The singularity probability of a random symmetric matrix is exponentially small

Campos, Marcelo; Jenssen, Matthew; Michelen, Marcus; Sahasrabudhe, Julian

THE SINGULARITY PROBABILITY OF A RANDOM SYMMETRIC MATRIX IS EXPONENTIALLY SMALL

MARCELO CAMPOS, MATTHEW JENSSEN, MARCUS MICHELEN AND JULIAN SAHASRABUDHE

Abstract. Let A be drawn uniformly at random from the set of all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. We show that

$$\mathbb{P}(\det(A) = 0) \leqslant e^{-cn},$$

where $c > 0$ is an absolute constant, thereby resolving a well-known conjecture.

1. Introduction

Let B be a random $n \times n$ matrix whose entries are chosen independently and uniformly from $\{-1, 1\}$. It is an old problem, likely stemming from multiple origins, to determine the probability that B is singular. While a moment's thought reveals the lower bound of $(1+o(1))2n^2 2^{-n}$, the probability that two rows or columns are equal up to sign, establishing the corresponding *upper bound* remains an extremely challenging open problem. Indeed, it is widely believed that

$$\mathbb{P}(\det(B) = 0) = (1 + o(1))2n^2 2^{-n}. \tag{1}$$

While this precise asymptotic has so far eluded researchers, a huge amount is now known about this fascinating problem. The first advances were made by Komlós [22] in the 1960s, who showed that the singularity probability is $O(n^{-1/2})$ (see also [23] and [3]).

Nearly 30 years later Kahn, Komlós and Szemerédi [19], in a remarkable paper, showed that the singularity probability is exponentially small. At the heart of their paper is an ingenious argument with the Fourier transform that allows them to give vastly more efficient descriptions of "structured" subspaces of $\mathbb{R}^n$ that are spanned by $\{-1, 1\}$-vectors. Their method was then developed by Tao and Vu [44, 45] who showed a bound of $(3/4 + o(1))^n$, by proving an interesting link between the ideas of [19] and the structure of set addition and, in particular, Freiman's theorem. This trajectory was then developed further by Bourgain, Vu and Wood [5], who proved a bound of $(2^{-1/2} + o(1))^n$, and by Tao and Vu [49], who pioneered the development of "inverse Littlewood-Offord theory," now an integral aspect of random matrix theory (see Section 1.1).

In 2007, Rudelson and Vershynin, in an important and influential paper [33], gave a different proof of the exponential upper bound on the singularity probability of B. The key idea was to construct efficient ε-nets for points on the sphere that have special anti-concentration properties and are thus more likely to be in the kernel of B. This then

The first named author is partially supported by CNPq.

led them to prove an elegant inverse Littlewood-Offord type result, inspired by [49], in a geometric setting.

This perspective was then developed further in the 2018 breakthrough work of Tikhomirov [50], who proved

$$\mathbb{P}(\det(B) = 0) = (1/2 + o(1))^n,$$

thereby essentially proving the conjectured upper bound. One of the key innovations in [50] was to adopt a probabilistic viewpoint of the (discretized) sphere: instead of directly proving that efficient nets exist by latching onto some sort of structure, he shows that the probability of randomly selecting a "structured" point on the discrete sphere is incredibly unlikely. While this change in perspective may not immediately sound useful, Tikhomirov's "inversion of randomness" gives him access to a whole host of probabilistic tools.

Another major advance on the problem was made recently by Jain, Sah and Sawhney [17,18], who (building on the recent work of Litvak and Tikhomirov [26]), proved the natural analogue of (1) for random matrices with independent entries chosen from a finite set S, for any *non-uniform* distribution on S. For the case of $\{-1, 1\}$-matrices, however, they do not improve on the bound of Tikhomirov.

While the problem for matrices B with all entries independent is now very well understood, the situation for *symmetric* random matrices remains somewhat more mysterious. Indeed all of the previously mentioned works on random matrices depend deeply on the fact that the entries of B are independent, and often treat B as n independent copies of a row, thus allowing for an essentially "one-dimensional" treatment of the problem. In the symmetric case, no such perspective is available.

Let A be a random $n \times n$ symmetric matrix, uniformly drawn from all symmetric matrices with entries in $\{-1, 1\}$. Again, it is generally believed that $\mathbb{P}(\det A = 0) = \Theta(n^2 2^{-n})$ (see, e.g. [8, 53, 54]) but progress has come more slowly. The problem of showing that A is almost surely non-singular goes back, at least, to Weiss in the early 1990s but was not resolved until 2005 by Costello, Tao and Vu [8], who obtained the bound

$$\mathbb{P}(\det(A) = 0) \leqslant n^{-1/8+o(1)}. \tag{2}$$

The first super-polynomial bounds were obtained by Nguyen [31] and, simultaneously, Vershynin [51], the latter obtaining a bound of the form $\exp(-n^c)$. Nguyen [31] developed the quadratic Littlewood-Offord theory introduced in [8], while Vershynin [51] worked in the geometric framework pioneered in his work with Rudelson [33–35].

In 2019, a more combinatorial perspective for inversion of random discrete matrices was introduced by Ferber, Jain, Luh and Samotij [11] and applied by Ferber and Jain [10] to show

$$\mathbb{P}(\det A = 0) \leqslant \exp(-cn^{1/4}(\log n)^{1/2}).$$

In a similar spirit, Campos, Mattos, Morris and Morrison [7] then improved this bound to

$$\mathbb{P}(\det A = 0) \leqslant \exp(-cn^{1/2}), \tag{3}$$

by proving a "rough" inverse Littlewood-Offord theorem, inspired by the theory of hypergraph containers (see [2, 40]). This bound was then improved by Jain, Sah and Sawhney [16],

who improved the exponent to $-cn^{1/2}\log^{1/4} n$, and, simultaneously, by the authors of this paper [6] who improved the exponent to $-c(n\log n)^{1/2}$.

The convergence of these results onto the exponent of $-c(n\log n)^{1/2}$ is no coincidence and in fact represents a natural barrier in the problem. Indeed, all of the results up to now have treated "structured" vectors by only using the top-half of the matrix (i.e. the half above the diagonal), which conveniently consists of independent entries. However, as pointed out in [7], if one is restricted to working in the top-half of A one cannot obtain an exponent better than $-c(n\log n)^{1/2}$. Thus to get beyond this obstruction, somehow the randomness of the matrix must "reused".

In this paper we prove an exponential upper-bound on the singularity probability of a symmetric random matrix, thereby breaking though this barrier and giving the optimal bound, up to the constant in the exponent.

Theorem 1.1. *Let A be uniformly drawn from all $n\times n$ symmetric matrices with entries in $\{-1,1\}$. Then*

$$\mathbb{P}(\det(A)=0)\leqslant e^{-cn}, \tag{4}$$

where $c>0$ is an absolute constant.

The main technical innovations of this paper are a new inverse Littlewood-Offord type theorem for "conditioned" random walks and a new "inversion of randomness" technique that allows us to "reuse" the randomness of our matrix by pushing some of the randomness onto the random selection of a vector from our discretized sphere. In fact, there is a delicate tradeoff between these two ingredients; a loss in the second ingredient allows for an improvement in the first, *unless* some specific "arithmetic" structure arises (see Section 1.2).

We note that there is a natural sister problem regarding the probability that the *least singular value* of A is at most ε. Our methods also apply to this problem and we will detail these results in a forthcoming paper, where more general coefficient distributions are also treated. In this paper we have opted to sacrifice generality to reduce clutter and technicality and, hopefully, to make the new ideas as clear as possible.

1.1. **Inverse Littlewood-Offord theory.** For $v\in\mathbb{R}^n$, we define the concentration function (one of several to come) as

$$\rho(v):=\max_{b\in\mathbb{R}}\mathbb{P}\left(\sum_{i=1}^n \varepsilon_i v_i=b\right),$$

where $\varepsilon_1,\dots,\varepsilon_n\in\{-1,1\}$ are uniform and independent. The study of $\rho(v)$ goes back at least to the classical work of Littlewood and Offord [24, 25] on the zeros of random polynomials, but perhaps begins in earnest with the beautiful 1945 result of Erdős [9]: if $v\in\mathbb{R}^n$ has all non-zero coordinates then

$$\rho(v)\leqslant 2^{-n}\binom{n}{\lfloor n/2\rfloor}=O(n^{-1/2}).$$

This was then developed by Szemerédi and Sárközy [39], who showed that if all of the v_i are *distinct* then one can obtain the much stronger bound of $O(n^{-3/2})$, and by Stanley [41] who

determined the *exact* maximum, using algebraic tools. A higher-dimensional version of this problem also received considerable attention (going under the name of *the* Littlewood-Offord problem) and was studied by several authors [14, 20, 21, 38] before it was ultimately resolved in the work of Frankl and Füredi [13] (see also [47]).

Of these early results, the most important for us here is the work of Halász [15] who made an important connection with the Fourier transform to prove (among other things) the following beautiful result: if there are N_k solutions to $x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$ among the entries of v, then $\rho(v) = O(n^{-2k-1/2} N_k)$.

More recently the question has been turned on its head by Tao and Vu [49], who pioneered the study of "inverse" Littlewood-Offord theory. They suggested that if $\rho(v)$ is "large" then v must exhibit some particular arithmetic structure. For example, Tao and Vu [46, 49], and Nguyen and Vu [30, 32] proved that if v is such that $\rho(v) > n^{-C}$ then $O(n^{1-\varepsilon})$ of the elements v_i of v can be efficiently covered with a generalized arithmetic progression of rank $r = O_{\varepsilon,C}(1)$.

While these results provide a very detailed picture in the range $\rho(v) > n^{-C}$, they begin to break down[1] if $\rho(v) = n^{-\omega(1)}$ and therefore are of limited direct use in showing that the singularity probability is exponentially small. Inverse results which work for smaller ρ bring us to the "counting" Littlewood-Offord theorem of Ferber, Jain, Luh and Samotij [11], and the "weak" inverse Littlewood-Offord theorems of Campos, Mattos, Morris and Morrison [7] and of the present authors in [6], which are useful for $\rho(v)$ as small as $\exp(-c(n \log n)^{1/2})$, but afford less structure.

Our novel inverse Littlewood-Offord theorem in this paper is most similar to that of Rudelson and Vershynin [33], also developed in [51] and [16], who showed that if $\rho(v) \gg e^{-cn}$ then there exists $\phi > 0$ with[2] $\phi \approx 1/\rho(v)$ for which the dilated vector $\phi \cdot v$ is exceptionally close to the integer lattice $\mathbb{Z}^n$. These Littlewood-Offord theorems, styled after Rudelson-Vershynin, tend to be a little bit subtler; instead of determining the structure of the whole vector, we only show there is some "correlation" with the rigid object $\mathbb{Z}^n$.

To state our inverse Littlewood-Offord theorem, we formulate an important notion introduced by Rudelson and Vershynin. We switch to working in $\mathbb{R}^d$, for $d \approx cn$, hinting at the later context of these results. If $A \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$ then define $d(x, A) := \inf_{a \in A}\{\|x - a\|_2\}$. Now, for $\alpha \in (0, 1)$, define the *least common denominator* of a vector $v \in \mathbb{R}^d$ to be the smallest $\phi > 0$ for which $\phi \cdot v$ is within $\sqrt{\alpha d}$ of a non-zero integer point. That is,

$$D_\alpha(v) = \inf\left\{\phi > 0 : \ d(\phi \cdot v, \mathbb{Z}^d \setminus \{0\}) \leqslant \sqrt{\alpha d}\right\}.$$

Note here that we have excluded the origin from $\mathbb{Z}^d$ in the definition since $\phi \cdot v \approx 0$ does not tell us any interesting about v. Indeed, given *any* $v \in \mathbb{S}^{d-1}$, one could always set $\phi < \sqrt{\alpha d}$ and obtain $d(\phi \cdot v, \mathbb{Z}^d) \leqslant d(\phi \cdot v, 0) \leqslant \sqrt{\alpha d}$, and so this degenerate case needs to be excluded somehow. In fact, in the course of the paper, we will work with a slightly different non-degeneracy condition (see (19)).

[1] Technically these results break down if $\rho(v) < n^{-\log\log n}$.

[2] In what follows, we will be somewhat vague with our use of $\approx$.

Our Littlewood-Offord theorem shows that a similar conclusion to that of [33] can be obtained in the presence of a large number ($k \approx n$) of additional "soft" constraints on the random walk. In particular we prove the following result, which is in fact weaker than what we really need (see Lemma 3.1), but captures its essence. We say that a random vector with entries in $\{-1, 0, 1\}$ is μ-lazy if each entry is independent and is equal to 0 with probability $1 - \mu$ and is equal to each of $-1, 1$ with probability $\mu/2$.

Theorem 1.2. *There exist $R, c_1, c_2 > 0$, for which the following holds for every $d \in \mathbb{N}$, $\alpha \in (0, 1)$, $0 \leqslant k \leqslant c_1 \alpha d$ and $t \geqslant \exp(-c_1 \alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \ldots, w_k \in \mathbb{S}^{d-1}$ be orthogonal and let W be the matrix with rows $w_1, \ldots, w_k$.*

If $\tau \in \{-1, 0, 1\}^d$ is a 1/4-lazy random vector and

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \text{ and } \|W\tau\|_2 \leqslant c_2\sqrt{k}\right) \geqslant Rte^{-c_1 k}, \tag{5}$$

then $D_\alpha(v) \leqslant 16/t$.

Here we interpret $\|W\tau\|_2 \leqslant c_2\sqrt{k}$ as encoding the "soft" constraints and $|\langle \tau, v\rangle| \leqslant t$ as the "hard" constraint. It is useful to think of $t \approx \rho(v)$, although we actually set t relative to a related notion tailored specifically to our application.

To understand the quantitative aspect of Theorem 1.2, it is best to consider the contrapositive of Theorem 1.2, which roughly says that if v is "unstructured at scale t" (that is, $D_\alpha(v) > 16/t$) then the soft and hard constraints are roughly negatively dependent[3]. Indeed, if v is sufficiently "unstructured at scale t" then we might expect $\langle \tau, v\rangle$ to approximate a Gaussian and, in particular, to have

$$\mathbb{P}(|\langle \tau, v\rangle| \leqslant t) \approx t.$$

On the other hand, since $w_1, \ldots, w_k \in \mathbb{S}^{d-1}$ are orthogonal, it turns out that (see Lemma 5.7)

$$\mathbb{P}(\|W\tau\|_2 \leqslant c_2\sqrt{k}) \leqslant e^{-c_1 k},$$

where $c_1 > 0$ is a suitably small constant depending on $c_2 > 0$. If these two events were negatively dependent then we would expect a bound of

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \text{ and } \|W\tau\|_2 \leqslant c_2\sqrt{k}\right) \leqslant te^{-c_1 k}.$$

Theorem 1.2 says something *almost* as strong as this, giving the inequality up to a constant R and the value of c_1.

For us, the main difficulty lies in "decoupling" the soft and hard constraints, which is ultimately achieved by a somewhat complicated geometric argument on the Fourier side. However, we should point out that Theorem 1.2 is non-trivial even in the case of $k = 0$ and in fact reduces, in this case, to the inverse Littlewood-Offord result proved by Rudelson and Vershynin in [33].

[3]Here, we say events S, T are negatively dependent if $\mathbb{P}(S \cap T) \leqslant \mathbb{P}(S)\mathbb{P}(T)$.

In fact, the $k = 0$ case is useful for understanding the sort of structure that the conclusion $D_\alpha(v) < c/t$ provides. It is not hard to show that if one chooses $v \in \mathbb{S}^{n-1}$ very close to a point on the lattice $(Ct)\mathbb{Z}^n$, where $C \gg 1$, then v satisfies

$$\mathbb{P}(|\langle v, \tau\rangle| \leqslant t) \gg t. \tag{6}$$

Thus the inverse theorem of [33, 35] says, roughly speaking, that *all* vectors satisfying (6) must have this structure. Our Theorem 1.2 says the same is true even in the presence of a large number of additional constraints.

1.2. **Proof sketch and a new "inversion of randomness" technique.** Here we briefly sketch how our inverse Littlewood-Offord result is used alongside a novel scheme for "reusing randomness" to prove Theorem 1.1. As hinted at before, we will be helped along by treating the discretized sphere as a probability space, which will allow us to "recover" some of the randomness lost due to the symmetry of A. We keep our discussion here loose and impressionistic and we will take up our careful study in the following section.

Our first move will be to "locally replace" A with a random matrix M that has many of the entries zeroed out. This will allow us to untangle some of the more subtle and complicated dependencies and has the advantage that various associated Fourier transforms are non-negative. Indeed let[4]

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}, \tag{7}$$

where $d = cn$ and H_1 is a $(n-d) \times d$ random matrix with i.i.d. entries that are μ-lazy, meaning that $(H_1)_{i,j} = 0$ with probability $1 - \mu$ and $(H_1)_{i,j} = \pm 1$ with probability $\mu/2$. We stress here that we cannot "globally" replace A with M, and we may need to permute coordinates, depending on what part of the sphere we are working on.

We now follow the strategy of [33, 50] and partition the sphere $\mathbb{S}^{n-1}$ based on the anti-concentration properties of the various $v \in \mathbb{S}^{n-1}$. Indeed, for each $v \in \mathbb{S}^{n-1}$, we find a corresponding "scale" $\varepsilon \in (0, 1)$ for which

$$\mathbb{P}(\|Mv\|_2 < \varepsilon\sqrt{n}) \approx (L\varepsilon)^n, \tag{8}$$

where L is a large constant. Notice here that we have defined this "scale" relative to the symmetric matrix M, rather than A or $\rho(v)$, and so we expect it to capture the anti-concentration properties of v, specific to the matrix M. This ε should be interpreted as "the scale at which the anti-concentration properties of v just start to be felt", as we imagine gradually decreasing ε from 1 to 0. For example, if v is a random point on the sphere, it is not hard to see that v will typically have $\varepsilon \leqslant e^{-cn}$, which is in fact *so* small that we can safely ignore v (due to previous work). On the other hand, the constant vector $n^{-1/2}(1, \ldots, 1)$ will have $\varepsilon \approx n^{-1/2}$. Interestingly, this latter fact is not easy to establish rigorously, but is heuristically not hard to guess in analogy with the modified setting where M has iid entries.

[4]Here we use the notation $[n] := \{1, \ldots, n\}$; for a vector $v \in \mathbb{R}^n$ and $S \subseteq [n]$, we use the notation $v_S := (v)_{i\in S}$ and for a $n \times m$ matrix A, and $R \subseteq [m]$, we use the notation $A_{S\times R}$ for the $|S| \times |R|$ matrix $(A_{i,j})_{i\in S, j\in R}$.

We now study all vectors $v \in \mathbb{S}^{n-1}$ at a given scale $\varepsilon \geqslant e^{-cn}$. While this is an uncountable set, we build an efficient ε-net for these vectors in two steps. We first discretize the whole sphere by taking an ε-net for $\mathbb{S}^{n-1}$, which we call Λ_ε. We can then say something like

$$\mathbb{P}(Av = 0 \text{ for some } v \text{ at scale } \varepsilon) \leqslant \mathbb{P}\left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n} \text{ for some } v \in \Lambda_\varepsilon\right).$$

One's first instinct might be to simply union bound over all $v \in \Lambda_\varepsilon$; however it turns out that even the most efficient epsilon nets have $|\Lambda_\varepsilon| \approx (C/\varepsilon)^n$, which is too large to say anything.

The key insight here is that most of Λ_ε is not used when approximating $v \in \mathbb{S}^{n-1}$ *at scale* ε and so we can refine our net Λ_ε by discarding all vectors $w \in \Lambda_\varepsilon$ with $\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}) \ll (L\varepsilon)^n$. So if we let $\mathcal{N}_\varepsilon \subseteq \Lambda_\varepsilon$ be the collection of vectors with $\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n$, our problem reduces to showing that

$$|\mathcal{N}_\varepsilon| \leqslant L^{-2n}|\Lambda_\varepsilon| \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n, \tag{9}$$

which brings us to the technical heart of the paper (see Theorem 7.1). We point out that the factor of L^{-2n}, rather than L^{-n}, in (9) is important for us as it allows us to drown out the L^n coming from (8) *and* the factor C^n in (9), when we union bound over $\mathcal{N}_\varepsilon$.

To prove (9) we take a probabilistic perspective inspired by [50]; although we stress that our methods are considerably different. To show (9) it is enough to show, for $v \in \Lambda_\varepsilon$ chosen uniformly at random

$$\mathbb{P}_{v\in\Lambda_\varepsilon}\left(v \in \mathcal{N}_\varepsilon\right) \approx \mathbb{P}_{v\in\Lambda_\varepsilon}\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\right) \geqslant (L\varepsilon)^n\right) \leqslant L^{-2n}, \tag{10}$$

(see Lemma 7.3, for the rigorous statement). To get a feel for how we tackle this, let us consider the event $\|Mv\|_2 \leqslant \varepsilon n^{1/2}$. Indeed recalling (7), the definition of M, we have that

$$Mv = \begin{bmatrix} H_1 v_{[d]} \\ H_1^T v_{[d+1,n]} \end{bmatrix}$$

and so to control the event $\|Mv\|_2 \leqslant \varepsilon\sqrt{n}$, it is enough to control the intersection of events $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ and $\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2}$. Note that if we simply ignore the second event and bound

$$\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon n^{1/2}) \leqslant \mathbb{P}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}),$$

we land in a situation very similar to previous works; where half of the matrix is neglected entirely and we are thus limited by the $(n\log n)^{1/2}$ obstruction, described above. So to overcome this barrier, we need to control these two events simultaneously.

