligned}$$

where the second equality follows from the definition of $\mathcal{N}_{s,t}(N)$ and the last equality follows by direct computation.

To obtain our improvement, we note that above computation may be viewed in the following two steps:

- $|\mathcal{N}_{s,t}(N)| \leq 2^{o(n^2)}|\mathcal{S}_{\beta n,(s,t)}(N)| \prod_{k=\beta n+1}^{n} 2^{\max\{2(n-k-1)-\mathrm{rank}(T_k),0\}}$, which is true for any $0 < \beta < 1$
- For $0 < \beta < 1$ such that $\beta n < s$,

$$\begin{aligned} |\mathcal{S}_{\beta n,(s,t)}(N)| &\leq 2^{o(n^2)} \prod_{k=0}^{\beta n} 2^{\max\{2(n-k-1)-\mathrm{rank}(T_k),0\}} \\ &= 2^{o(n^2)} \prod_{k=0}^{\beta n} 2^{2(n-k)-\mathrm{rank}(T_k)} \\ &= 2^{o(n^2)} \prod_{k=0}^{\beta n} 2^{2n-3k} \\ &= 2^{(2\beta-\beta^2)n^2-\beta^2 n^2/2+o(n^2)} \end{aligned}$$

In particular, by our assumption on s, we know that this holds for (say) $\beta = 2^{-10}$.

However, by Proposition 3.10 and Remark 3.11, we already know that for $\beta = 2^{-10}$,

$$\begin{aligned} |\mathcal{S}_{\beta n,(s,t)}(N)| &\leq |\mathcal{S}_{\beta n}(N)| \\ &\leq 2^{(2\beta-\beta^2)n^2-\beta^2 n^2+o(n^2)} \end{aligned}$$

Using this improved bound in the previous computation, we get that $|\mathcal{N}_{s,t}(N)| \leq 2^{h(n,s,t)+o(n^2)}$, where

$$h(n,s,t) = g_1(n,s,t) - \frac{\beta^2 n^2}{2}.$$

Hence, we have showed that Case 6 can be replaced by the following two cases:

Case 6.1 $f(\alpha,n,s,t) = g_1(n,s,t)$ and $s \leq n/10$

Case 6.2 $f(\alpha,n,s,t) = h(n,s,t)$ and $n/10 \leq s \leq n/2$,

each of which place a restriction on β which must be larger than the constant c_{DV} obtained in [4]. This completes the proof of Theorem 1.18.