The key idea here is to use the *randomness in* H_1 to control the event $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ and to use the *randomness in* $v \in \Lambda_\varepsilon$ to control the event $\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2}$. To get this to work, we crucially partition the outcomes in H_1, based on a robust notion of rank. Indeed, let

$$\mathcal{E}_k = \left\{H_1 : \sigma_{d-k}(H_1) \geqslant c\sqrt{n} \text{ and } \sigma_{d-k+1}(H_1) < c\sqrt{n}\right\},$$

where $\sigma_1(H_1) \geqslant \cdots \geqslant \sigma_d(H_1)$ denote the singular values of H_1. We may then bound $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$ above by (only using the randomness in H_1, for the moment)

$$\sum_{k=0}^{d} \mathbb{P}_{H_1}\Big(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2} \,\Big|\, \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\Big) \cdot \mathbb{P}_{H_1}\Big(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\Big). \quad (11)$$

It is here that we can see the link with our inverse Littlewood-Offord theorem, Theorem 1.2, which we use (after a good deal of preparation) to bound the probabilities

$$\mathbb{P}_{H_1}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}, \mathcal{E}_k),$$

that appear in (11). The event $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ corresponds to the "hard" constraint $|\langle \tau, v\rangle| \leqslant t$ in Theorem 1.2, while the event $\mathcal{E}_k$ corresponds to the "soft" constraint $\|W\tau\|_2 \leqslant c_2\sqrt{k}$, where we think of τ as a single row of H_1. And so, after a certain amount of work with Theorem 1.2, we are able to conclude that

$$\mathbb{P}_{H_1}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n},\, \mathcal{E}_k) \leqslant (C\varepsilon e^{-ck})^{n-d} \quad (12)$$

unless $v_{[d]}$ is structured, in which case we do something different (and substantially easier). Thus, for all non-structured v, we have (11) is roughly at most

$$(C\varepsilon)^{n-d} \sum_{k=0}^{d} e^{-ck(n-d)} \mathbb{P}_{H_1}\Big(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2} \,\Big|\, \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\Big). \quad (13)$$

Up to this point, we have not appealed to the randomness in the choice of $v \in \Lambda_\varepsilon$, beyond demanding that v is non-structured. To see how we might take advantage of the randomness in v, let us consider the first moment of the quantity $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$, which we view as a random variable in v. Indeed, for $v \in \Lambda_\varepsilon$ taken uniformly at random, we show that

$$\mathbb{E}_{v\in\Lambda_\varepsilon}\, \mathbb{P}_{H_1}\left(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon\sqrt{n} \mid \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\right) \leqslant (C\varepsilon)^{d-k}. \quad (14)$$

We establish this bound by swapping expectations, and bounding the probabilities

$$\mathbb{P}_{v_{[d+1,n]}}(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2}), \quad (15)$$

where H_1 is a *fixed* matrix satisfying $\mathcal{E}_k \cap \{H_1 : \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}\}$. The idea here is that since H_1 has $d-k$ singular values of size $\approx n^{1/2}$, we should expect

$$\mathbb{P}_{v_{[d+1,n]}}(\|H v_{[d+1,k]}\|_2 \leqslant \varepsilon n^{1/2}) \approx (C\varepsilon)^{d-k}, \quad (16)$$

which is suggested, for example, by a Gaussian heuristic. This then results in the bound at (14). See Section 7.2 for details on this step. Putting (14) and (13) together, and using that $\varepsilon > e^{-cn}$, we have

$$\mathbb{E}_v\, \mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2}) \leqslant (C\varepsilon)^n.$$

Observe that the loss from the rank at (16) is compensated by the gain afforded by the extra constraint added to our Littlewood-Offord step.

While this is a good bound on the expectation, this is *not* enough for our purposes, as the first moment only tells us, via Markov, that

$$\mathbb{P}_{v\in\Lambda_\varepsilon}\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon n^{1/2}\right) \geqslant (L\varepsilon)^n\right) \leqslant L^{-n}\,,$$

falling short of our desired L^{-2n} bound.

So to prove our result, we instead study[5] the *second moment* of $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$,

$$\mathbb{E}_v\left(\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})\right)^2,$$

in much the same way, but with a few added technicalities.

To say a few words about how the second moment is different, we will see (Fact 7.7)

$$\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon n^{1/2}\right)\right)^2 \leqslant \mathbb{P}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \|H_2 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2} \text{ and } \|H^T v_{[d+1,n]}\|_2 \leqslant 2\varepsilon n^{1/2})\,,$$

where H_2 is an independent copy of H_1 and $H := [H_1, H_2]$ is the concatenation of these two matrices. We then proceed in much the same way as above, but treating H, in place of H_1, and carrying forward the two "hard" constraints resulting from the two copies of $v_{[d]}$. This explains the shape of our "real" inverse Littlewood-Offord theorem, Lemma 3.1, where we allow for these two hard constraints. Ultimately, we arrive at the bound

$$\mathbb{E}_v\left(\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})\right)^2 \leqslant (C\varepsilon)^{2n},$$

which implies the desired conclusion at (10).

1.3. **A few remarks on presentation.** This paper is roughly divided into three parts. The first part consists of Sections 3-5 which are dedicated to proving our conditioned inverse Littlewood-Offord result, Lemma 3.1, which is the "real" version of Theorem 1.2. These sections lay the groundwork for Section 6, where we prove Theorem 6.1, which is the only result we carry forward into later sections.

The second part of the paper consists of Section 7 and Section 8. In Section 7, we obtain our crucial bound on the size of our net $\mathcal{N}_\varepsilon$ using our novel "inversion of randomness" technique, as outline above. On the other hand, Section 8 contains the less exciting proof that $\mathcal{N}_\varepsilon$ is in fact a net for Σ_ε.

In the final section, Section 9, we pull together the various elements of this paper, state the reductions we will use from previous work and prove Theorem 1.1.

In most cases, we have highlighted the main results of each section at the start. So if one does not want to delve into the details of a particular element of the proof, one can simply inspect the top of the section to glean what is needed for going forward.

2. Central Definitions

We now turn to give a proper treatment of the proof, by laying out the key definitions that will concern us in this paper. We begin by partitioning the sphere $\mathbb{S}^{n-1}$ into "structured"

[5]Actually, we need a slight variant, where we cut out structured vectors.

and "unstructured" vectors. Formally, we set $\gamma = e^{-cn}$, for sufficiently small $c > 0$, and then define the "structured" vectors as

$$\Sigma := \left\{ v \in \mathbb{S}^{n-1} : \rho(v) \geqslant \gamma \right\}.$$

The invertibility of a random symmetric matrix on the set of "unstructured" vectors $v \in \mathbb{S}^{n-1} \setminus \Sigma$ is already well understood and so we can restrict our attention to this set of structured vectors. We refer the reader to Section 9 for the details here.

Following Rudelson and Vershynin [33], we make a further reduction to working with vectors that are reasonably "flat" on a large part of their support. For $D \subseteq [n]$, with $|D| = d$, we define

$$\mathcal{I}(D) := \left\{ v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D \right\}, \tag{17}$$

where $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants, fixed throughout the paper and defined in Section 2.1. We will set $d := c_0^2 n/2$, where c_0 is defined below in Section 2.1.

Now set

$$\mathcal{I} := \bigcup_{D \subseteq [n], |D| = d} \mathcal{I}(D).$$

The case of non-flat v is already taken care of in the work of Vershynin [51] (see Section 9) and so it is enough to work with $\mathcal{I} \cap \Sigma$. Since we will ultimately union bound over D, it is enough to work with $\mathcal{I}(D) \cap \Sigma$, for *some* fixed set D, and so, by symmetry it is enough to restrict our attention to vectors $v \in \mathcal{I}([d]) \cap \Sigma$.

Now, with this in mind, we further partition the set $\mathcal{I}([d]) \cap \Sigma \subseteq \mathbb{S}^{n-1}$, but for this we need to introduce another distribution on symmetric matrices. Define the probability space $\mathcal{M}_n(\mu)$ by defining $M \sim \mathcal{M}_n(\mu)$ to be the random matrix

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix},$$

where H_1 is a $(n-d)\times d$ random matrix with i.i.d. entries that are μ-lazy (that is, $(H_1)_{i,j} = 0$ with probability $1-\mu$ and $(H_1)_{i,j} = \pm 1$ with probability $\mu/2$). In fact, we will fix $\mu = 1/4$ throughout the paper.

Now, given $v \in \mathcal{I}([d])$ and $L > 0$, in the spirit of [50], we define the *threshold*

$$\mathcal{T}_L(v) = \sup \left\{ t \in [0,1] : \mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) \geqslant (4Lt)^n \right\},$$

or the "scale" of v, as we called it in Section 1.2. Observe carefully here that we are defining $\mathcal{T}_L$ *relative to the matrix* M, rather than our original distribution A.

We may now define our partition of $\mathcal{I}([d]) \cap \Sigma$. For $\varepsilon \in (0,1)$, let

$$\Sigma_\varepsilon := \{ v \in \mathcal{I}([d]) : \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon] \}.$$

We shall show (as it is not obvious) that indeed

$$\Sigma \cap \mathcal{I}([d]) \subseteq \bigcup_{\varepsilon > \gamma^4} \Sigma_\varepsilon.$$

With the definition of Σ_ε in hand, we are able to define $\mathcal{N}_\varepsilon$ which will be an efficient net for Σ_ε at scale ε. It turns out that *defining* this net is not hard, although showing that it satisfies the desired properties will be the main challenge of this paper. For this, we first define the *trivial net at scale* ε to be[6]

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

which is a natural net for $\mathcal{I}([d])$. Where $\mathcal{I}'(D)$ is similar to $\mathcal{I}(D)$ but with slightly looser constraints and relative to $\mathbb{R}^n$;

$$\mathcal{I}'(D) := \left\{v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in D\right\}.$$

Since we are only interested in approximating vectors in Σ_ε, we can get away with a significantly more efficient net. For this we introduce two more concentration functions. First, we define the *Lévy concentration function*: if X is a random vector taking values in $\mathbb{R}^n$, define

$$\mathcal{L}(X,t) := \max_{w \in \mathbb{R}^n} \mathbb{P}\left(\|X - w\|_2 \leqslant t\right).$$

Second, we define a variant of this concentration function for the uniform distribution on random symmetric matrices with capped operator norm[7].

$$\mathcal{L}_{A,op}(v,t) := \max_{w \in \mathbb{R}^n} \mathbb{P}\left(\{\|Av - w\|_2 \leqslant t\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right).$$

Here we are just cutting out the slightly irritating event that A has large operator norm. Intuitively this is an acceptable move as the probability that $\|A\| \geqslant 4\sqrt{n}$, is exponentially small (see Lemma 9.5), however some care is needed as we are mostly concerned with far less likely events.

We now introduce our nets $\mathcal{N}_\varepsilon$,

$$\mathcal{N}_\varepsilon := \left\{v \in \Lambda_\varepsilon : \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n\right\}.$$

The reader should view the lower bound $\mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n$ as the real core of this definition, while the upper bound for $\mathcal{L}_{A,op}$ is less important. The two main tasks of this paper will be to show that $\mathcal{N}_\varepsilon$ is indeed a net for Σ_ε (an easier task) and secondly that $|N_\varepsilon|/|\Lambda_\varepsilon|$ is smaller than $\approx L^{-2n}$, where L is a large constant.

2.1. **Discussion of constants and parameters.** We will treat the constants κ_0, κ_1 (seen at (17)) as absolute throughout the paper, and we allow other absolute constants $C, C', \cdots$ to depend on these exact quantities. In particular, we set $\kappa_0 = \rho$ and $\kappa_1 = \delta^{-1/2}/2$, where δ, ρ are as in Lemma 9.2 (which is a lemma from [51]). While we have not computed these constants, it would not be too much work to do so.

We also note our treatment of c_0, which, for most of the paper, will be presented as a parameter and dependencies involving c_0 will be explicitly noted. However, we will ultimately fix $c_0 = \min\{2^{-24}, \rho\delta^{1/2}\}$ where, again, δ, ρ are as in Lemma 9.2. Thus it is no harm for the reader to view c_0 as an absolute constant which is fixed throughout the paper. The reason for

[6]Here and throughout, $B_n(x,r)$ is the ℓ^2 ball centered at x with radius r.

[7]For a matrix A, we use the notation $\|A\| := \max_{x:\|x\|_2=1} \|A\|_2$ to denote the usual $2 \to 2$ operator norm.

the extra care with c_0 comes from its delicate relationship to d/n. Indeed, we will ultimately set $d := \lceil c_0^2 n/2 \rceil$.

Another point to note is our use of R, which represents related, but different constants throughout the paper. Roughly speaking, these related values of R increase as we get deeper into the proof.

3. Inverse Littlewood-Offord for conditioned random walks I: Statement of result and setting up the proof

This section is the first of three sections where we lay out and prove our main inverse Littlewood-Offord type theorem, Lemma 3.1, which works in the presence of a large number ($k \approx n$) of relatively soft constraints on our random walk. As mentioned before, the conclusion of our Littlewood-Offord theorem will be similar to that of Rudelson and Vershynin [33], who showed that vectors v, for which the random walk $\langle v, \tau \rangle$ concentrates, admit non-trivial least common denominators. As we will see, the proof of Lemma 3.1 is rather involved and consists mainly of a geometric argument on the Fourier side to "decouple" the many soft constraints from the few hard constraints.

Given a $2d \times \ell$ matrix W (which encodes these soft constraints on our walk, as in Theorem 1.2) and a vector $Y \in \mathbb{R}^d$, we define the Y-augmented matrix W_Y as

$$W_Y = \left[W, \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix}, \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} \right]. \tag{18}$$

Here $Y \approx v/t$ will be a re-scaled version of v from Theorem 1.2. We define, for $\alpha \in (0,1)$, the *least common denominator* of a vector $v \in \mathbb{R}^d$ to be

$$D_\alpha(v) := \inf \left\{ \phi > 0 : \ \|\phi \cdot v\|_{\mathbb{T}} \leqslant \min \left\{ \phi \|v\|_2/2, \sqrt{\alpha d} \right\} \right\}, \tag{19}$$

where $\|x\|_{\mathbb{T}} := \inf\{\|x-y\|_2 : y \in \mathbb{Z}^d\}$, for $x \in \mathbb{R}^d$, denotes the minimum distance to an integer point. Note the definition at (19) is a bit different from the definition presented in the introduction, in that the "non-degeneracy condition" is now $\|\phi \cdot v\|_{\mathbb{T}} \leqslant \phi \|v\|_2/2$. We will stick with this definition throughout the paper.

We let $\|A\|_{\mathrm{HS}}$ denote the Hilbert-Schmidt norm of a matrix A, that is, $\|A\|_{\mathrm{HS}}^2 := \sum_{i,j} |A_{i,j}|^2$ and for $\mu \in (0,1)$, $m \in \mathbb{N}$, define the m-dimensional μ-lazy random vector $\tau \sim \mathcal{Q}(m, \mu)$ to be the vector with independent entries $(\tau_i)_{i=1}^m$, satisfying

$$\mathbb{P}(\tau_i = -1) = \mathbb{P}(\tau_i = +1) = \mu/2 \qquad \text{and} \qquad \mathbb{P}(\tau_i = 0) = 1 - \mu.$$

We now state our main inverse Littlewood-Offord type theorem, which is our "real" (and strengthened) version of Theorem 1.2, from Section 1.1.

Lemma 3.1. *For $d \in \mathbb{N}$ and $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $t \geqslant \exp(-2^{-9}\alpha d)$. For $0 < c_0 \leqslant 2^{-24}$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\|_2 \geqslant 2^{-10} c_0/t$, let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$.*

If $\tau \sim \mathcal{Q}(2d, 1/4)$ and $D_\alpha(Y) > 16$ then

$$\mathcal{L}\left(W_Y^T \tau, c_0^{1/2} \sqrt{k+1} \right) \leqslant (Rt)^2 \exp(-c_0 k), \tag{20}$$

where $R = 2^{32}c_0^{-2}$.

Before we start working towards the proof of Lemma 3.1, we make a few informal remarks on its statement and its connection to Theorem 1.2. The main difference to note is that there are now two "hard" constraints encoded in the left-hand side of (20); these are, in the notation of Theorem 1.2,

$$|\langle (v, 0_{[d]}), \tau\rangle| < t \text{ and } |\langle (0_{[d]}, v), \tau\rangle| < t.$$

The "soft" constraints are now encoded as the columns $w_1, \dots, w_k$ of W.

To combine the "hard" and "soft" constraints into a single matrix inequality, we rescale v, thinking of $|\langle (v, 0_{[d]}), \tau\rangle| < t$ as $|\langle c_0^{1/2} t^{-1}(v, 0_{[d]}), \tau\rangle| < c_0^{1/2}$. This explains the scaling on Y, which is unusually written as $\|Y\|_2 \geqslant 2^{-10}c_0/t$, where t should be thought of a very small number $\approx e^{-cn}$.

The scaling of $D_\alpha(Y)$ in Lemma 3.1, in contrast with the statement of Theorem 1.2, is explained in a similar way. If $\phi \cdot Y \sim \mathbb{Z}^d$, where $\phi = O(1)$ then $(\phi/t) = O(1/t)$ satisfies $(\phi/t) \cdot v \sim \mathbb{Z}^d$, as we think of $Y \approx v/t$.

This also makes the numerology of Lemma 3.1 a little more transparent. If Y is a random vector with $\|Y\|_2 \approx 1/t$, we have $|Y_i| \approx t^{-1}n^{-1/2}$ and thus we expect the one dimensional random walk $\langle Y, \tau\rangle$ to have

$$\mathcal{L}\left(\langle Y, \tau\rangle, c_0^{1/2}\right) \approx t.$$

Thus we expect Y to have some special structure if $\mathcal{L}\left(\langle Y, \tau\rangle, c_0^{1/2}\right) \gg t$. On the other hand, for each w_i we expect that $|\langle w_i, \tau\rangle| \approx 1$ and, since the w_i must be "approximately orthogonal" (due to the assumption $\|W\| \leqslant 2$), we should expect

$$\mathcal{L}\left(W\tau, c_0^{1/2}\sqrt{k}\right) \approx e^{-ck},$$

being somewhat vague about this constant $c > 0$. Second, note that Lemma 3.1 is still interesting even in the case $k = 0$, where it is not hard to see that it reduces to

$$\mathcal{L}\left(\langle Y, \tau\rangle,\, c_0^{1/2}\right) \leqslant Rt,$$

whenever $D_\alpha(Y) \leqslant 16$, which is essentially the statement of the main inverse Littlewood-Offord theorem of Rudelson and Vershynin in [33].

Finally, we point out that the contrapositive of Lemma 3.1 is more conducive to the "inverse Littlewood-Offord" reading:

$$\text{if } \ \mathcal{L}(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}) \geqslant (Rt)^2 \exp(-c_0 k) \ \text{ then } \ D_\alpha(Y) \leqslant 16.$$

For the remainder of this section, we take some first steps towards the proof of Lemma 3.1. We first pass to the Fourier side and set up our problem there, describing our goal in terms of a certain "level set". We then make a first reduction, by getting some basic control on the fibers of this level set. In the following section, Section 4, we make a more significant reduction about the geometry of our level set. In Section 5 we prove the key Lemma 5.1, the statement of which is very similar to that of Lemma 3.1, but with a more complicated

quantity replacing the right-hand side of (20). Finally, with one further step, we conclude Section 5, with the proof of Lemma 3.1.

3.1. **Passing to the Fourier side.** To prove Lemma 3.1 we will prove the contrapositive; assume (20) fails and then obtain an upper bound on the least common denominator by finding a non-trivial $\phi > 0$ that satisfies $\phi = O(1)$ and $\|\phi \cdot Y\|_{\mathbb{T}} \leqslant \sqrt{\alpha d}$. Our first step in proving Lemma 3.1 is to use the lower bound in the negation of (20) to obtain a lower bound on a level set of an appropriate Fourier transform. This manoeuvre was pioneered by Halász [15] and has been a key step in all of the Fourier approaches to inverse Littlewood-Offord theory.

For a $2d \times \ell$ matrix W, we define the *W-level set*, for $t \geqslant 0$, to be

$$S_W(t) := \left\{\theta \in \mathbb{R}^\ell : \|W\theta\|_{\mathbb{T}} \leqslant \sqrt{t}\right\}$$

and we define γ_ℓ to be the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1} I_\ell)$ and I_ℓ denotes the $\ell \times \ell$ identity matrix.

The following Esseen-type lemma, allows us relate the quantity seen at the left-hand side of (20) with the Gaussian volume of a level-set.

Lemma 3.2. *Let $\beta > 0$, $\nu \in (0, 1/4]$, let W be a $2d \times \ell$ matrix and let $\tau \sim \mathcal{Q}(2d, \nu)$. Then there exists $m > 0$ so that*

$$\mathcal{L}(W^T\tau, \beta\sqrt{\ell}) \leqslant 2\exp\left(2\beta^2\ell - \nu m/2\right)\gamma_\ell(S_W(m)).$$

The proof of this Lemma is a straightforward exercise with the characteristic function of $W^T\tau$ and is postponed to Appendix A.

We can now describe how our least common denominator can be spotted in Fourier space. From Lemma 3.2 along with the negation of (20), we obtain $m > 0$ and a set $S_{W_Y}(m) \subseteq \mathbb{R}^{k+2}$ with Gaussian volume bounded below by $(Rt)^2 \exp(c_1 m - c_2 k)$. Now, for reasons that we will not explain here (since it is just a consequence of the Fourier transform), the first k-coordinates of the space, correspond to the k "soft" constraints while the final two coordinates correspond to the two "hard" constraints.

With this in mind, the idea is to find an element $\psi \in S_{W_Y}(m)$ for which $\|\psi_{[k]}\|_2 = O(\sqrt{k})$, and one of ψ_{k+1}, ψ_{k+2} is $O(1)$ and "non-trivial". It will turn out that one of ψ_{k+1}, ψ_{k+2} is a good candidate for our desired least common denominator. The condition on the $\psi_{[k]}$ should be thought of as just getting these coordinates "out of the way".

To find this desired $\psi \in S_{W_Y}(m)$, for $r, s > 0$, we define the *cylinder*

$$\Gamma_{r,s} := \left\{\theta \in \mathbb{R}^{k+2} : \ \left\|\theta_{[k]}\right\|_2 \leqslant r, |\theta_{k+1}| \leqslant s \text{ and } |\theta_{k+2}| \leqslant s\right\}. \tag{21}$$

We now restate our condition on ψ in terms of $\Gamma_{r,s}$: we want to show that there exists an $x \in S_{W_Y}(m)$ for which

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S_{W_Y}(m) \neq \emptyset, \tag{22}$$

where s is chosen depending on the non-triviality condition we need. We shall then ultimately see that if $y \in (\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x)$, where $x \in S_{W_Y}(m)$, then $(x - y)$ is a good candidate

for ψ (see Claims 5.4-5.6). In what remains in this section, we warm up by making a first easy reduction on the structure of $S_{W_Y}(m)$ under the assumption that (22) fails.

3.2. **A first reduction: controlling the density on fibers.** For our first reduction, we first record the following easy fact.

Fact 3.3. *For $s > 0$, let $S \subseteq \mathbb{R}^2$ be such that $\gamma_2(S) \geqslant 8s^2$, then there exists $x, y \in S$ so that $s < \|x - y\|_\infty \leqslant 16$.*

Proof. We prove the contrapositive and assume there is no pair $x, y \in S$ with $s < \|x-y\|_\infty \leqslant 16$. We cover $\mathbb{R}^2 = \bigcup_{p \in 16\cdot\mathbb{Z}^2} Q_p$ where $Q_p := p + [-8, 8]^2$. Thus $\gamma_2(S) \leqslant \sum_{p\in 16\cdot\mathbb{Z}^2} \gamma_2(S \cap Q_p)$. Since there is no $x, y \in S$ so that $s < \|x-y\|_\infty \leqslant 16$, then for each Q_p there is $x = x(p) \in Q_p$ so that

$$\gamma_2(S \cap Q_p) \leqslant \gamma_2(S \cap Q_p \cap (x(p) + [-s, s]^2)) \leqslant \gamma_2(x(p) + [-s, s]^2).$$

Letting $g \sim \mathcal{N}(0, (2\pi)^{-1})$, we have

$$\gamma_2(x + [-s, s]^2) \leqslant \mathbb{P}(x_1 - s \leqslant g \leqslant x_1 + s)\mathbb{P}(x_2 - s \leqslant g \leqslant x_2 + s) \leqslant 4s^2 \exp(-\pi\|p\|_2^2/16),$$

where we have used that $(x_i - s)^2 \geqslant p_i^2/8$, which holds since we may assume that $s \leqslant 1$ (else the statement holds trivially). Now we may bound

$$\gamma_2(S) \leqslant \sum_{p\in 16\cdot\mathbb{Z}^2} \gamma_2(S \cap Q_p) \leqslant 4s^2 \sum_{p\in 16\cdot\mathbb{Z}^2} \exp(-\pi\|p\|_2^2/16) < 8s^2,$$

which completes the proof. □

Now for $S \subseteq \mathbb{R}^{k+2}$, and $\theta_{[k]} \in \mathbb{R}^k$, we define the "vertical fiber"

$$S(\theta_{[k]}) := \left\{(\theta_{k+1}, \theta_{k+2}) \in \mathbb{R}^2 : (\theta_{[k]}, \theta_{k+1}, \theta_{k+2}) \in S\right\}. \tag{23}$$

The following lemma tells us that if we are unable to find a point in our desired intersection $(\Gamma_{r,16} \setminus \Gamma_{r,s} + x) \cap S$, for all $x \in S$, we can obtain good control on the measure of the vertical fibers of S.

Lemma 3.4. *For $k \in \mathbb{N}$, $r > 0$ and $s > 0$, let $S \subset \mathbb{R}^{k+2}$ be such that for all $x \in S$ we have*

$$(\Gamma_{r,16} \setminus \Gamma_{r,s} + x) \cap S = \emptyset.$$

Then

$$\max_{\theta_{[k]}\in\mathbb{R}^k} \gamma_2(S(\theta_{[k]})) \leqslant 8s^2.$$

Proof. We prove the contrapositive; let $\psi_{[k]}$ be such that $\gamma_2\left(S(\psi_{[k]})\right) > 8s^2$. This implies (Fact 3.3) that there exists $(\theta_{k+1}, \theta_{k+2}), (\theta'_{k+1}, \theta'_{k+2}) \in S(\psi_{[k]})$ with

$$s \leqslant \max\{|\theta_{k+1} - \theta'_{k+1}|, |\theta_{k+2} - \theta'_{k+2}|\} \leqslant 16.$$

Unpacking what this means in the full space $\mathbb{R}^{k+2}$: we have $\theta, \theta' \in S$ so that $\theta_{[k]}, \theta'_{[k]} = \psi_{[k]}$, and $s \leqslant \max\{|\theta_{k+1} - \theta'_{k+1}|, |\theta_{k+2} - \theta'_{k+2}|\} \leqslant 16$. Thus

$$\theta \in (\theta' + \Gamma_{r,16} \setminus \Gamma_{r,s}),$$

as desired. □

In the next section we go on to obtain a more complicated reduction of this form, that will ultimately be key in proving Lemma 3.1.

4. Inverse Littlewood Offord II: A geometric inequality

We now turn to make a more intricate and subtle reduction from that seen in Section 3.2, that will be key in finding our least common denominator. The lemma we prove here is purely geometric, but one should always think of it as being applied to an appropriate level set $S = S_{W_Y}(m)$, as seen in Lemma 3.2.

Given a set $S \subset \mathbb{R}^{k+2}$ and $y \in \mathbb{R}^{k+2}$, define the "translated horizontal fiber",

$$F_y(S; a, b) := \{\theta_{[k]} = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k : (\theta_1, \dots, \theta_k, a, b) \in S - y\}.$$

Our main goal of this section tells us that under the assumption

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S = \emptyset,$$

for all $x \in S$, the total measure of S can be controlled by the measure of the k-dimensional fibers $F_y(S; a, b)$. We state it in the contrapositive form to make the application (in Section 5) a little easier to spot.

Lemma 4.1. *For $k \in \mathbb{N}$ and $s > 0$, let $S \subset \mathbb{R}^{k+2}$ be a measurable set which satisfies*

$$8s^2 e^{-k/8} + 64s^2 \max_{a,b,y} \left(\gamma_k(F_y(S; a, b) - F_y(S; a, b))\right)^{1/4} < \gamma_{k+2}(S). \tag{24}$$

Then there is an $x \in S$ so that[8]

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S \neq \emptyset. \tag{25}$$

To prove this lemma, we will need a few facts about Gaussian space, which we collect in Sections 4.1 and 4.2, before moving on to prove Lemma 4.1 in Section 4.3.

4.1. A few facts about Gaussian space.

Recall that for $\ell \in \mathbb{N}$, γ_ℓ is the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1} I_\ell)$.

Lemma 4.2. *Let $k \geqslant 0$, $r > 0$ and $S \subset \mathbb{R}^{k+2}$ be measurable. Then there exists $x \in S$, and $h \in \Gamma_{r,8}$ so that*

$$\gamma_{k+2}(S \cap B) \leqslant 8\gamma_{k+2}((S - x) \cap \Gamma_{2r,16} + h),$$

where $B := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leqslant r\}$.

Proof. Consider translates $\Gamma_{r,8} + y$ where $y_{k+1}, y_{k+2} \in 16\mathbb{Z}^2$ to write

$$\gamma_{k+2}(S \cap B) = \sum_{y \in \{0\}^k \times 16\mathbb{Z}^2} \gamma_{k+2}(S \cap (\Gamma_{r,8} + y)). \tag{26}$$

We express $\gamma_{k+2}(S \cap (\Gamma_{r,8} + y))$ as

$$\int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\theta \in S \cap (\Gamma_{r,8} + y)\big] e^{-\pi\|\theta\|_2^2/2}\, d\theta = \int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\phi \in (S - y) \cap \Gamma_{r,8}\big] e^{-\pi\|\phi + y\|_2^2/2}\, d\phi. \tag{27}$$

[8]Note, in particular, that Lemma 4.1 says that *if* (24) is satisfied then we must have $s < 16$.

Rewriting the exponent in the integrand at (27)

$$-\|\phi+y\|_2^2 = -\|\phi\|_2^2 - 2\phi_{k+1}y_{k+1} - 2\phi_{k+2}y_{k+2} - y_{k+1}^2 - y_{k+2}^2,$$

we use that $|\phi_{k+1}|, |\phi_{k+2}| \leqslant 8$ whenever $\mathbb{1}[\phi \in (S-y)\cap\Gamma_{r,8}] \neq 0$, to see

$$\gamma_{k+2}(S\cap(\Gamma_{8,r}+y)) \leqslant \exp\left(-\frac{\pi}{2}y_{k+1}^2 - \frac{\pi}{2}y_{k+2}^2 + 8\pi|y_{k+1}| + 8\pi|y_{k+2}|\right)\gamma_{k+2}((S-y)\cap\Gamma_{r,8})\,. \quad (28)$$

So, apply (28) to (26) to get

$$\begin{aligned}\gamma_{k+2}(S\cap B) &\leqslant \sum_{y\in\{0\}^k\times 16\mathbb{Z}^2} \gamma_{k+2}((S-y)\cap\Gamma_{r,8})e^{-\frac{\pi}{2}y_{k+1}^2-\frac{\pi}{2}y_{k+2}^2+8\pi|y_{k+1}|+8\pi|y_{k+2}|}\\ &\leqslant \max_y \gamma_{k+2}((S-y)\cap\Gamma_{r,8}) \sum_{y_{k+1},y_{k+2}\in 16\mathbb{Z}} e^{-\frac{\pi}{2}y_{k+1}^2-\frac{\pi}{2}y_{k+2}^2+8\pi|y_{k+1}|+8\pi|y_{k+2}|}\\ &\leqslant 16\max_y \gamma_{k+2}((S-y)\cap\Gamma_{r,8})\,.\end{aligned}$$

Let y be a vector at which the above maximum is attained. Now observe that if $S\cap(\Gamma_{r,8}+y) = \emptyset$ then $(S-y)\cap\Gamma_{r,8} = \emptyset$ and thus $\gamma_{k+2}(S\cap B) = 0$; so there is nothing to prove. Thus we may assume $S\cap(\Gamma_{r,8}+y) \neq \emptyset$ and let $x \in S\cap(\Gamma_{8,r}+y)$. Define $h := x-y \in \Gamma_{r,8}$ and notice that

$$(S-y)\cap\Gamma_{r,8} - h = (S-y-h)\cap(\Gamma_{r,8}-h) \subseteq (S-x)\cap\Gamma_{2r,16},$$

where the inclusion holds since $h \in \Gamma_{r,8}$. Therefore $(S-y)\cap\Gamma_{r,8} \subseteq (S-x)\cap\Gamma_{2r,16}+h$, allowing us to conclude that

$$\gamma_{k+2}(S\cap B) \leqslant 16\gamma_{k+2}((S-y)\cap\Gamma_{r,8}) \leqslant 16\gamma_{k+2}((S-x)\cap\Gamma_{2r,16}+h),$$

as desired. □

We also need the following standard tail estimate on a k-dimensional Gaussian.

Fact 4.3. $\gamma_k\big(\{x\in\mathbb{R}^k : \|x\|_2^2 \geqslant k\}\big) \leqslant \exp(-k/8)$.

Proof. For any $\varepsilon \in (0,1)$ the *standard* Gaussian measure of the set $\{x \in \mathbb{R}^k : \|x\|_2^2 \geqslant k/(1-\varepsilon)\}$ is at most $\exp(-\varepsilon^2 k/4)$. Recalling that γ_k has standard deviation $(2\pi)^{-1/2}$ and taking $\varepsilon = 1-(2\pi)^{-1}$, gives the desired bound. □

4.2. **A Gaussian Brunn-Minkowski type theorem.** We now lay out a useful tool which gives us some control of the Gaussian measure of the sum set $A+B$, relative to the Gaussian measures of A and B. Indeed, the following theorem due to Borell [4], can be viewed as a Brunn-Minkowski-type theorem for Gaussian space.

For this, let $\Phi(x)$ be the cumulative probability function $\Phi(x) := \mathbb{P}(Z \leqslant x)$, for the *standard* one dimensional Gaussian $Z \sim \mathcal{N}(0,1)$, while γ_k is (still) the k-dimensional Gaussian with covariance matrix $(2\pi)^{-1}I_k$.

Theorem 4.4 (Borell)**.** *Let $A, B \subseteq \mathbb{R}^k$ be Borel. Then*

$$\gamma_k(A+B) \geqslant \Phi\Big(\Phi^{-1}(\gamma_k(A)) + \Phi^{-1}(\gamma_k(B))\Big).$$

Proof. In [4] Theorem 4.4 is proved for the standard Gaussian measure rather than γ_k. However we can change the standard deviation of the measure by taking dilates of the sets A and B. $\square$

We will use the following simple consequence of Theorem 4.4.

Lemma 4.5. *Let $A \subseteq \mathbb{R}^k$ be Borel. Then*

$$\gamma_k(A - A) \geqslant \gamma_k(A)^4 .$$

Proof. By Theorem 4.4, we have

$$\gamma_k(A - A) \geqslant \Phi(2\Phi^{-1}(\gamma_k(A))) = \Phi(2x), \tag{29}$$

where we have set $x = \Phi^{-1}(\gamma_k(A))$. Note that

$$\Phi(2x) = \mathbb{P}(Z \leqslant 2x) = \mathbb{P}\left(Z_1 + Z_2 + Z_3 + Z_4 \leqslant 4x\right) \geqslant \mathbb{P}(Z \leqslant x)^4 = \Phi(x)^4 \tag{30}$$

where Z_j are i.i.d. copies of $Z \sim \mathcal{N}(0,1)$. Combining (29) and (30) completes the proof. $\square$

4.3. **Proof of Lemma 4.1.** With these pieces now in place, we can move on to prove Lemma 4.1, our key geometric lemma on the Fourier side.

Proof of Lemma 4.1. Write $r = \sqrt{k}$ for simplicity. We prove the contrapositive and assume for every $x \in S$ we have

$$(\Gamma_{2r,16} \setminus \Gamma_{2r,s} + x) \cap S = \emptyset. \tag{31}$$

We define

$$B := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leqslant r\}.$$

and proceed to bound $\gamma_{k+2}(S)$ from above by first bounding $\gamma_{k+2}(S \setminus B)$ and then bounding $\gamma_{k+2}(S \cap B)$.

Step 1: Upper bound for $\gamma_{k+2}(S \setminus B)$. For $\theta_{[k]} \in \mathbb{R}^k$, let $S(\theta_{[k]})$ be as defined at (23)

$$S(\theta_{[k]}) = \left\{(\theta_{k+1}, \theta_{k+2}) \in \mathbb{R}^2 : (\theta_{[k]}, \theta_{k+1}, \theta_{k+2}) \in S\right\}.$$

We may write

$$\gamma_{k+2}(S \setminus B) = \int_{\|\theta_{[k]}\|_2 \geqslant r} \gamma_2\left(S(\theta_{[k]})\right) \, d\gamma_k \tag{32}$$

and thus

$$\gamma_{k+2}(S \setminus B) \leqslant \left(\max_{\theta_{[k]} \in \mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right)\right) \gamma_k\big(\{\|\theta_{[k]}\|_2 \geqslant r\}\big). \tag{33}$$

Lemma 3.4 and (31) shows

$$\max_{\theta_{[k]} \in \mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right) \leqslant 8s^2. \tag{34}$$

Fact 4.3 bounds

$$\gamma_k\big(\{\|\theta_{[k]}\|_2 \geqslant r\}\big) \leqslant \exp(-k/8) \tag{35}$$

and so from (33), (34) and (35) we learn

$$\gamma_{k+2}(S \setminus B) \leqslant 8s^2 e^{-k/8}. \tag{36}$$

Step 2: Upper bound for $\gamma_{k+2}(S \cap B)$. By Lemma 4.2, there exists $x \in S$ and $h \in \Gamma_{r,8}$ such that

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-x) \cap \Gamma_{2r,16} + h). \tag{37}$$

Now since we are assuming the claim is false, and $x \in S$, we use (31) to deduce that

$$(S-x) \cap \Gamma_{2r,16} \subseteq (S-x) \cap \Gamma_{2r,s} \tag{38}$$

and so letting $y = x - h$, we see

$$(S-x) \cap \Gamma_{2r,s} + h = (S - x + h) \cap (\Gamma_{2r,s} + h) = (S-y) \cap (\Gamma_{2r,s} + h). \tag{39}$$

Thus by (37), (38) and (39), we have

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)). \tag{40}$$

Bound

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant \int_{|a-h_{k+1}|,|b-h_{k+2}| \leqslant s} \gamma_k\big(F_y(S;a,b)\big)\, d\gamma_2 \tag{41}$$

and apply Lemma 4.5 to obtain

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant 4s^2 \max_{a,b,y}(\gamma_k(F_y(S;a,b) - F_y(S;a,b)))^{1/4}. \tag{42}$$

Combining (40) and (42) gives

$$\gamma_{k+2}(S \cap B) \leqslant 64s^2 \max_{a,b,y} \big(\gamma_k\left(F_y(S;a,b) - F_y(S;a,b)\right)\big)^{1/4} \tag{43}$$

Putting Step 1 and Step 2 together : (43) together with (36) implies

$$\gamma_{k+2}(S) \leqslant 8s^2 e^{-k/8} + 64s^2 \max_{a,b,y}(\gamma_k(F_y(S;a,b) - F_y(S;a,b)))^{1/4},$$

completing the proof of the contrapositive. □

5. Inverse Littlewood-Offord III: Comparison to a lazier walk and Proof of Lemma 3.1

In Section 4 we proved our key geometric ingredient, Lemma 4.1, to deal with the geometry of our level set (as seen in Section 3.1). We now use this lemma to take the following big step towards Lemma 3.1.

Lemma 5.1. *For $d \in \mathbb{N}$ and $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $t \geqslant \exp(-2^{-10}\alpha d)$. For $0 < c_0 \leqslant 2^{-24}$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\| \geqslant 2^{-10}c_0/t$ and let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$. Also let $\tau \sim \mathcal{Q}(2d, 1/4)$ and $\tau' \sim \mathcal{Q}(2d, 2^{-9})$ and $\beta \in [c_0/2^{10}, \sqrt{c_0}]$, $\beta' \in (0, 1/2)$.*

If

$$\mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}) \geqslant (Rt)^2 \exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} \tag{44}$$

then $D_\alpha(Y) \leqslant 16$. Here we have set $R = 2^{31}/c_0^2$.

Of course, Lemma 5.1 looks quite a bit like Lemma 3.1 save for quantity

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k), \tag{45}$$

on the right-hand side of (44). One should view this quantity as an approximation of the contribution that the "soft" constraints make. Indeed, if one reads this lemma in the contrapositive, it says that we can successfully "decouple" the "soft" constraints from the "hard" constraints, provided Y is sufficiently "unstructured", meaning $D_\alpha(Y) > 16$. Of course, this story is not quite an honest one; we have to use the lazier vector τ', rather than τ, to get things to work out, and we also take a loss in the exponent of $1/4$. The key here is that we obtain the correct power of t in our bound, which is deeply important for our application. We also note that our use of "decoupling" should not be confused with the "decoupling" step in Costello, Tao and Vu [8], which is used to deal with very unstructured vectors.

We prove this lemma in Section 5.2 after laying out a few facts on level sets in Section 5.1. We will then conclude this section in Section 5.3 with a proof of Lemma 3.1, by combining Lemma 5.1 with one further ingredient to bound (45).

5.1. **Working with level sets.** To prepare for the proof of Lemma 5.1, we record two basic facts about level sets. First off, we note a sort of converse to the Esseen-type inequality that we saw in Section 3, Lemma 3.2. Again, we will postpone the straightforward proof of this lemma to Appendix A. Recall that we defined, for a $2d \times \ell$ matrix W, the *W-level set*, for $t \geqslant 0$, to be

$$S_W(t) := \left\{\theta \in \mathbb{R}^\ell : \|W\theta\|_{\mathbb{T}} \leqslant \sqrt{t}\right\}.$$

Lemma 5.2. *Let $\beta > 0, \mu \in (0, 1/4]$, let W be a $2d \times \ell$ matrix, and let $\tau \sim \mathcal{Q}(2d, \mu)$. Then for all $t \geqslant 0$, we have*

$$\gamma_\ell(S_W(t))e^{-32\mu t} \leqslant \mathbb{P}_\tau\big(\|W^T \cdot \tau\|_2 \leqslant \beta\sqrt{\ell}\big) + \exp\left(-\beta^2\ell\right).$$

We need also need the following basic fact about level sets. Recall that, for a set $S \subset \mathbb{R}^{k+2}$ and $y \in \mathbb{R}^{k+2}$, we defined the "translated horizontal fiber",

$$F_y(S; a, b) := \{\theta_{[k]} = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k : (\theta_1, \ldots, \theta_k, a, b) \in S - y\}.$$

Fact 5.3. *For any $2d \times (k+2)$ matrix W. If $m > 0$ we have*

$$S_W(m) - S_W(m) \subseteq S_W(4m).$$

Similarly, for any $y \in \mathbb{R}^{k+2}$ and $a, b \in \mathbb{R}$ we have

$$F_y(S_W(m); a, b) - F_y(S_W(m); a, b) \subseteq F_0(S_W(4m); 0, 0). \tag{46}$$

Proof. Notice that if $x, y \in S_W(m)$ then by definition $\|Wx\|_{\mathbb{T}}, \|Wy\|_{\mathbb{T}} \leqslant \sqrt{m}$. Thus, by the triangle inequality,

$$\|W(x-y)\|_{\mathbb{T}} \leqslant \|Wx\|_{\mathbb{T}} + \|Wy\|_{\mathbb{T}} \leqslant 2\sqrt{m}.$$

For (46), let $\theta_{[k]}, \theta'_{[k]} \in F_y(S; a, b)$. We have that

$$(\theta_1, \ldots, \theta_k, a, b), (\theta'_1, \ldots, \theta'_k, a, b) \in S_W(m) - y$$

and so $\theta'' := (\theta_1 - \theta'_1, \ldots, \theta_k - \theta'_k, 0, 0) \in S_W(4m)$. Thus $\theta_{[k]} - \theta'_{[k]} \in F_0(S_W(4m); 0, 0)$, implying (46). □

5.2. **Proof of 5.1.** We may now turn to prove Lemma 5.1, our big step towards Lemma 3.1.

Proof of Lemma 5.1. Apply Lemma 3.2 to find $m > 0$ such that the level set

$$S := S_{W_Y}(m) = \{\theta \in \mathbb{R}^{k+2} : \|W_Y \theta\|_{\mathbb{T}} \leqslant \sqrt{m}\},$$

satisfies

$$e^{-\frac{1}{8}m + 2\beta^2 k} \gamma_{k+2}(S) \geqslant \mathcal{L}(W_Y^T \tau, \beta\sqrt{k+1}). \tag{47}$$

Thus (47) together with our hypothesis (44) gives a lower bound

$$\gamma_{k+2}(S) \geqslant \frac{1}{4} e^{\frac{1}{8}m - 2\beta^2 k} (Rt)^2 T^{1/4}, \tag{48}$$

where we have set

$$T := \mathbb{P}(\|W^T \tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k).$$

We now make the following important designations,

$$r_0 := \sqrt{k} \qquad \text{and} \qquad s_0 := 2^{16} c_0^{-1} (\sqrt{m} + \sqrt{k}) t. \tag{49}$$

Recall from (21) that for $r, s > 0$ we defined the *cylinder*

$$\Gamma_{r,s} := \left\{\theta \in \mathbb{R}^{k+2} : \ \left\|\theta_{[k]}\right\|_2 \leqslant r \ \text{ and } |\theta_{k+1}| \leqslant s, \ |\theta_{k+2}| \leqslant s, \right\}.$$

Claim 5.4. *There exists $x \in S \subseteq \mathbb{R}^{k+2}$ so that*[9]

$$\left(\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x\right) \cap S \neq \emptyset. \tag{50}$$

Proof of Claim 5.4. We look to apply Lemma 4.1 with $s = s_0$. For this, we bound

$$M := \max_{a,b,y} \left\{ \gamma_k \Big(F_y(S; a, b) - F_y(S; a, b) \Big) \right\},$$

above by $e^{m/4} T$, thus giving a lower bound on $\gamma_{k+2}(S)$ and allowing us to apply Lemma 4.1. Use Fact 5.3 to see that for any y, a, b, we have

$$F_y(S; a, b) - F_y(S; a, b) \subseteq F_0(S_{W_Y}(4m); 0, 0). \tag{51}$$

Now carefully observe that

$$F_0(S_{W_Y}(4m); 0, 0) = \left\{\theta_{[k]} \in \mathbb{R}^k : \|W\theta_{[k]}\|_{\mathbb{T}} \leqslant \sqrt{4m}\right\} = S_W(4m),$$

which is a level-set corresponding to the ("decoupled") event $\mathbb{P}_{\tau'}(\|W^T \tau'\|_2 \leqslant \beta'\sqrt{k})$, where $\tau' \sim \mathcal{Q}(2d, 2^{-9})$ and $\beta' \in (0, 1/2)$ is as in the hypothesis. Thus we may apply Lemma 5.2 along with (51) to obtain

$$M \leqslant \gamma_k(F_0(S_{W_Y}(4m), 0, 0)) = \gamma_k(S_W(4m)) \leqslant e^{m/4} T. \tag{52}$$

[9]Note that this claim shows, in particular, that $s_0 < 16$.

So combining (52) with (48), gives

$$\gamma_{k+2}(S) \geqslant (1/4)e^{m/16+2\beta^2 k}(Rt)^2 M^{1/4} \geqslant 8s_0^2 e^{-k/8} + 64 s_0^2 M^{1/4}, \tag{53}$$

allowing us to apply Lemma 4.1 and complete the proof of the claim. The last inequality at (53) follows from a simple check: each term on the right-hand side of (53) is at most half of the left-hand side. First note that

$$s_0^2 = 2^{32} c_0^{-2}(\sqrt{m}+\sqrt{k})^2 t^2 < 2^{33}(k+m)(t/c_0)^2 \tag{54}$$

and so

$$8s_0^2 e^{-k/8} \leqslant \frac{1}{8} e^{m/8-2\beta^2 k}(Rt)^2 e^{-\beta'^2 k/4}$$

follows from $\beta' \leqslant 1/2$ and the definition of R. On the other hand, use (54) to bound

$$64 s_0^2 e^{m/16} \leqslant 2^{39} t^2 c_0^{-2}(2^{20} c_0^{-2}\beta^2 k + 8(m/8)) \leqslant \frac{1}{8}(Rt)^2 e^{m/8+\beta^2 k}$$

thus showing the second inequality at (53) and finishing the proof of the claim. □

We now observe the simple consequence of Claim 5.4.

Claim 5.5. *We have that* $S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0}) \neq \emptyset$.

Proof of Claim 5.5. By Claim 5.4, there exists $x, y \in S = S_{W_Y}(m)$ so that $y \in (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x) \cap S$. Set $\phi := y - x$ and observe that $\phi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$, by Fact 5.3. □

We now conclude the proof of Lemma 5.1 with the following claim.

Claim 5.6. *If* $\psi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$ *then there exists* $i \in \{k+1, k+2\}$ *so that*

$$\|\psi_i Y\|_{\mathbb{T}} < \min\{\psi_i \|Y\|_2/2, \sqrt{\alpha d}\}\,.$$

Proof of Claim 5.6. Note that since $\psi \in S_{W_Y}(4m)$ there is a $p \in \mathbb{Z}^{2d}$ so that $W_Y\psi \in B_{2d}(p, 2\sqrt{m})$. So if we express

$$W_Y\psi = W\psi_{[k]} + \psi_{k+1}\begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2}\begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix},$$

we have that

$$\psi_{k+1}\begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2}\begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \in B_{2d}(p, 2\sqrt{m}) - W\psi_{[k]} \subseteq B_{2d}(p, 2\sqrt{m} + 4\sqrt{k}), \tag{55}$$

where the last inclusion holds because $\psi \in \Gamma_{2r_0,16}$ and so $\|\psi_{[k]}\|_2 \leqslant 2r_0 \leqslant 2\sqrt{k}$ and $\|W\| \leqslant 2$.

Since $\psi \notin \Gamma_{2r_0,s_0}$ we have that at least one of $|\psi_{k+1}|, |\psi_{k+2}|$ are $> s_0$. So, assume without loss that $|\psi_{k+1}| > s_0$ and that $\psi_{k+1} > 0$ (otherwise replace ψ with $-\psi$). Now project (55) onto the first d coordinates, to obtain

$$\psi_{k+1} Y \in B_d(p_{[d]}, 2\sqrt{m} + 4\sqrt{k}). \tag{56}$$

We now observe that $\|\psi_{k+1}Y\|_{\mathbb{T}} < \frac{\psi_{k+1}\|Y\|_2}{2}$. Indeed,

$$\frac{\psi_{k+1}\|Y\|_2}{2} > \frac{s_0\|Y\|_2}{2} \geqslant \left(\frac{2^{15}(\sqrt{m}+\sqrt{k})t}{c_0}\right)\left(2^{-10}\frac{c_0}{t}\right) > (2\sqrt{m}+4\sqrt{k}), \tag{57}$$

where we have used the definition of s_0 and that $\|Y\|_2 > 2^{-10}c_0/t$.

Finally, we note that $m \leqslant 2^{-4}\alpha d$. To see this, we use (48), $\gamma_{k+2}(S) \leqslant 1$ and our lower bound $t \geqslant \exp(-2^{-9}\alpha d)$ to see

$$e^{-m/8} \geqslant \gamma_{k+2}(S)e^{-m/8} \geqslant (Rt)^2 e^{-2\beta'^2 k} \geqslant \exp(-2^{-7}\alpha d),$$

where we have used $k \leqslant 2^{-9}\alpha d$ and $\beta' < 1$ for the last inequality, thus $m \leqslant 2^{-4}\alpha d$. Therefore from (56) and (57) we have

$$\|\psi_{k+1}Y\|_{\mathbb{T}} \leqslant 2\sqrt{m}+4\sqrt{k} \leqslant \sqrt{\alpha d},$$

as desired. This completes the proof of the Claim 5.6. □

Let ψ and $i \in \{k+1, k+2\}$ be as guaranteed by Claim 5.6. Then $\psi_i \leqslant 16$, and

$$\|\psi_i Y\|_{\mathbb{T}} < \min\{\|\psi_i Y\|_2/2, \sqrt{\alpha d}\},$$

and so $D_\alpha(Y) \leqslant 16$ thus completing the proof of Lemma 5.1. □

5.3. **Proof of Lemma 3.1.** Before turning to prove Lemma 3.1, we require one further result which tells us that $\|W\sigma\|_2$ is anti-concentrated when σ is a random vector and W is a fixed matrix. While there are several interesting results of this type in the literature [12,15,36] (and we will encounter another in Subsection 7.2), we state here a variant of the Hanson-Wright inequality with an explicit constant. We derive this from Talagrand's classical inequality in Appendix D.

Lemma 5.7. *For $d \in \mathbb{N}$, $\nu \in (0,1)$, let $\delta \in (0, \sqrt{\nu}/16)$, let $\sigma \sim \mathcal{Q}(2d,\nu)$, and let W be a $2d \times k$ matrix satisfying $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Then*

$$\mathbb{P}(\|W^T\sigma\|_2 \leqslant \delta\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k) \tag{58}$$

We now turn to prove Lemma 3.1.

Proof of Lemma 3.1. Setting $\beta' := 4\sqrt{c_0}$, we look to apply Lemma 5.1. For this, note that the hypotheses in Lemma 3.1 imply the hypotheses in Lemma 5.1 with respect to c_0, d, α, k, Y, W and τ (and we have the extra condition on $\|W\|_{\mathrm{HS}}$). So if we additionally assume $D_\alpha(Y) > 16$, we may apply Lemma 5.1 (in the contrapositive) to obtain

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (2^{31}c_0^{-2}t/2)^2 e^{4\beta^2 k}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + e^{-\beta'^2 k}\right)^{1/4}. \tag{59}$$

To deal with the right-hand side, we apply Lemma 5.7 to take care of the quantity involving $\tau' \in \{-1,0,1\}^{2d}$, our $\nu = 2^{-9}$ lazy random vector. Note that $4\sqrt{c_0} \leqslant 2^{-10} \leqslant \sqrt{\nu}/16$, and

that our given W satisfies $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Thus we may apply Lemma 5.7, with $\delta = \beta'$ and $\sigma = \tau'$, to see

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k). \tag{60}$$

Plugging this into the right-hand side of (59) yields

$$\begin{aligned}\exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} &\leqslant 2\exp(4c_0 k - 2^{-21}k) + 2\exp(2c_0 k - 4c_0 k)\\ &\leqslant 4\exp(-c_0 k).\end{aligned}$$

Putting this together with (59), yields

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (Rt)^2\exp(-c_0 k),$$

as desired. □

6. Inverse Littlewood-Offord for conditioned random matrices

In this section we lift the main result of the previous sections (Lemma 3.1) to study the concentration of the vector H_1X, where H_1 is a random $(n-d)\times d$ matrix, conditioned on having k singular values which are much smaller than "typical" and X is a fixed vector for which $|X_i| \approx N$ for each i.

Here N should be thought of as $\approx 1/\varepsilon$, in the context of the proof (see Section 1.2) and H_1 comes from its appearance in our matrix M,

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}.$$

The main result of this section is the following theorem[10].

Theorem 6.1. *For $n \in \mathbb{N}$ and $0 < c_0 \leqslant 2^{-24}$, let $d \leqslant c_0^2 n$, and for $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $N \leqslant \exp(2^{-10}\alpha d)$. Let $X \in \mathbb{R}^d$ satisfy $\|X\|_2 \geqslant c_0 2^{-10} n^{1/2} N$, and let H be a random $(n-d)\times 2d$ matrix with i.i.d. $(1/4)$-lazy entries in $\{-1,0,1\}$.*

If $D_\alpha(r_n \cdot X) > 16$ then

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n\right) \leqslant e^{-c_0 nk/4}\left(\frac{R}{N}\right)^{2n-2d}, \tag{61}$$

where we have set $H_1 := H_{[n-d]\times[d]}$, $H_2 := H_{[n-d]\times[d+1,2d]}$, $r_n := \frac{c_0}{16\sqrt{n}}$ and $R := 2^{39}c_0^{-3}$.

To understand the numerology in Theorem 6.1, notice that if we only consider the "soft" constraints on the singular values (without the constraints imposed by X) we would expect something like

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n}\right) \approx c^{nk}, \tag{62}$$

[10]For convenience, we define $\sigma_j(H) = 0$ for $j > \mathrm{rk}(H)$.

for some absolute $c \in (0,1)$, which depends on the value of c_0. Here we are using, crucially, that H is a *rectangular* matrix with aspect ratio bounded away from 1. Indeed, if H were a square matrix then $\sigma_{\min}(H) \approx n^{-1/2}$, with high probability[11].

On the other hand, the inverse Littlewood-Offord theorem of Rudelson and Vershynin [33] (with a bit of extra work) tells us that if X is such that $|X_i| \approx N$ for all $i \in [d]$, and

$$\mathbb{P}(\|H_1 X\|_2, \|H_2 X\|_2 \leqslant n) \geqslant \left(\frac{R}{N}\right)^{2n-2d},$$

then $D_\alpha(n^{-1/2}X) = O(1)$. Thus Theorem 6.1 is telling us that we maintain an inverse Littlewood-Offord type theorem even in the presence of many additional constraints imposed by the condition on the least singular values.

6.1. **A tensorization step.** We need the following basic fact.

Fact 6.2. *If $r \geqslant t > 0$ and X is a random variable taking values in $\mathbb{R}^{k+2}$, then*

$$\mathcal{L}(X,t) \leqslant \mathcal{L}(X,r) \leqslant (1+2r/t)^{k+2}\mathcal{L}(X,t).$$

Proof. The lower bound is trivial. The upper bound follows from the fact that a ball of radius r in $\mathbb{R}^{k+2}$ can be covered by $(1+2r/t)^{k+2}$ balls of radius t. □

We now prove a "tensorization" lemma which shows that anti-concentration of a single row in a random matrix H (with iid rows) implies the anti-concentration of matrix products involving H.

Lemma 6.3. *For $d < n$ and $k \geqslant 0$, let W be a $2d \times (k+2)$ matrix and let H be a $(n-d) \times 2d$ random matrix with i.i.d. rows. Let $\tau \in \mathbb{R}^{2d}$ be a random vector with the same distribution as the rows of H. If $\beta \in (0, 1/8)$ then*

$$\mathbb{P}_H\big(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\big) \leqslant \Big(2^5 e^{2\beta^2 k}\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\Big)^{n-d}.$$

Proof. Apply Markov's inequality to see that

$$\mathbb{P}\big(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\big) \leqslant \exp\big(2\beta^2(k+1)(n-d)\big)\,\mathbb{E}_H e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2}. \tag{63}$$

Letting $\tau_1, \ldots, \tau_{n-d}$ denote the i.i.d. rows of H, we may rewrite

$$\mathbb{E}_H\, e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2} = \prod_{i=1}^{n-d} \mathbb{E}_{\tau_i}\, e^{-2\|W^T\tau_i\|^2/\beta^2} = \Big(\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2}\Big)^{n-d}. \tag{64}$$

Observe now that

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} = \int_0^\infty \mathbb{P}\left(e^{-2\|W^T\tau\|^2/\beta^2} > u\right) du = \int_0^\infty 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2/\beta \leqslant u\big)\, du.$$

[11] While we can refer the reader to [34,35] for more on the singular values of rectangular random matrices, we were not able to find any result such as (62) in the literature. However, it is not so hard to deduce (62) from the Hanson-Wright inequality [36] along with a "random rounding" step similar to that in Appendix E.

Splitting the integral on the right-hand side gives

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} = \int_0^{\sqrt{k+1}} 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2 \leqslant \beta u\big) + \int_{\sqrt{k+1}}^{\infty} 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2 \leqslant \beta u\big).$$

We then appeal to Fact 6.2 to write

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} \leqslant \mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\left(\int_0^{\sqrt{k+1}} 4ue^{-2u^2}\,du + \int_{\sqrt{k+1}}^{\infty}\left(1+\frac{2u}{\sqrt{k+1}}\right)^{k+2} 4ue^{-2u^2}\,du\right).$$

Here the first integral is $\leqslant 1$, while the second integral is $\leqslant 8$ (see Fact D.3 in Appendix D) and thus

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} \leqslant 9\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\,. \tag{65}$$

Combining lines (65) with (64) and (63) gives

$$\mathbb{P}_H(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}) \leqslant \left(9\exp(2\beta^2(k+1))\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\right)^{n-d},$$

and the result follows. □

6.2. **Approximating matrices W with nets.** Note that in Theorem 6.1, the least singular values of the matrix H could, a priori, correspond to any of a huge number of possible directions. To limit the number of directions we need to consider, we build nets for k-tuples of these directions. Luckily, the construction of these nets is rendered relatively simple (unlike the nets $\mathcal{N}_\varepsilon$) by appealing to a randomized-rounding technique pioneered in the context of random matrices by Livshyts [27] (also see Section 3 of [28]).

With this in mind, let $\mathcal{U}_{2d,k}$ be the set of all $2d \times k$ matrices with orthonormal columns. The following theorem provides a net for $\mathcal{U}_{2d,k}$, when viewed as a subset of $\mathbb{R}^{[2d]\times[k]}$. The proof is deferred to Appendix E.

Lemma 6.4. *For $k \leqslant d$ and $\delta \in (0, 1/2)$, there exists $\mathcal{N} = \mathcal{N}_{2d,k} \subset \mathbb{R}^{[2d]\times[k]}$ with $|\mathcal{N}| \leqslant (2^6/\delta)^{2dk}$ so that for any $U \in \mathcal{U}_{2d,k}$, any $r \in \mathbb{N}$ and $r \times 2d$ matrix A there exists $W \in \mathcal{N}$ so that*

(1) $\|A(W-U)\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|A\|_{\mathrm{HS}}$,
(2) $\|W-U\|_{\mathrm{HS}} \leqslant \delta\sqrt{k}$ and
(3) $\|W-U\| \leqslant 8\delta$.

Recall, for a $2d \times k$ matrix W and $Y \in \mathbb{R}^d$, we defined (at (18)) the augmented matrix

$$W_Y = \left[W, \begin{bmatrix}\mathbf{0}_d\\ Y\end{bmatrix}, \begin{bmatrix}Y\\ \mathbf{0}_d\end{bmatrix}\right].$$

6.3. **Proof of Theorem 6.1.** We recall a standard fact from linear algebra, reworded to suit our context.

Fact 6.5. *For $3d < n$, let H be a $(n-d)\times 2d$ matrix. If $\sigma_{2d-k+1}(H) \leqslant x$ then there exist k orthogonal unit vectors $w_1, \dots, w_k \in \mathbb{R}^{2d}$ so that $\|Hw_i\|_2 \leqslant x$. In particular, there exists $W \in \mathcal{U}_{2d,k}$ so that $\|HW\|_{\mathrm{HS}} \leqslant x\sqrt{k}$.*

We also note that if H is a $(n-d)\times 2d$ matrix with entries in $\{-1,0,1\}$ then we immediately have $\|H\|_{\mathrm{HS}} \leqslant \sqrt{2d(n-d)}$.

Proof of Theorem 6.1. Write $Y := \frac{c_0}{16\sqrt{n}} \cdot X$. We use Fact 6.5 to upper bound the left-hand-side of (61) as

$$\begin{aligned}\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1 X\|_2, \|H_2 X\|_2 \leqslant n) \\ \leqslant \mathbb{P}(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant 3c_0\sqrt{n(k+1)}/16).\end{aligned}$$

Set $\delta := c_0/16$, and let $\mathcal{W}$ be the δ-net for $\mathcal{U}_{2d,k}$, given by Lemma 6.4.

We fix a matrix H for a moment. If there exists a matrix $U \in \mathcal{U}_{2d,k}$ so that $\|HU_Y\|_{\mathrm{HS}} \leqslant 3c_0\sqrt{n(k+1)}/16$, apply Lemma 6.4 to find $W \in \mathcal{W}$ so that

$$\|HW_Y\|_{\mathrm{HS}} \leqslant \|H(W_Y - U_Y)\|_{\mathrm{HS}} + \|HU_Y\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|H\|_{\mathrm{HS}} + 3c_0\sqrt{n(k+1)}/16$$

which is at most $c_0\sqrt{n(k+1)}/4$, since $\|H\|_{\mathrm{HS}} \leqslant \sqrt{2nd}$. Thus

$$\mathbb{P}\left(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{16}\sqrt{n(k+1)}\right) \leqslant \mathbb{P}\left(\exists W \in \mathcal{W} : \|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right).$$

So by the union bound, we have

$$\mathbb{P}\left(\exists W \in \mathcal{W} : \ \|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}\right) \leqslant \sum_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}\right).$$

Now

$$|\mathcal{W}| \leqslant (2^6/\delta)^{2dk} \leqslant \exp(32dk\log c_0^{-1}) \leqslant \exp(c_0 k(n-d)/4),$$

where the last inequality holds since $d \leqslant c_0^2 n$, and so

$$\sum_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right) \leqslant e^{c_0 k(n-d)/4} \max_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right). \tag{66}$$

Let $W \in \mathcal{W}$ be such that the maximum in (66) is attained, apply Lemma 6.3 with $\beta := \sqrt{c_0}/2$ to obtain

$$\mathbb{P}(\|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}) \leqslant \left(2^5 e^{c_0 k/2}\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big)\right)^{n-d}. \tag{67}$$

We now look to apply Lemma 3.1. We define $t := 16/(c_0 N) \geqslant \exp(-2^{-9}\alpha d)$ and $R_0 := 2^{-7}c_0 R = 2^{-7}c_0(2^{39}c_0^{-3}) = 2^{32}c_0^{-2}$ so that we have

$$\|Y\|_2 = c_0\|X\|_2/(16n^{1/2}) \geqslant 2^{-14}c_0^2 N = 2^{-10}c_0/t.$$

By the construction of $\mathcal{W}$ in Lemma 6.4 we have $\|W\| \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$. We also have $k \leqslant 2^{-10}\alpha d$ and $D_\alpha(\frac{c_0}{16\sqrt{n}}X) = D(Y) > 16$, therefore we may apply Lemma 3.1 to see that

$$\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big) \leqslant (R_0 t)^2\exp(-c_0 k) \leqslant \left(\frac{R}{8N}\right)^2\exp(-c_0 k).$$

Substituting this bound in (67) we get

$$\max_{W\in\mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leqslant (c_0/4)\sqrt{n(k+1)}) \leqslant \left(\frac{R}{N}\right)^{2n-2d} \exp(-c_0k(n-d)/2)$$

and finally combining it with the previous bounds gives

$$\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0\sqrt{n}/16 \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n) \leqslant \left(\frac{R}{N}\right)^{2n-2d} \exp(-c_0k(n-d)/4).$$

This completes the proof of Theorem 6.1. □

7. Nets for structured vectors: Size of the Net

In this section we take a important step towards Theorem 1.1 by bounding the size of our net

$$\mathcal{N}_\varepsilon := \left\{v \in \Lambda_\varepsilon : (L\varepsilon)^n \leqslant \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8L\varepsilon)^n\right\},$$

where we recall that

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]).$$

In particular, our main goal of this section will be to prove the following theorem on the size of $\mathcal{N}_\varepsilon$.

Theorem 7.1. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-24}$, let $n \geqslant L^{64/c_0^2}$, let $d \in [c_0^2n/4, c_0^2n]$ and let $\varepsilon > 0$ be such that $\log \varepsilon^{-1} \leqslant nL^{-32/c_0^2}$. Then*

$$|\mathcal{N}_\varepsilon| \leqslant \left(\frac{C}{c_0^6L^2\varepsilon}\right)^n,$$

where $C > 0$ is an absolute constant.

As the geometry of the set Λ_ε is a bit complicated, we follow an idea of Tikhomirov [50], by working with the intersection of $\mathcal{N}_\varepsilon$ with a selection of "boxes" which cover (an appropriately re-scaled) Λ_ε.

Definition 7.2. *Define a (N,κ,d)-*box *to be a set of the form $\mathcal{B} = B_1 \times \ldots \times B_n \subset \mathbb{Z}^n$ where $|B_i| \geqslant N$ for all $i \geqslant 1$; $B_i = [-\kappa N, -N] \cup [N, \kappa N]$, for $i \in [d]$; and $|\mathcal{B}| \leqslant (\kappa N)^n$.*

The advantage of working with these boxes is that they lend themselves naturally to a probabilistic interpretation, which we now adopt. We ask "what is the probability that

$$\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n,$$

where X is chosen uniformly at random from $\mathcal{B}$?". This interpretation was used to ingenious effect in the work of Tikhomirov, who called this the "inversion of randomness". While we do take this vantage point, our path forward is considerably different from that of Tikhomirov.

We now state our key "box" version of Theorem 7.1, in this probabilistic framework. Indeed, almost all of the work in proving Theorem 7.1 goes into proving the following variant for boxes.

Lemma 7.3. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-24}$, let $n > L^{64/c_0^2}$ and let $\frac{1}{4}c_0^2 n \leqslant d \leqslant c_0^2 n$. For $N \geqslant 2$, satisfying $\log N \leqslant c_0 L^{-8n/d} d$, and $\kappa \geqslant 2$, let $\mathcal{B}$ be a (N, κ, d)-box and let X be chosen uniformly at random from $\mathcal{B}$. Then*

$$\mathbb{P}_X \left(\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n \right) \leqslant \left(\frac{R}{L}\right)^{2n},$$

where $R := Cc_0^{-3}$ and $C > 0$ is an absolute constant.

7.1. **Counting with the least common denominator.** In this subsection, we prove the following simple lemma, which says that the probability of choosing $X \in \mathcal{B}$ with "large" least common denominator is super-exponentially small. This will ultimately allow us to apply Theorem 6.1, which requires an upper-bound on the $D_\alpha(X)$ for application.

We point out that in Lemma 7.4, we rescale by a factor of $r_n = c_0 2^{-4} n^{-1/2}$, despite the fact we are working in $d < n$ dimensions. This is just a trace of the fact that $\mathbb{R}^n$ is our true point of reference. Additionally we will only need Lemma 7.4 when $K = 16$.

Lemma 7.4. *For $\alpha \in (0,1), K \geqslant 1$ and $\kappa \geqslant 2$, let $n \geqslant d \geqslant K^2/\alpha$ and let $N \geqslant 2$ be so that $KN < 2^d$. Let $\mathcal{B} = ([-\kappa N, -N] \cup [N, \kappa N])^d$ and let X be chosen uniformly at random from $\mathcal{B}$. Then*

$$\mathbb{P}_X \left(D_\alpha\big(r_n \cdot X\big) \leqslant K \right) \leqslant (2^{20}\alpha)^{d/4}, \tag{68}$$

where we have set $r_n := c_0 2^{-4} n^{-1/2}$.

Proof. If $D_\alpha\big(r_n \cdot X\big) \leqslant K$ then let $\psi \in (0, K]$ be the minimum[12] in the definition of least common denominator. Set $\phi := r_n \psi$ and observe that ϕ satisfies

$$\|\phi X\|_{\mathbb{T}} \leqslant \sqrt{\alpha d} \quad \text{and} \quad \phi \in [(2\kappa N)^{-1}, r_n K]. \tag{69}$$

To see the bound $\phi \geqslant (2\kappa N)^{-1}$, note that if $\phi < (2\kappa N)^{-1}$ then each coordinate of $\phi \cdot X$ lies in $(-1/2, 1/2)$ which would imply $\|\phi X\|_{\mathbb{T}} = \|\phi X\|_2 = \phi \|X\|_2$. Using the non-triviality condition in the definition of least common denominator (19), this would imply

$$\phi \|X\|_2 = \|\phi \cdot X\|_{\mathbb{T}} = \|\psi(r_n \cdot X)\|_{\mathbb{T}} \leqslant \psi \|r_n \cdot X\|_2/2 = \phi \|X\|_2/2,$$

which is a contradiction. Thus the bounds in (69) hold.

Now to calculate the probability in (68), we discretize the range of possible ϕ. For each integer $i \in [1/\alpha, 2KN/\alpha] =: I$ we define $\phi_i := i\alpha/(2\kappa N)$ and note that if X, ϕ satisfy (69) then there exists ϕ_i for which

$$\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d} \quad \text{and } \phi_i \in [(2\kappa N)^{-1}, r_n K],$$

by simply choosing ϕ_i for which $|\phi_i - \phi| \leqslant \alpha/(\kappa N)$ and using triangle inequality

$$\|\phi_i X\|_{\mathbb{T}} \leqslant \|\phi X\|_{\mathbb{T}} + \|(\phi_i - \phi)X\|_2 \leqslant \sqrt{\alpha d} + |\phi_i - \phi| \cdot \sqrt{d}(\kappa N) \leqslant 2\sqrt{\alpha d}. \tag{70}$$

[12]Technically the least common denominator is defined in terms of an infimum, however the minimum is always attained for non-zero vectors.

Thus we have that

$$\mathbb{P}_X(D_\alpha(r_n \cdot X) \leqslant K) \leqslant \sum_{i \in I} \mathbb{P}_X \left(\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d} \right). \tag{71}$$

To bound the terms on the right-hand side, note that if $\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d}$ then

$$\frac{1}{d} \sum_{j=1}^{d} \|\phi_i X_j\|_{\mathbb{T}}^2 \leqslant 4\alpha\,.$$

By averaging, there is a set $S(X,i) \subset [d]$ with $|S(X,i)| \geqslant d/2$ for which $\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}$ for all $j \in S(X,i)$. Union bounding over all sets $S \subseteq [d]$ and using the independence of the coordinates X_j we have

$$\mathbb{P}_X(D_\alpha(r_n \cdot X) \leqslant K) \leqslant 2^d \sum_{i \in I} \prod_{j=1}^{d/2} \mathbb{P}_{X_j} \left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha} \right). \tag{72}$$

We now claim that

$$\mathbb{P}_{X_j} \left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha} \right) \leqslant 32\sqrt{\alpha}. \tag{73}$$

For this, note that if $\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}$, then $|\phi_i X_j - p| \leqslant 4\sqrt{\alpha}$, where $p \in \mathbb{Z}$ satisfies $|p| \leqslant |\phi_i X_j| + 1 \leqslant \phi_i \kappa N + 1 =: T_i$. And so

$$\mathbb{P}_{X_j}(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}) \leqslant \sum_{p=-T_i}^{T_i} \mathbb{P}_{X_j}(|X_j - p\phi_i^{-1}| \leqslant 4\sqrt{\alpha}\phi_i^{-1}) \leqslant \frac{(2T_i+1)(8\alpha^{1/2}\phi_i^{-1}+1)}{2(\kappa-1)N}.$$

where we have used that X_j is uniform on $[-\kappa N, -N] \cup [N, \kappa N]$ and the lower bound $\kappa N \phi_i \geqslant 1/2$ from (70) along with the assumption $\kappa \geqslant 2$. Also note that $8\alpha^{1/2}\phi_i^{-1} \geqslant 1$ since $\phi \leqslant r_n K \leqslant d^{-1/2} K$, allowing us to conclude (73).

Now, plugging (73) into (72) and bounding $|I| \leqslant (2KN/\alpha + 1) \leqslant 3^d$ completes the proof of Lemma 7.4. □

7.2. **Anti-concentration for linear projections of random vectors.** In this subsection we prove the following anti-concentration result for random variables HX, where H is a *fixed* matrix and X is a random vector with independent entries. One small remark regarding notation: H as stated in Lemma 7.5 will actually be H^T in Section 7.3.

Lemma 7.5. *Let $N \in \mathbb{N}$, $n, d, k \in \mathbb{N}$ be such that $n - d \geqslant 2d > 2k$, H be a $2d \times (n-d)$ matrix with $\sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16$ and $B_1, \ldots, B_{n-d} \subset \mathbb{Z}$ with $|B_i| \geqslant N$. If X is taken uniformly at random from $\mathcal{B} := B_1 \times \ldots \times B_{n-d}$, then*

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \left(\frac{Cn}{dc_0 N} \right)^{2d-k},$$

where $C > 0$ is an absolute constant.

We derive this from the following anti-concentration result of Rudelson and Vershynin. This is essentially Corollary 1.4 along with Remark 2.3 in their paper [37], but we have restated their result slightly to better suit our context.

Theorem 7.6. *Let $N \in \mathbb{N}$ and let $n, d, k \in \mathbb{N}$ be such that $n - d \geqslant 2d > k$. Let P be an orthogonal projection of $\mathbb{R}^{n-d}$ onto a $(2d-k)$-dimensional subspace and let $X = (X_1, \ldots, X_{n-d})$ be a random vector with independent entries for which*

$$\mathcal{L}\big(X_i, 1/2\big) \leqslant N^{-1},$$

for all $i \in [n-d]$. Then, for all $K \geqslant 1$,

$$\max_{y \in \mathbb{R}^{n-d}} \mathbb{P}_X\big(\|PX - y\|_2 \leqslant K\sqrt{2d-k}\big) \leqslant \left(\frac{CK}{N}\right)^{2d-k},$$

where $C > 0$ is a absolute constant.

We can now deduce Lemma 7.5.

Proof of Lemma 7.5. Since $H^T H$ is a symmetric $(n-d) \times (n-d)$ matrix with $\mathrm{rk}(H) \leqslant 2d$, by the spectral theorem we have $H^T H = \sum_{i=1}^{2d} \sigma_i(H)^2 v_i v_i^T$, where $v_1, \ldots, v_{2d} \in \mathbb{R}^{n-d}$ are orthonormal. Define the orthogonal projection $P := \sum_{i=1}^{2d-k} v_i v_i^T$. Then we have

$$\|HX\|_2^2 = \langle X, H^T H X\rangle = \sum_{j=1}^{2d} \sigma_j(H)^2 \langle X, v_j\rangle^2 \geqslant \sigma_{2d-k}(H)^2 \sum_{j=1}^{2d-k} \langle X, v_j\rangle^2 \geqslant 2^{-8} c_0^2 n \|PX\|_2^2.$$

Therefore

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \mathbb{P}_X(\|PX\|_2 \leqslant 16 c_0^{-1}\sqrt{n}). \tag{74}$$

We now apply Theorem 7.6 to the orthogonal projection P with $K = 16c_0^{-1}\sqrt{n/(2d-k)}$ to see

$$\mathbb{P}_X(\|PX\|_2 \leqslant K\sqrt{2d-k}) \leqslant \left(\frac{Cn}{dc_0 N}\right)^{2d-k}, \tag{75}$$

which together with (74) completes the proof of Lemma 7.5. □

7.3. **Proof of Theorem 7.3.** We take a moment to prepare the ground for the proof of Theorem 7.3. We express our random matrix M, as in the statement of Theorem 7.3, as

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[n-d]\times[n-d]}, \end{bmatrix}$$

Where H_1 is a $(n-d) \times d$ random matrix with iid 1/4-lazy entries in $\{-1, 0, 1\}$. We shall also let H_2 be an independent copy of H_1 and define H to be the $(n-d) \times 2d$ matrix

$$H := \begin{bmatrix} H_1 & H_2 \end{bmatrix}.$$

For a vector $X \in \mathbb{R}^n$, we define the event $\mathcal{A}_1 = \mathcal{A}_1(X)$ by

$$\mathcal{A}_1 := \big\{H : \|H_1 X_{[d]}\|_2 \leqslant n \text{ and } \|H_2 X_{[d]}\|_2 \leqslant n\big\}$$

and let $\mathcal{A}_2 = \mathcal{A}_2(X)$ be the event

$$\mathcal{A}_2 := \big\{H : \|H^T X_{[d+1,n]}\|_2 \leqslant 2n\big\}.$$

We now note a simple inequality linking H, $\mathcal{A}_1$ and $\mathcal{A}_2$ with the event $\{\|MX\|_2 \leqslant n\}$.

Fact 7.7. *For $X \in \mathbb{R}^n$, let $\mathcal{A}_1 = \mathcal{A}_1(X)$, $\mathcal{A}_2 = \mathcal{A}_2(X)$ be as above. We have*

$$(\mathbb{P}_M(\|MX\|_2 \leqslant n))^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2).$$

Proof. Let M' be an independent copy of M. Expand $\mathbb{1}(\|MX\|_2 \leqslant n)$ as a sum of indicators, apply $\mathbb{E}_M$ and square to see

$$(\mathbb{P}_M(\|MX\|_2 \leqslant n))^2 = \sum_{M,M'} \mathbb{P}(M')\mathbb{P}(M)\mathbb{1}(\|MX\|_2, \|M'X\|_2 \leqslant n),$$

which is at most

$$\sum_{H_1,H_2} \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{1}\big(\|H_1X_{[d]}\|_2 \leqslant n, \|H_2X_{[d]}\|_2 \leqslant n \text{ and } \|H^TX_{[d+1,n]}\|_2 \leqslant 2n\big),$$

which is exactly $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2)$. □

We shall also need a "robust" notion of the rank of the matrix H: Define $\mathcal{E}_k$ to be the event

$$\mathcal{E}_k := \big\{H : \sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16 \text{ and } \sigma_{2d-k+1}(H) < c_0\sqrt{n}/16\big\}$$

and note that always exactly one of the events $\mathcal{E}_0, \ldots, \mathcal{E}_{2d}$ holds.

We now set

$$\alpha := 2^{13}L^{-8n/d}, \tag{76}$$

and, given a box $\mathcal{B}$, we define the set of *typical* vectors $T(\mathcal{B}) \subseteq \mathcal{B}$ to be

$$T = T(\mathcal{B}) := \big\{X \in \mathcal{B} : D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) > 16\big\}. \tag{77}$$

Now set $K := 16$ and note that Lemma 7.4 implies that if X is chosen uniformly from $\mathcal{B}$ and $n \geqslant L^{64/c_0^2} \geqslant 2^8/\alpha$ we have

$$\mathbb{P}_X(X \notin T) = \mathbb{P}_X\big(D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) \leqslant 16\big) \leqslant \big(2^{33}L^{-8n/d}\big)^{d/4} \leqslant \left(\frac{2}{L}\right)^{2n}. \tag{78}$$

Proof of Lemma 7.3. Let M, H_1, H_2, H, $\mathcal{A}_1, \mathcal{A}_2$, $\mathcal{E}_k$, α and $T := T(\mathcal{B})$ be as above. We denote

$$\mathcal{E} := \Big\{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant (L/N)^n\Big\}$$

and write

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X(\mathcal{E} \cap \{X \in T\}) + \mathbb{P}_X(X \notin T).$$

Now define

$$f(X) := \mathbb{P}_M(\|MX\|_2 \leqslant n)\mathbb{1}(X \in T)$$

and apply (78), the bound on $\mathbb{P}_X(X \notin T)$, to obtain

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X\left(f(X) \geqslant (L/N)^n\right) + (2/L)^{2n} \leqslant (N/L)^{2n}\mathbb{E}_X\, f(X)^2 + (2/L)^{2n}, \tag{79}$$

where the last inequality follows from Markov's inequality. So to prove Lemma 7.3, it is enough to prove $\mathbb{E}_X\, f(X)^2 \leqslant 2(R/N)^{2n}$.

From Fact 7.7 we may write

$$\mathbb{P}_M(\|MX\|_2 \leqslant n)^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2) = \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2|\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \tag{80}$$

and so

$$f(X)^2 \leqslant \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2|\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{1}(X \in T). \tag{81}$$

We now look to apply Lemma 6.1 to obtain upper bounds for the quantities $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)$, when $X \in T$. For this, note that $d \leqslant c_0^2 n$, $N \leqslant \exp(L^{-8n/d}d) \leqslant \exp(2^{-10}\alpha n)$ and set $R_0 := 2^{39}c_0^{-3}$ (This is the "R" in Theorem 6.1). Also note that, by the definition of a (N, κ, d)-box and the fact that $d \geqslant \frac{1}{4}c_0^2 n$, we have that $\|X_{[d]}\|_2 \geqslant d^{1/2}N \geqslant c_0 2^{-10}\sqrt{n}N$. Now set $\alpha' := 2^{-10}\alpha$ to see that for $X \in T$ and $0 \leqslant k \leqslant \alpha' d$,

$$\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant \exp(-c_0 nk/4)\left(\frac{R_0}{N}\right)^{2n-2d}.$$

Moreover by Theorem 6.1,

$$\sum_{k \geqslant \alpha' d} \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant \mathbb{P}_H\big(\{\sigma_{2d-\alpha' d}(H) \leqslant c_0\sqrt{n}/16\} \cap \mathcal{A}_1\big) \leqslant \exp(-c_0\alpha' dn/4).$$

Thus, for all $X \in \mathcal{B}$, we have

$$f(X)^2 \leqslant \sum_{k=0}^{\alpha' d} \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k)\exp(-c_0 nk/4)\left(\frac{R_0}{N}\right)^{2n-2d} + \exp(-c_0\alpha' dn/4)\,. \tag{82}$$

We now consider the quantities $g_k(X) := \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k)$ appearing in (82). Indeed,

$$\mathbb{E}_X[g_k(X)] = \mathbb{E}_X\mathbb{E}_H\big[\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k\big] = \mathbb{E}_{X_{[d]}}\,\mathbb{E}_H\left[\mathbb{E}_{X_{[d+1,n]}}\mathbb{1}[\mathcal{A}_2]\,\big|\,\mathcal{A}_1 \cap \mathcal{E}_k\right].$$

We now consider a fixed $H \in \mathcal{A}_1 \cap \mathcal{E}_k$ for $k \leqslant \alpha' d$. Each such H has $\sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16$ and thus we may apply Lemma 7.5 to see that

$$\mathbb{E}_{X_{[d+1,n]}}\,\mathbb{1}[\mathcal{A}_2] = \mathbb{P}_{X_{[d+1,n]}}(\|H^T X_{[d+1,n]}\|_2 \leqslant n) \leqslant \left(\frac{C'n}{c_0 dN}\right)^{2d-k} \leqslant \left(\frac{4C'}{c_0^3 N}\right)^{2d-k},$$

for an absolute constant $C' > 0$, using that $d \geqslant \frac{1}{4}c_0^2 n$. And so for each $0 \leqslant k \leqslant \alpha' d$, taking $R := \max\{8C'c_0^{-3}, 2R_0\}$, we have

$$\mathbb{E}_X[g_k(X)] \leqslant \left(\frac{R}{2N}\right)^{2d-k}. \tag{83}$$

We apply $\mathbb{E}_X$ to (82) and then use (83) to obtain

$$\mathbb{E}_X f(X)^2 \leqslant \left(\frac{R}{2N}\right)^{2n}\sum_{k=0}^{\alpha' d}\left(\frac{2N}{R}\right)^{k}\exp(-c_0 nk/4) + \exp(-c_0\alpha' dn/4).$$

Using that $N \leqslant \exp(c_0 n/4)$ and $N \leqslant \exp(c_0 L^{-8n/d} d) = \exp(c_0 \alpha' d/8)$ gives

$$\mathbb{E}_X f(X)^2 \leqslant 2\left(\frac{R}{2N}\right)^{2n}. \tag{84}$$

Combining (84) with (79) completes the proof of Lemma 7.3. □

7.4. **Proof of Theorem 7.1.** The main work of this section is now complete with the proof of Lemma 7.3. We now just need to go from X in a "box" to X in a "sphere" Λ_ε. To accomplish this step, we simply cover the sphere with boxes. Recall that

$$\mathcal{I}'([d]) := \left\{v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in [d]\right\},$$

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

and that $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants defined in Section 2.

Lemma 7.8. *For all $\varepsilon \in [0,1]$, $\kappa \geqslant \max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$, there exists a family $\mathcal{F}$ of (N,κ,d)-boxes with $|\mathcal{F}| \leqslant \kappa^n$ so that*

$$\Lambda_\varepsilon \subseteq \bigcup_{\mathcal{B}\in\mathcal{F}} (4\varepsilon n^{-1/2}) \cdot \mathcal{B}, \tag{85}$$

where $N = \kappa_0/(4\varepsilon)$.

Proof. For $\ell \geqslant 1$ define the interval of integers $I_\ell := \left[-2^\ell N, 2^\ell N\right] \setminus \left[-2^{\ell-1}N, 2^{\ell-1}N\right]$ and $I_0 := [-N, N]$. Also take $J := [-\kappa N, \kappa N] \setminus [-N, N]$. For $(\ell_{d+1}, \ldots, \ell_n) \in \mathbb{Z}^n_{\geqslant 0}$ we define the box $B(\ell_{d+1}, \ldots, \ell_n) := J^d \times \prod_{j=d+1}^n I_{\ell_j}$ and the family of boxes

$$\mathcal{F} := \left\{B(\ell_{d+1}, \ldots, \ell_n) : \sum_{j:\ell_j>0} 2^{2\ell_j} \leqslant 8n/\kappa_0^2\right\}.$$

We claim that $\mathcal{F}$ is the desired family. For this, we first show the inclusion at (85). Let $v \in \Lambda_\varepsilon$. Since $v \in 4\varepsilon n^{-1/2}\mathbb{Z}^n$, $X := vn^{1/2}/(4\varepsilon) \in \mathbb{Z}^n$. For $i \in [d+1, n]$, define ℓ_i so that $X_i \in I(\ell_i)$. We claim $X \in B(\ell_{d+1}, \ldots, \ell_n)$. For this, observe that $X_i \in J$ for $i \in [d]$: since $v \in \mathcal{I}'([d])$, we have $\kappa_0 \leqslant |v_i| n^{1/2} \leqslant \kappa_1$, for $i \in [d]$. So $\kappa_0/(4\varepsilon) \leqslant |X_i| \leqslant \kappa_1/(4\varepsilon)$, for $i \in [d]$. Thus $X_i \in J$ since $N = \kappa_0/(4\varepsilon)$ and $\kappa \geqslant \kappa_1/\kappa_0$. Thus $v \in B(\ell_{d+1}, \ldots, \ell_n)$. We now observe that $B(\ell_{d+1}, \ldots, \ell_n) \in \mathcal{F}$, since

$$\sum_{j:\ell_j>0} 2^{2(\ell_j-1)} N^2 \leqslant \sum_{j=1}^n X_j^2 \leqslant n/(4\varepsilon)^2 \left(\sum_i v_i^2\right) \leqslant 4nN^2/\kappa_0^2.$$

Thus we have (85).

We now show $|\mathcal{F}| \leqslant \kappa^n$. For this we only need to count the number of sequences $(\ell_{d+1}, \ldots, \ell_n)$ of non-negative integers for which $\sum_{\ell_i>0} 4^{\ell_i} \leqslant 8n/\kappa_0^2$. For each $t \geqslant 0$ are at most $8n/(4^t\kappa_0^2)$ values of $i \in [d+1, n]$ for which $\ell_i = t$ and there are at most $\binom{n}{\leqslant 8n/(4^t\kappa_0^2)}$ choices for these values of i. Hence, there are at most

$$\prod_{t\geqslant 0} \binom{n}{\leqslant 8n/(\kappa_0^2 4^t)} \leqslant (\kappa_0/4)^{-4n} < \kappa^n$$

such tuples.

It only remains to show an upper bound on the size of $B(\ell_{d+1},\dots,\ell_n)\in\mathcal{F}$. We have

$$|B(\ell_{d+1},\dots,\ell_n)|\leqslant N^n\kappa^d 2^{n+\sum_j \ell_j}\leqslant \kappa^d(16/\kappa_0^2)^n N^n\leqslant(\kappa N)^n$$

where the second inequality holds due to the fact $\prod_j 2^{\ell_j}\leqslant\left(\frac{1}{n}\sum_j 2^{2\ell_j}\right)^n\leqslant(8/\kappa_0^2)^n$ and the last inequality holds due to the choice of κ. □

We may now use our covering Lemma 7.8 to apply Theorem 7.3 to deduce Theorem 7.1, the main result of this section.

Proof of Theorem 7.1. Apply Lemma 7.8 with $\kappa=\max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$ and use the fact that $\mathcal{N}_\varepsilon\subseteq\Lambda_\varepsilon$ to write

$$\mathcal{N}_\varepsilon\subseteq\bigcup_{\mathcal{B}\in\mathcal{F}}\left((4\varepsilon n^{-1/2})\cdot\mathcal{B}\right)\cap\mathcal{N}_\varepsilon$$

and so

$$|\mathcal{N}_\varepsilon|\leqslant\sum_{\mathcal{B}\in\mathcal{F}}|(4\varepsilon n^{-1/2}\cdot\mathcal{B})\cap\mathcal{N}_\varepsilon|\leqslant|\mathcal{F}|\cdot\max_{\mathcal{B}\in\mathcal{F}}|(4\varepsilon n^{-1/2}\cdot\mathcal{B})\cap\mathcal{N}_\varepsilon|.$$

By rescaling by $\sqrt{n}/(4\varepsilon)$ and applying Lemma 7.3, we have

$$|(4\varepsilon n^{-1/2}\cdot\mathcal{B})\cap\mathcal{N}_\varepsilon|\leqslant\left|\left\{X\in\mathcal{B}:\mathbb{P}_M(\|MX\|_2\leqslant n)\geqslant(L\varepsilon)^n\right\}\right|\leqslant\left(\frac{R}{L}\right)^{2n}|\mathcal{B}|.$$

Here the application of Lemma 7.3 is justified as $0<c_0\leqslant 2^{-24}$, $c_0^2n/2\leqslant d\leqslant c_0^2n$; $\kappa\geqslant 2$; we have $\log 1/\varepsilon\leqslant n/L^{32/c_0^2}$ and therefore

$$\log N=\log\kappa_0/(4\varepsilon)\leqslant n/L^{32/c_0^2}\leqslant c_0L^{-8n/d}d,$$

as specified in Lemma 7.3, since $\kappa_0<1$, $d\geqslant L^{-1/c_0^2}n$, $c_0\geqslant L^{-1/c_0^2}$ and $8n/d\leqslant 16/c_0^2$. So, using that $|\mathcal{F}|\leqslant\kappa^n$ and $|\mathcal{B}|\leqslant(\kappa N)^n$ for each $\mathcal{B}\in\mathcal{F}$, we have

$$|\mathcal{N}_\varepsilon|\leqslant\kappa^n\left(\frac{R}{L}\right)^{2n}|\mathcal{B}|\leqslant\kappa^n\left(\frac{R}{L}\right)^{2n}(\kappa N)^n\leqslant\left(\frac{C}{c_0^6L^2\varepsilon}\right)^n,$$

where $C=\kappa^2R^2c_0^6$, thus completing the proof of Theorem 7.1. □

8. Nets for structured vectors: approximating with the net

While we have spent considerable energy up to this point showing that $\mathcal{N}_\varepsilon$ is small, we have so far not shown that it is in fact a *net*. We now show just this, by showing that vectors in Σ_ε are approximated by elements of $\mathcal{N}_\varepsilon$. As we will see, this is considerably easier and is taken care of in Lemma 8.2, which, in a similar spirit to Lemma 6.4, is based on randomized rounding. For this, we recall that we defined

$$\Sigma_\varepsilon=\{v\in\mathcal{I}([d]):\mathcal{T}_L(v)\in[\varepsilon,2\varepsilon]\}\subset\mathbb{S}^{n-1}\,, \tag{86}$$

where $\mathcal{T}_L(v)=\sup\{t\in[0,1]:\mathbb{P}(\|Mv\|_2\leqslant t\sqrt{n})\geqslant(4Lt)^n\}$, and $d=c_0^2n<2^{-32}n$. Also recall the definition of our net

$$\mathcal{N}_\varepsilon=\left\{v\in\Lambda_\varepsilon:\mathbb{P}(\|Mv\|_2\leqslant 4\varepsilon\sqrt{n})\geqslant(L\varepsilon)^n\text{ and }\mathcal{L}_{A,op}(v,\varepsilon\sqrt{n})\leqslant(2^8L\varepsilon)^n\right\}.$$

We also make the basic observation that if $\mathcal{T}_L(v) = s$, then

$$(2sL)^n \leqslant \mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n}) \leqslant (8sL)^n .$$

Until now, we have almost entirely been working with the matrix M. The following lemma allows us to make a comparison between M and our central object of study: A, a uniform $n\times n$ symmetric matrix with entries in $\{-1, 1\}$. The proof of the lemma is based on a comparison of Fourier transforms and is deferred to Appendix B. This is similar to the replacement step in the work of Kahn Komlós and Szemerédi [19] and subsequent works [5,45]. However here, we only need to "break even", whereas they are looking for a substantial gain at this step.

Lemma 8.1. *For $v \in \mathbb{R}^n$ and $t \geqslant \mathcal{T}_L(v)$ we have*

$$\mathcal{L}(Av, t\sqrt{n}) \leqslant (50Lt)^n .$$

We now prove Lemma 8.2 which tells us that $\mathcal{N}_\varepsilon$ is a net for Σ_ε.

Lemma 8.2. *Let $\varepsilon \in (0, \kappa_0/8)$, $d \leqslant n/32$. If $v \in \Sigma_\varepsilon$ then there is $u \in \mathcal{N}_\varepsilon$ with $\|u - v\|_\infty \leqslant 4\varepsilon n^{-1/2}$.*

Proof. Given $v \in \Sigma_\varepsilon$, we define a random variable $r = (r_1, \dots, r_n)$ where the r_i are independent, $\mathbb{E}\, r_i = 0$, $|r_i| \leqslant 4\varepsilon n^{-1/2}$ and such that $v - r \in 4\varepsilon n^{-1/2}\mathbb{Z}^n$, for all r. We then define the random variable $u := v - r$. We will show that with positive probability there is a choice of $u \in \mathcal{N}_\varepsilon$.

Note that $\|r\|_\infty = \|u - v\|_\infty \leqslant 4\varepsilon n^{-1/2}$ for all u. Also, $u \in \mathcal{I}'([d])$ for all u, since $v \in \mathcal{I}([d])$ and $\|u - v\|_\infty \leqslant 4\varepsilon/\sqrt{n} \leqslant \kappa_0/(2\sqrt{n})$. So, from the definition of $\mathcal{N}_\varepsilon$, we need only show that there exists such a u satisfying

$$\mathbb{P}(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n. \tag{87}$$

We first show that *all* u satisfy the upper bound at (87). To see this, write $\mathcal{E} = \{\|A\| \leqslant 4\sqrt{n}\}$ and let $w(u) \in \mathbb{R}^n$, be such that

$$\begin{aligned}\mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) &= \mathbb{P}\left(\|Av - Ar - w(u)\| \leqslant \varepsilon\sqrt{n} \text{ and } \mathcal{E}\right)\\ &\leqslant \mathbb{P}\left(\|Av - w(u)\| \leqslant 5\varepsilon\sqrt{n} \text{ and } \mathcal{E}\right)\\ &\leqslant \mathcal{L}_{A,op}(v, 5\varepsilon\sqrt{n}) \leqslant \mathcal{L}(Av, 5\varepsilon\sqrt{n}).\end{aligned}$$

Since $v \in \Sigma_\varepsilon$, Lemma 8.1 bounds

$$\mathcal{L}(Av, 5\varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n . \tag{88}$$

We now show that

$$\mathbb{E}_u\, \mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (1/2)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \geqslant (1/2)(2\varepsilon L)^n , \tag{89}$$

where the last inequality holds by the fact $v \in \Sigma_\varepsilon$. From (89), it follows that there exists $u \in \Lambda_\varepsilon$ satisfying (87).

So to prove the first inequality in (87), we define the event $\mathcal{E} := \{M : \|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}\}$. For all u, we have

$$\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) = \mathbb{P}_M(\|Mv - Mr\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n} \text{ and } \mathcal{E});$$

Thus

$$\begin{aligned}\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) &\geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n} \,|\mathcal{E})\mathbb{P}(\mathcal{E}) \\ &\geqslant \left(1 - \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n} \,|\mathcal{E})\right) \mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \,.\end{aligned}$$

Taking expectations with respect to u gives,

$$\mathbb{E}_u\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \left(1 - \mathbb{E}_u\mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n} \,|\mathcal{E})\right)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \tag{90}$$

and exchanging the expectations reveals that it is enough to show

$$\mathbb{E}_M\left[\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \,\middle|\, \mathcal{E}\right] \leqslant 1/2.$$

We will show that $\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \leqslant 1/4$ for all $M \in \mathcal{E}$, by Markov's inequality. For this, fix a $n \times n$ matrix M with entries $|M_{i,j}| \leqslant 1$ and $M_{i,j} = 0$, if $(i,j) \in [d+1,n] \times [d+1,n]$, and note that

$$\mathbb{E}_r \|Mr\|_2^2 = \sum_{i,j} \mathbb{E}\,(M_{i,j} r_i)^2 = \sum_i \mathbb{E}\, r_i^2 \sum_j M_{i,j}^2 \leqslant 32\varepsilon^2 d \leqslant \varepsilon^2 n,$$

where, for the second equality, we have used that the r_i are mutually independent and $\mathbb{E}\, r_i = 0$, for the third inequality, we used $\|r\|_\infty \leqslant 4\varepsilon/\sqrt{n}$ and for the final inequality we used $d \leqslant n/32$. Thus by Markov, we have

$$\mathbb{P}_r(\|Mr\|_2 \geqslant 2\varepsilon\sqrt{n}) \leqslant (2\varepsilon\sqrt{n})^{-2}\mathbb{E}_r \|Mr\|_2^2 \leqslant 1/4. \tag{91}$$

Putting (91) together with (90) proves (89), completing the proof of (87). □

9. Proof of Theorem 1.1

In this section we put together our results to prove Theorem 1.1. But before we get to this, we note a few reductions afforded by previous work. Let us define

$$q_n(\gamma) := \max_{w \in \mathbb{R}^n} \mathbb{P}_A(\exists v \in \mathbb{R}^n \setminus \{0\} : Av = w,\ \rho(v) \geqslant \gamma), \tag{92}$$

where

$$\rho(v) = \max_{w \in \mathbb{R}} \mathbb{P}\left(\sum_{i=1}^n \varepsilon_i v_i = w\right)$$

and $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$ are i.i.d. and uniform. One slightly irritating aspect of the definition (92) is that the existential quantifies over *all non-zero* $v \in \mathbb{R}^n$, rather than all $v \in \mathbb{S}^{n-1}$, as we have been working with. So, as we will shortly see, we will need to approximate this extra dimension of freedom with a net.

These small issues aside, we will use the following inequality, which effectively allows us to remove very unstructured vectors from consideration.

Lemma 9.1. *Let A be a random $n \times n$ symmetric $\{-1,1\}$-matrix. For all $\gamma > 0$ we have*

$$\mathbb{P}(\det(A) = 0) \leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma} \right)$$

We record the details of this lemma in Appendix C, although an almost identical lemma can be found in [7], which collected elements from [8, 10, 31].

9.1. **Non-flat vectors.** Here we note a lemma due to Vershynin [51] which tells us that it is enough for us to consider vectors $v \in \mathcal{I}$. For this, we reiterate the important notion of *compressible vectors*, introduced by Rudelson and Vershynin [33]. Say a vector in $\mathbb{S}^{n-1}$ is (δ, ρ)-compressible if it has distance $\leqslant \rho$ from a vector with support $\leqslant \delta n$. Let $\mathrm{Comp}\,(\delta, \rho)$ denote the set of such compressible vectors. In [51, Proposition 4.2], Vershynin provides the following lemma which allows us to disregard all compressible vectors.

Lemma 9.2. *There exist $\delta, \rho, c \in (0,1)$ so that for all $n \in \mathbb{N}$,*

$$\max_{w \in \mathbb{R}^n} \mathbb{P}_A \left(\bigcup_{v \in \mathbb{S}^{n-1} \setminus \mathrm{Comp}\,(\delta,\rho)} \left\{ \|Av - w\|_2 \leqslant c\sqrt{n} \right\} \right) \leqslant 2e^{-cn},$$

where A is a random $n \times n$ symmetric $\{-1,1\}$-matrix.

The following lemma of Rudelson and Vershynin [33, Lemma 3.4] tells us that incompressible vectors are "flat" for a constant proportion of coordinates.

Lemma 9.3. *For $\delta, \rho \in (0,1)$, let $v \in \mathrm{Incomp}\,(\delta, \rho)$. Then*

$$(\rho/2)n^{-1/2} \leqslant |v_i| \leqslant \delta^{-1/2} n^{-1/2}$$

for at least $\rho^2 \delta n/2$ values of $i \in [n]$.

Now recall that we defined

$$\mathcal{I}(D) = \left\{ v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D \right\}$$

and $\mathcal{I} = \bigcup_{D \subseteq [n], |D| = d} \mathcal{I}(D)$. Here we fix $\kappa_0 = \rho/3$ and $\kappa_1 = \delta^{-1/2} + \rho/6$, where δ, ρ are as in Lemma 9.2. We also fix $c_0 = \min\{2^{-24}, \rho\delta^{1/2}/2\}$.

The following lemma is what we will apply in the proof of Theorem 1.1.

Lemma 9.4. *For $n \in \mathbb{N}$, let $d < c_0^2 n$. Then*

$$\max_{w \in \mathbb{R}^n} \mathbb{P}_A \left(\bigcup_{v \in \mathbb{S}^{n-1} \setminus \mathcal{I}} \left\{ Av \in \{t \cdot w\}_{t>0},\ \|A\| \leqslant 4\sqrt{n} \right\} \right) \leqslant 16c^{-1}e^{-cn}.$$

Proof. Apply Lemma 9.3 along with the definitions of κ_1, κ_2 and $\mathcal{I}$ to see $\mathbb{S}^{n-1} \setminus \mathcal{I} \subseteq \mathrm{Comp}\,(\delta, \rho)$. Now take a $c\sqrt{n}$-net $\mathcal{X}$ for $\{t \cdot w\}_{0 < t \leqslant 4\sqrt{n}}$ of size $8c^{-1}$. Then

$$\left\{ A : Av \in \{t \cdot w\}_{t>0},\ \|A\| \leqslant 4\sqrt{n} \right\} \subset \bigcup_{w' \in \mathcal{X}} \left\{ A : \|Av - w'\|_2 \leqslant c\sqrt{n} \right\}.$$

Union bounding over $\mathcal{X}$ and applying Lemma 9.2 completes the lemma. □

9.2. **Proof of Theorem 1.1.** As we noted in Section 2, matrices A with $\|A\| \geqslant 4\sqrt{n}$ will be a slight nuisance for us. The following concentration inequality for the operator norm of a random matrix will allow us to remove all such matrices A from consideration.

Lemma 9.5. *Let A be uniformly drawn from all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. Then for n sufficiently large,*

$$\mathbb{P}\left(\|A\| \geqslant 4\sqrt{n}\right) \leqslant 4e^{-n/32}.$$

This follows from a classical result of Bai and Yin [1] (see also [43, Theorem 2.3.23]) which implies that the median of $\|A\|$ is equal to $(2 + o(1))\sqrt{n}$, combined with a concentration inequality due to Meckes [29, Theorem 2]. A version of Lemma 9.5 without explicit constants, is well-known and straightforward, though we have included a version with explicit constants for concreteness.

We will also need the following, rather weak, relationship between the threshold $\mathcal{T}_L$, defined in terms of the matrix M, and $\rho(v)$, the "one-dimensional" concentration function of v. For this we define one more bit of (standard) notation

$$\rho_\varepsilon(v) := \max_{b \in \mathbb{R}^n} \mathbb{P}\left(\sum_i v_i \varepsilon_i \in (b - \varepsilon, b + \varepsilon)\right).$$

Lemma 9.6. *Let $v \in \mathbb{S}^{n-1}$ and $\varepsilon = \mathcal{T}_L(v)$. Then $\rho_\varepsilon(v)^4 \leqslant 2^{12} L\varepsilon$.*

We postpone the proof of this lemma to Appendix B and move on to the proof of Theorem 1.1.

Proof of Theorem 1.1. It is not hard to see that $\mathbb{P}(\det(A) = 0) < 1$ for all n, and therefore it is enough to prove Theorem 1.1 for all sufficiently large n.

Now, as in Section 2, we set $\gamma = e^{-cn}$, where we now define, $c := L^{-32/c_0^2}/8$, $L := \max\{2^{26}C_1, 16/\kappa_0\}$, where $C_1 = C/c_0^6$ is the constant appearing in Theorem 7.1. We also let $c_0 > 0$ be as defined above and $d := \lceil c_0^2 n/2 \rceil$.

From Lemma 9.1 we have

$$\mathbb{P}(\det(A) = 0) \leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma}\right)$$

and so it is enough to bound $q_n(\gamma)$ for all large n. Let $\Sigma = \{v \in \mathbb{S}^{n-1} : \rho(v) \geqslant \gamma\}$, as defined in Section 2, and note that

$$\{A : \exists v \in \mathbb{R}^n,\ Av = w,\ \rho(v) \geqslant \gamma\} \subset \{A : \exists v \in \Sigma,\ Av \in \{t \cdot w\}_{t>0}\}.$$

Since $d = \lceil c_0^2 n/2 \rceil$, by Lemma 9.4 and Lemma 9.5, we have

$$q_n(\gamma) \leqslant \max_{w \in \mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I} \cap \Sigma :\ Av \in \{t \cdot w\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) + 32c^{-1}e^{-cn} \tag{93}$$

and so it is enough to show the first term on the right-hand-side is $\leqslant 2^{-n}$. Using that $\mathcal{I} = \bigcup_D \mathcal{I}(D)$, we have the first term of (93) is

$$\leqslant 2^n \max_{D\in[n]^{(d)}} \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}(D)\cap\Sigma:\ Av \in \{t\cdot w\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) \tag{94}$$

$$= 2^n \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}([d])\cap\Sigma:\ Av \in \{t\cdot w\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right), \tag{95}$$

where the last line holds by symmetry of the coordinates. Thus it is enough to show that the maximum at (95) is at most 4^{-n}.

Now, for $v \in \Sigma$ we have $\rho(v) \geqslant \gamma$ and so, by Lemma 9.6, we have that

$$\gamma^4 \leqslant \rho(v)^4 \leqslant \rho_{\mathcal{T}_L(v)}(v)^4 \leqslant 2^{12} L\mathcal{T}_L(v).$$

Define $\eta := \gamma^4/(2^{12}L) \leqslant \mathcal{T}_L(v)$. Also note that by definition, $\mathcal{T}_L(v) \leqslant 1/L \leqslant \kappa_0/8$.

Now, recalling definition (86) of $\Sigma_\varepsilon = \Sigma_\varepsilon([d])$ from Section 2, we may write

$$\mathcal{I}([d])\cap\Sigma \subseteq \bigcup_{i=1}^{n} \left\{v \in \mathcal{I} : \mathcal{T}_L(v) \in [2^{j-1}\eta, 2^j\eta]\right\} = \bigcup_{j=0}^{\log_2(\kappa_0/16\eta)} \Sigma_{2^j\eta}$$

and so by the union bound, it is enough to show

$$\max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \Sigma_\varepsilon:\ Av \in \{t\cdot w\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) \leqslant 8^{-n},$$

for all $\varepsilon \in [\eta, \kappa_0/16]$. Fix an $\varepsilon\sqrt{n}$-net $\mathcal{X}$ for $\{t\cdot w\}_{0<t\leqslant 4\sqrt{n}}$ of size $8/\varepsilon \leqslant 2^n$ to get

$$\{A:\ Av \in \{t\cdot w\}_{t>0},\ \|A\| \leqslant 4\sqrt{n}\} \subset \bigcup_{w'\in\mathcal{X}} \{A:\ \|Av - w'\|_2 \leqslant \varepsilon\sqrt{n},\ \|A\| \leqslant 4\sqrt{n}\}.$$

So by taking the union bound over $\mathcal{X}$ it is enough to prove that

$$Q_\varepsilon := \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \Sigma_\varepsilon:\ \|Av - w\|_2 \leqslant \varepsilon\sqrt{n}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) \leqslant 2^{-4n}. \tag{96}$$

Let $w \in \mathbb{R}^n$ be such that the maximum at (96) is attained. Now, since $\varepsilon < \kappa_0/8$ for $v \in \Sigma_\varepsilon$, we apply Lemma 8.2, to find a $u \in \mathcal{N}_\varepsilon = \mathcal{N}_\varepsilon([d])$ so that $\|v - u\|_2 \leqslant 4\varepsilon$. So if $\|A\| \leqslant 4\sqrt{n}$ and $\|Av - w\| \leqslant \varepsilon\sqrt{n}$, we see that

$$\|Au - w\|_2 \leqslant \|Av - w\|_2 + \|A(v-u)\|_2 \leqslant \|Av - w\|_2 + \|A\|\|(v-u)\|_2 \leqslant 32\varepsilon\sqrt{n}$$

and thus

$$\{A : \exists v \in \Sigma_\varepsilon:\ \|Av - w\| \leqslant \varepsilon\sqrt{n}\} \cap \{\|A\| \leqslant 4\sqrt{n}\} \subseteq \{A : \exists u \in \mathcal{N}_\varepsilon : \|Au - w\| \leqslant 32\varepsilon\sqrt{n}, \|A\| \leqslant 4\sqrt{n}\}.$$

So, by union bounding over our net $\mathcal{N}_\varepsilon$, we see that

$$Q_\varepsilon \leqslant \mathbb{P}_A\left(\exists u \in \mathcal{N}_\varepsilon : \|Au - w\| \leqslant 32\varepsilon\sqrt{n} \text{ and } \|A\| \leqslant 4\sqrt{n}\right) \leqslant \sum_{u\in\mathcal{N}_\varepsilon} \mathcal{L}_{A,op}\left(u, 32\varepsilon\sqrt{n}\right).$$

Now note that if $u \in \mathcal{N}_\varepsilon$, then $\mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n$ and so by Fact 6.2 we have that $\mathcal{L}_{A,op}\left(u, 32\varepsilon\sqrt{n}\right) \leqslant (2^{16}L\varepsilon)^n$. As a result,

$$Q_\varepsilon \leqslant |\mathcal{N}_\varepsilon|(2^{16}L\varepsilon)^n \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n (2^{16}L\varepsilon)^n \leqslant 2^{-4n}.$$

where the second to last inequality follows from our Theorem 7.1 and the last inequality holds for our choice of $L = \max\{2^{26}C_1, 16/\kappa_0\}$. To see that the application of Theorem 7.1 is valid, note that

$$\log 1/\varepsilon \leqslant \log 1/\eta = \log 2^{12}L/\gamma^4 \leqslant nL^{-32/c_0^2}/2 + \log 2^{12}L \leqslant nL^{-32/c_0^2},$$

where the last inequality hold for all sufficiently large n. This completes the proof. □

Acknowledgments

We thank Rob Morris for many helpful comments on the presentation of this paper. We also thank Vishesh Jain, Natasha Morrison, Ashwin Sah, Mehtaab Sawhney and Van Vu for helpful remarks on the first preprint.

Appendix A. The Proofs of two Esseen-type lemmas

In this section we prove our two Esseen-type lemmas, Lemma 3.2 and Lemma 5.2, for random variables of the form $W^T\tau$, where τ is a μ-lazy random vector in $\{-1,0,1\}^{2d}$ and W is a (fixed) $2d \times \ell$ matrix for some $\ell \in \mathbb{N}$. Recall that for a vector $u \in \mathbb{R}^\ell$, we let $\|u\|_{\mathbb{T}}$ denote the Euclidean distance from u to the integer lattice $\mathbb{Z}^\ell$.

A.1. **Basics of Fourier representation.** As above, we let τ be a μ-lazy random vector in $\{-1,0,1\}^{2d}$ and let W be a $2d \times \ell$ matrix. Recall the characteristic function φ_X of a vector valued random variable X is defined as

$$\varphi_X(\theta) = \mathbb{E}\exp(2\pi i\langle X, \theta\rangle),$$

and so we may express characteristic function of $W^T\tau$ as

$$\varphi(\theta) = \mathbb{E}\exp(2\pi i\langle \tau, W\theta\rangle) = \prod_{j=1}^{2d}\left((1-\mu) + \mu\cos(2\pi(W\theta)_j)\right).$$

We note the elementary fact that for $\mu \in [0, 1/4]$ we have

$$\mu\|x\|_{\mathbb{T}}^2 \leqslant -\log\left(1-\mu+\mu\cos(2\pi x)\right) \leqslant 32\mu\|x\|_{\mathbb{T}}^2, \tag{97}$$

from which we deduce

$$\exp\left(-32\mu\,\|W\theta\|_{\mathbb{T}}^2\right) \leqslant \varphi(\theta) \leqslant \exp\left(-\mu\,\|W\theta\|_{\mathbb{T}}^2\right). \tag{98}$$

We now note a standard fact regarding Fourier inversion (see [48] p.290).

Fact A.1 (Fourier inversion)**.** *Let X be a random vector in $\mathbb{R}^\ell$, then for $w \in \mathbb{R}^\ell$ we have*

$$\mathbb{E}\exp\left(-\frac{\pi\|X-w\|_2^2}{2}\right) = \int_{\mathbb{R}^\ell} e^{-\pi\|\theta\|_2^2}\cdot e^{-2\pi i\langle w,\theta\rangle}\varphi_X(\theta)\,d\theta\,.$$

In particular, letting $g \sim \mathcal{N}(0,(2\pi)^{-1}I_\ell)$, we have

$$\mathbb{E}\exp\left(-\frac{\pi\|X-w\|_2^2}{2}\right) = \mathbb{E}_g(e^{-2\pi i\langle w,g\rangle}\varphi_X(g))\,.$$

A.2. **Proof of Lemma 3.2 and Lemma 5.2.** Recall that for $\ell \in \mathbb{N}$, γ_ℓ denotes the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1} I_\ell)$. We begin with the proof of Lemma 3.2.

Proof of Lemma 3.2. Let $w \in \mathbb{R}^\ell$. We apply Markov's inequality to obtain

$$\mathbb{P}_\tau\big(\|W^T\tau - w\|_2 \leqslant \beta\sqrt{\ell}\big) \leqslant \exp\left(\frac{\pi}{2}\beta^2\ell\right) \mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \cdot \tau - w\|_2^2}{2}\right).$$

As above, let φ be the characteristic function of $W^T\tau$. We apply Fact A.1 and (98) to obtain

$$\mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \cdot \tau - w\|_2^2}{2}\right) = \mathbb{E}_g[e^{-2\pi i\langle w,g\rangle}\varphi(g)] \leqslant \mathbb{E}_g[\exp(-\nu\|Wg\|_{\mathbb{T}}^2)].$$

The right-hand-side of the above may be rewritten as

$$\int_0^1 \mathbb{P}_g(\exp(-\nu\|Wg\|_{\mathbb{T}}^2) \geqslant t)\, dt = \nu \int_0^\infty \mathbb{P}_g(\|Wg\|_{\mathbb{T}}^2 \leqslant u) e^{-\nu u}\, du = \nu \int_0^\infty \gamma_\ell(S_W(u)) e^{-\nu u}\, du,$$

where for the first equality we made the change of variable $t = e^{-\nu u}$.

Choosing m to maximize $\gamma_\ell(S_W(u))e^{-\nu u/2}$ (as a function of u), we may bound

$$\nu \int_0^\infty \gamma_\ell(S_W(u))e^{-\nu u} du \leqslant \nu\gamma_\ell(S_W(m))e^{-\nu m/2} \int_0^\infty e^{-\nu u/2} du = 2\gamma_\ell(S_W(m))e^{-\nu m/2}.$$

Putting everything together we obtain

$$\mathbb{P}_\tau(\|W^T\tau - w\|_2 \leqslant 2\beta\sqrt{\ell}) \leqslant 2e^{\pi\beta^2\ell/2}e^{-\nu m/2}\gamma_\ell(S_W(m)).$$

□

The proof of Lemma 5.2 proceeds in much the same way.

Proof of Lemma 5.2. Let us set $X = \|W^T \cdot \tau\|_2$ and write

$$\mathbb{E}_X e^{-\pi X^2/2} = \mathbb{E}_X\, \mathbb{1}(X \leqslant \beta\sqrt{\ell})e^{-\pi X^2/2} + \mathbb{E}_X\, \mathbb{1}\big(X \geqslant \beta\sqrt{\ell}\big)e^{-\pi X^2/2} \leqslant \mathbb{P}_X(X \leqslant \beta\sqrt{\ell}) + e^{-\pi\beta^2\ell/2}$$

and therefore, using that $\exp(-\pi\beta^2\ell/2) \leqslant \exp(-\beta^2\ell)$,

$$\mathbb{E}_\tau \exp\left(\frac{-\pi\|W^T \cdot \tau\|_2^2}{2}\right) \leqslant \mathbb{P}_\tau(\|W^T \cdot \tau\|_2 \leqslant \beta\sqrt{\ell}) + e^{-\beta^2\ell}.$$

As before, we let φ be the characteristic function of $W^T\tau$, and let g be a standard ℓ-dimensional Gaussian random variable with standard deviation $(2\pi)^{-1/2}$. By Fact A.1 and (98) we obtain

$$\mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \cdot \tau\|_2^2}{2}\right) = \mathbb{E}_g[\varphi(g)] \geqslant \mathbb{E}_g[\exp(-32\mu\|Wg\|_{\mathbb{T}}^2)].$$

Similar to the proof of Lemma 3.2, we write

$$\mathbb{E}_g[\exp(-32\mu\|Wg\|_{\mathbb{T}}^2)] = 32\mu \int_0^\infty \gamma_\ell(S_W(u))e^{-32\mu u} du \geqslant 32\mu\gamma_\ell(S_W(t)) \int_t^\infty e^{-32\mu u}\, du,$$

where we have used that $\gamma_\ell(S_W(b)) \geqslant \gamma_\ell(S_W(a))$ for all $b \geqslant a$. This completes the proof of Lemma 5.2. □

Appendix B. Relating A to the zeroed out matrix M.

In this section we prove Lemma 8.1 and Lemma 9.6. To prove these results, we compare Fourier transforms (that is the *characteristic functions*) of the random variables Mv and Av, for fixed v. We first record the characteristic functions of these random variables. For $\xi \in \mathbb{R}^n$ we have

$$\psi_v(\xi) := \mathbb{E}\, e^{2\pi i \langle Av, \xi\rangle} = \left(\prod_{k=1}^{n} \cos(2\pi v_k \xi_k)\right) \cdot \left(\prod_{j<k} \big(2\pi(\xi_j v_k + \xi_k v_j)\big)\right)$$

and

$$\chi_v(\xi) := \mathbb{E}\, e^{2\pi i \langle Mv, \xi\rangle} = \prod_{j=1}^{d} \prod_{k=d+1}^{n} \left(\frac{3}{4} + \frac{1}{4}\cos\big(2\pi(\xi_j v_k + \xi_k v_j)\big)\right).$$

Our comparison is based on two main points. First we have that $\chi_v(\xi) \geqslant 0$. Second, we have

$$\psi_v(\xi) \leqslant \chi_v(2\xi)\,, \tag{99}$$

which follows from $|\cos(t)| \leqslant \frac{3}{4} + \frac{1}{4}\cos(2t)$ and $|\cos(t)| \leqslant 1$.

Fact B.1. *For $v \in \mathbb{R}^n$, and $t \geqslant \mathcal{T}_L(v)$, we have*

$$\mathbb{E}\exp(-\pi\|Mv\|_2^2/t^2) \leqslant (9Lt)^n.$$

Proof. Now $\mathbb{E}\exp(-\pi\|Mv\|_2^2/t^2)$ is at most

$$\mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) + \sqrt{n}\int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\, ds\,. \tag{100}$$

and since $t \geqslant \mathcal{T}_L(v)$, we have $\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n}) \leqslant (8Ls)^n$ for all $s \geqslant t$, and so we may bound

$$\sqrt{n}\int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\, ds \leqslant \sqrt{n}(8Lt)^n \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)(s/t)^n\, ds\,.$$

Changing variables $u = s/t$, the right hand side is equal to

$$t^{-1}\sqrt{n}(8Lt)^n \int_1^\infty \exp(-u^2 n) u^n\, du \leqslant t^{-1}\sqrt{n}(8Lt)^n \int_1^\infty \exp(-u^2/2)\, du \leqslant (9Lt)^n,$$

as desired. □

Proof of Lemma 8.1. Apply Markov's inequality to bound

$$\mathbb{P}(\|Av - w\|_2 \leqslant t\sqrt{n}) \leqslant \exp(\pi n/2)\mathbb{E}\,\exp\left(-\pi\|Av - w\|_2^2/2t^2\right)\,. \tag{101}$$

Using the Fourier inversion formula in Fact A.1 we write

$$\mathbb{E}_A\,\exp\left(-\pi\|Av - w\|_2^2/2t^2\right) = \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2} \cdot e^{-2\pi i t^{-1}\langle w, \xi\rangle}\psi_v(t^{-1}\xi)\, d\xi\,. \tag{102}$$

Rescaling, applying (99) and non-negativity of χ_v yields that the RHS of (102) is at most

$$\int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\chi_v(2t^{-1}\xi)\, d\xi \leqslant \mathbb{E}_M \exp(-2\pi\|Mv\|_2^2/t^2).$$

Now use Fact B.1 along with the assumption $t \geqslant \mathcal{T}_L(v)$ to obtain

$$\mathbb{E}_M \exp(-2\pi \|Mv\|_2^2/t^2) \leqslant (9Lt)^n,$$

as desired. □

We prove Lemma 9.6 in a similar manner. Recall $\rho_\varepsilon(v) = \max_{b\in\mathbb{R}^n} \mathbb{P}\left(\sum_i v_i\varepsilon_i \in (b-\varepsilon, b+\varepsilon)\right)$.

Proof of Lemma 9.6. Set $\varepsilon = \mathcal{T}_L(v)$ and let B be a $n \times n$ matrix uniformly drawn from all matrices with entries in $\{\pm 1\}$ and apply Markov's inequality to bound

$$\rho_\varepsilon(v)^n \leqslant \max_{w\in\mathbb{R}^n} \mathbb{P}(\|Bv - w\|_2 \leqslant \varepsilon\sqrt{n}) \leqslant \max_{w\in\mathbb{R}^n} \exp(\pi n/2)\mathbb{E}\exp\left(-\pi\|Bv-w\|_2^2/2\varepsilon^2\right). \quad (103)$$

Apply Fact A.1 to write

$$\mathbb{E}\exp\left(-\pi\|Bv-w\|_2^2/2\varepsilon^2\right) = \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2} \cdot e^{-2\pi i\varepsilon^{-1}\langle w,\xi\rangle} \prod_{1\leqslant j,k\leqslant n} \cos(2\pi\varepsilon^{-1}v_j\xi_k)\, d\xi \quad (104)$$

and use Hölder's inequality to bound the RHS of (104)

$$\leqslant \left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2/3}\, d\xi\right)^{3/4} \left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2} \prod_{1\leqslant j,k\leqslant n} \cos(2\pi\varepsilon^{-1}v_j\xi_k)^4\, d\xi\right)^{1/4}. \quad (105)$$

Now use $\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2/3}\, d\xi = \left(\frac{3}{2}\right)^{n/2}$ and $(\cos(a)\cos(b))^4 \leqslant \frac{3}{4} + \frac{1}{4}\cos(2(a+b))$, to see (105) is

$$\leqslant \left(\frac{3}{2}\right)^{3n/8} \left(2^{-n/2}\int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\chi_v(\sqrt{2}\varepsilon^{-1}\xi)\, d\xi\right)^{1/4} \leqslant \left(\frac{27}{128}\right)^{n/8} \left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4}. \quad (106)$$

Taken together, lines (103), (104), (105), (106) tell us that

$$\rho_\varepsilon(v)^n \leqslant (3/2)^{3n/8}(\exp(\pi/2)/\sqrt{2})^n \left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4}. \quad (107)$$

Now apply Fact B.1 to bound $\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right) \leqslant (9L\varepsilon)^n$ and so

$$\rho_\varepsilon(v)^n \leqslant (2^{12}L\varepsilon)^{n/4}.$$

□

Appendix C. Dealing with unstructured vectors

In this short section we will prove Lemma 9.1. This lemma is very similar to Lemma 2.1 in [7], which is essentially the same but over $\mathbb{Z}_p$. Their lemma, in turn, uses ideas from previous sources, mainly [8] but also [10, 31]. No originality is claimed on our part, for the contents of this appendix. Here we let A_m denote a $m \times m$ symmetric random matrix with entries in $\{-1, 1\}$, coupled so that A_{m-1} is A_m with the first row and columns removed.

We define the quantity

$$q_n(\gamma) = \max_{w\in\mathbb{R}^n} \mathbb{P}(\exists v \in \mathbb{R}^n \setminus \{0\} :\ A_n v = w,\ \rho(v) \geqslant \gamma)$$

and bound the singularity probability of A_n in terms of $q_n(\gamma)$.

Lemma C.1. *For $\gamma > 0$, we have*

$$\mathbb{P}(\det(A_n) = 0) \leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma} \right).$$

We prove Lemma 9.1 with aid of three further lemmas. The first is essentially [7, Lemma A.1] (see also [31, Section 2]).

Lemma C.2. *We have that*

$$\mathbb{P}(\det(A_n) = 0) \leqslant 4n \sum_{m=n}^{2n-2} \mathbb{P}(\mathrm{rk}(A_m) = m-1 \ \cap \ \mathrm{rk}(A_{m-1}) \in \{m-1, m-2\}). \tag{108}$$

Proof. This lemma appears as [7, Lemma A.1] but over $\mathbb{Z}_p$. If we apply that lemma with $p \gg n^n$ then all ranks are unchanged when viewed mod p and thus the lemma holds over $\mathbb{R}$. □

We deal with the sum (108), in two different cases: $\mathrm{rk}(A_{m-2}) = \mathrm{rk}(A_{m-1}) = m-1$ and $\mathrm{rk}(A_{m-1}) = m-2$, $\mathrm{rk}(A_m) = m-1$. We deal with this latter case first.

Lemma C.3. *We have*

$$\mathbb{P}(\mathrm{rk}(A_n) = n-1 \ \cap \ \mathrm{rk}(A_{n-1}) = n-2) \leqslant q_{n-1}(\gamma) + \gamma.$$

Proof. We define a map

$$\varphi : \{A_n : \mathrm{rk}(A_n) = n-1 \ \cap \ \mathrm{rk}(A_{n-1}) = n-2\} \to \mathbb{R}^{n-1},$$

so that $A = A_n$ satisfies $A_{[n]\times[2,n]}\varphi(A) = 0$. For this, let A_n be such that $\mathrm{rk}(A_n) = n-1$ and $\mathrm{rk}(A_{n-1}) = n-2$ and let X be the first row of A_n with the first entry removed. Notice that $X \in \mathrm{Im}(A_{n-1})$, otherwise it is not hard to see that $\mathrm{rk}(A_n) = \mathrm{rk}(A_{n-1}) + 2 = n$, using that A_n is symmetric. Thus there is $a \in \mathbb{R}^{n-1} \setminus \{0\}$ such that $A_{n-1} \cdot a = 0$ and $\langle a, X \rangle = 0$; further, since $\mathrm{rk}(A_{n-1}) = n-2$, this vector a is unique up to scalar multiples. Thus define $\varphi(A_n) := a$.

With φ in hand, we now bound

$$\mathbb{P}(\mathrm{rk}(A_n) = n-1 \ \cap \ \mathrm{rk}(A_{n-1}) = n-2) \leqslant \mathbb{P}(\mathrm{rk}(A_{n-1}) = n-2, \langle X, \varphi(A_{n-1}) \rangle = 0).$$

Considering the cases of $\rho(\varphi(A_{n-1})) \geqslant \gamma$ and $\rho(\varphi(A_{n-1})) < \gamma$ separately allows us to write

$$\mathbb{P}(\mathrm{rk}(A_{n-1}) = n-2, \langle X, \varphi(A_{n-1}) \rangle = 0) \leqslant q_{n-1}(\gamma) + \gamma,$$

as desired. □

We now treat the case when $\mathrm{rk}(A_n) = \mathrm{rk}(A_{n-1}) = n-1$. For this lemma, we adopt notation different from what we have been using in the body of the paper. If $v \in \mathbb{R}^m$ and $S \subset [m]$ we let $v_S \in \mathbb{R}^m$ be the vector $v_S = (v_k \mathbb{1}(k \in S))_{k \in [m]}$.

Lemma C.4. *For $t \in [n-2]$ we have*

$$\mathbb{P}(\mathrm{rk}(A_n) = n-1 \ \cap \ \mathrm{rk}(A_{n-1}) = n-1) \leqslant 3^t q_{n-1}(\gamma) + (2^t \gamma + 2^{-t})^{1/4}.$$

Proof. Let $I \cup J$ be a non-trivial partition of $[n-1]$ and set $\mathcal{A} = \{A_{n-1} : \det A_{n-1} \neq 0\}$. By a decoupling argument (e.g. line (9) of [10] or [7, Lemma A.9]) we have

$$\begin{aligned}
&\mathbb{P}(\mathrm{rk}(A_n) = \mathrm{rk}(A_{n-1}) = n-1) \\
&\leqslant \mathbb{E}_{A_{n-1}} \mathbb{P}\Big(\langle A_{n-1}^{-1}(X-X')_I, (X-X')_J \rangle = 0 \,\big|\, A_{n-1} \Big)^{1/4} \mathbb{1}(A_{n-1} \in \mathcal{A}) \qquad (109)
\end{aligned}$$

where X, X' are independent, and chosen uniformly at random from $\{1,-1\}^{n-1}$. Following [7], set

$$W(I) = \big\{ v \in \{-2,0,2\}^{n-1} : v_i = 0 \text{ for all } i \notin I \big\}$$

and

$$U_\gamma^{(I)} = \big\{ A_{n-1} : \rho(v) \leqslant \gamma \text{ for every } v \in \mathbb{R}^{n-1} \setminus \{0\} \text{ with } A_{n-1}v \in W(I) \big\} \,.$$

We bound (109) with three short claims.

Claim C.5. $\mathbb{P}(A_{n-1} \notin U_\gamma^{(I)}) \leqslant 3^{|I|} q_{n-1}(\gamma)$

Proof of Claim C.5. If $A_{n-1} \notin U_\gamma^{(I)}$ then there is a $v \in \mathbb{R}^{n-1} \setminus \{0\}$ with $\rho(v) \geqslant \gamma$ so that $A_{n-1}v \in W(I)$. Union bounding over $W(I)$ completes the claim. □

We now write $w = X - X'$ and for $A_{n-1} \in \mathcal{A}$, write $x = A_{n-1}^{-1} w_I$, as we see in (109).

Claim C.6. *For $A_{n-1} \in \mathcal{A}$ we have $\mathbb{P}(x = 0 \,|\, A_{n-1}) = 2^{-|I|}$.*

Proof of Claim C.6. Simply note that $\mathbb{P}(x = 0 \,|\, A_{n-1}) = \mathbb{P}(X_I = X'_I) = 2^{-|I|}$. □

Finally we have:

Claim C.7. *If $A_{n-1} \in U_\gamma^{(I)} \cap \mathcal{A}$ then*

$$\mathbb{P}_x(\langle x, w_J \rangle = 0 \text{ and } x \neq 0 \,|\, A_{n-1}) \leqslant 2^{|I|}\gamma \,.$$

Proof of Claim C.7. Since $A_{n-1}x = w_I$ with $x \neq 0$ and $A_{n-1} \in U_\gamma^{(I)}$, we have $\rho(x) \leqslant \gamma$. Conditioned on a given $x \neq 0$ with $\rho(x) \leqslant \gamma$ as well as A_{n-1} bound

$$\mathbb{P}_{X_J, X'_J}(\langle x, w_J \rangle = 0) = \mathbb{P}_{X_J, X'_J}(\langle x_J, X_J \rangle = \langle x_J, X'_J \rangle) \leqslant \rho(x_J) \leqslant 2^{|I|}\rho(x) \leqslant 2^{|I|}\gamma$$

where we have used [10, Lemma 2.9] for the bound $\rho(x_J) \leqslant 2^{|I|}\rho(x)$. □

Choosing $I = [t]$ and combining the previous three claims with (109) completes the proof of Lemma C.4. □

Proof of Lemma 9.1. For each $\gamma > 1/2$ we have $q_n(\gamma) \geqslant 2^{-n}$ and so we may assume that $\gamma < 1/n$ and $\gamma > 2^{-n}$. Set $t = \lfloor \log_4(1/\gamma) \rfloor \in [n-2]$ and apply Lemmas C.2, C.3 and C.4 to show

$$\begin{aligned}
\mathbb{P}(\det(A_n) = 0) &\leqslant 4n \sum_{m=n}^{2n-2} \left(\gamma + q_{m-1}(\gamma) + (3\gamma^{1/2})^{1/4} + \frac{q_{m-1}(\gamma)}{\gamma} \right) \\
&\leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma} \right) ,
\end{aligned}$$

thus completing the proof of Lemma 9.1. □

Appendix D. A few more details

We check the details on Lemma 5.7. We do this by deriving this lemma from a classical inequality due to Talagrand [42].

Theorem D.1 (Talagrand's Inequality)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex 1-Lipschitz function and $\sigma = (\sigma_1, \dots, \sigma_n)$ where σ_i are i.i.d. random variables with $|\sigma_i| \leqslant 1$. Then for any $t \geqslant 0$ we have*

$$\mathbb{P}\left(|F(\sigma) - m_F| \geqslant t\right) \leqslant 4 \exp\left(-t^2/16\right) ,$$

where m_F is the median of $F(\sigma)$.

From this, we derive the following consequence, which appears in [36] without explicit constants.

Lemma D.2. *For $d \in \mathbb{N}$, $\nu \in (0,1)$, let $\delta \in (0, \sqrt{\nu}/16)$, let $\sigma \sim \mathcal{Q}(2d, \nu)$, and let W be a $2d \times k$ matrix satisfying $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Then*

$$\mathbb{P}(\|W^T\sigma\|_2 \leqslant \delta\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k) \tag{110}$$

Proof. Note (110) is trivial if $k \leqslant 2^{12}/\nu$, so we assume $k > 2^{12}/\nu$. Define

$$F(x) := \|W\|^{-1}\|W^T x\|_2$$

and note that F is convex and 1-Lipschitz. Theorem D.1 immediately tells us that F is concentrated about the median m_F and so we only need to estimate m_F. For this, let us write down

$$m := \mathbb{E}\,\|W^T\sigma\|_2^2 = \sum_{i,j} W_{ij}^2 \mathbb{E}\,\sigma^2 = \nu\|W\|_{\mathrm{HS}}^2$$

and

$$m_2 := \mathbb{E}\,\|W^T\sigma\|_2^4 - (\mathbb{E}\,\|W^T\sigma\|_2^2)^2 = \sum_{i,j} W_{ij}^2\big(\mathbb{E}\,\sigma_i^4 - (\mathbb{E}\,\sigma_i^2)^2\big) \leqslant \nu\|W\|_{\mathrm{HS}}^2.$$

So for $t > 0$ we have, by Markov,

$$\mathbb{P}(\|W^T\sigma\|_2^2 \leqslant m - t) \leqslant t^{-2}\mathbb{E}\,\left(\|W^T\sigma\|_2^2 - m\right)^2 = t^{-2}m_2 \leqslant t^{-2}\nu\|W\|_{\mathrm{HS}}^2.$$

So set $t = \nu\|W\|_{\mathrm{HS}}^2/2$ to get

$$\mathbb{P}(\|W^T\sigma\|_2^2 \leqslant \nu\|W\|_{\mathrm{HS}}^2/2) \leqslant 4(\nu\|W\|_{\mathrm{HS}}^2)^{-1} < 1/2\,,$$

since $\|W\|_{\mathrm{HS}}^2 \geqslant k/4 > 8\nu^{-1}$. It follows that

$$m_F \geqslant \|W\|^{-1}\sqrt{\nu/2}\|W\|_{\mathrm{HS}}\,. \tag{111}$$

Now we may apply Talagrand's Inequality D.1 with $t = m_F - \delta\sqrt{k}\|W\|^{-1}$ to obtain

$$\mathbb{P}\left(\|W^T\sigma\|_2 \leqslant \delta\sqrt{k}\right) \leqslant 4\exp\left(-t^2/16\right) .$$

To complete the proof we note that

$$t = m_F - \delta\sqrt{k}\|W\|^{-1} \geqslant \|W\|^{-1}(\sqrt{\nu/2}\|W\|_{\mathrm{HS}} - \delta\sqrt{k}) \geqslant \sqrt{\nu k}/16\,,$$

where we have used (111), $\|W\| \leqslant 2$, $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\delta \leqslant \sqrt{\nu}/8$. □

The following integral appears in the proof of Lemma 6.3.

Fact D.3. *For $k \geqslant 0$, we have the integral inequality*

$$\int_{\sqrt{k+1}}^{\infty} \left(1 + \frac{2u}{\sqrt{k+1}}\right)^{k+2} u e^{-2u^2}\, du \leqslant 2.$$

Proof. Using the inequality $1 + x \leqslant e^{x^2/3}$ for $x \geqslant 2$, we have

$$\int_{\sqrt{k+1}}^{\infty} \left(1 + \frac{2u}{\sqrt{k+1}}\right)^{k+2} u e^{-2u^2}\, du \leqslant \int_{\sqrt{k+1}}^{\infty} \left(1 + \frac{2u}{\sqrt{k+1}}\right) u e^{-2u^2/3}\,.$$

Since $k \geqslant 0$, the right hand side is at most

$$\int_{1}^{\infty} u\,(1+2u)\, e^{-2u^2/3} \leqslant 2\,.$$

□

Appendix E. Proof of Lemma 6.4

We will require the following standard concentration result for the norm of a random matrix with independent, mean-zero, sub-Gaussian entries (see, e.g. [52, Theorem 4.4.5]).

Theorem E.1. *Let B be an $m \times n$ random matrix with independent, mean zero entries with $\max_{i,j} |B_{ij}| \leqslant K$. Then for $t > 0$ we have*

$$\mathbb{P}(\|B\| \geqslant 8K(\sqrt{m} + \sqrt{n} + t)) \leqslant 2\exp(-t^2)\,.$$

We remark that the explicit constant 8 does not appear in the statement of [52, Theorem 4.4.5], but can easily be extracted from the proof.

We also note the basic bound on the number of integer points in a ball,

$$|B_{r\sqrt{n}}(0) \cap \mathbb{Z}^n| \leqslant |B_{(r+1)\sqrt{n}}(0)| \leqslant (4r)^n,$$

where the first inequality holds by the fact that the boxes

$$p + \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]^{2kd} \text{ where } p \in \mathbb{Z}^{2kd} \cap B_{16\sqrt{kd}/\delta}(0)$$

are pairwise disjoint, have volume 1 and are contained in $B_{(r+1)\sqrt{n}}(0)$.

Proof of Lemma 6.4. We claim that the following net works for Lemma 6.4. Define

$$\mathcal{N} := \left\{ W \in \mathbb{R}^{[2d]\times[k]} :\ W_{i,j} \in \frac{\delta}{8\sqrt{d}} \cdot \mathbb{Z} \text{ and } \|W\|_{\mathrm{HS}} \leqslant 2\sqrt{k} \right\}.$$

To bound $|\mathcal{N}|$, note that $(8\sqrt{d}/\delta) \cdot \mathcal{N} \subseteq B_{16\sqrt{kd}/\delta}(0) \cap \mathbb{Z}^{2kd}$. Thus

$$|\mathcal{N}| \leqslant |B_{16\sqrt{kd}/\delta}(0) \cap \mathbb{Z}^{2kd}| \leqslant (2^6/\delta)^{2dk}.$$

Now let $U \in \mathcal{U}_{2d,k}$ and A be a $r \times 2d$ matrix. We show that there exists a $W \in \mathcal{N}$ satisfying properties (1), (2) and (3). We find such a W by randomly rounding W onto our net $\mathcal{N}$ and showing that the properties each hold with probability $> 3/4$.

For this, let $U' = \frac{8\sqrt{d}}{\delta} U$ and let W' be a $2d \times k$ random matrix with independent entries where $W'_{i,j} \in \{\lfloor U'_{i,j} \rfloor, \lceil U'_{i,j} \rceil\}$ and $\mathbb{E} W'_{i,j} = U'_{i,j}$ and let $W = \frac{\delta}{8\sqrt{d}} W'$. We remark that for *all* possible W, we have

$$\|W\|_{\mathrm{HS}} \leqslant \|U\|_{\mathrm{HS}} + \delta\sqrt{k} \leqslant 2\sqrt{k},$$

and so we always have $W \in \mathcal{N}$.

For Property (1), we write

$$\mathbb{E}_W \|A(W-U)\|_{\mathrm{HS}}^2 = \sum_{i,j} \mathbb{E} \langle A^T e_j, (W-U)e_i \rangle^2 = \sum_{i,j,\ell} A_{j\ell}^2 \mathbb{E} (W_{\ell i} - U_{\ell i})^2 \leqslant \frac{\delta^2 \|A\|_{\mathrm{HS}}^2 k}{64d},$$

using that $\mathbb{E} (W_{\ell i} - U_{\ell i})^2 \leqslant \frac{\delta^2}{64d}$ in the last inequality. Thus by Markov's inequality

$$\mathbb{P}_W \left(\|A(W-U)\|_{\mathrm{HS}} \leqslant \frac{\delta \|A\|_{\mathrm{HS}} \sqrt{k}}{\sqrt{2d}} \right) \geqslant 3/4. \tag{112}$$

For Property (3), we repeat the same argument, replacing A with a $2d \times 2d$ identity matrix. Thus Property (3) also holds with probability $> 3/4$ in the choice of W.

For property (2), note that $W - U$ is a random matrix with independent random entries satisfying $\mathbb{E}_W (W-U)_{ij} = 0$ and $|(W-U)_{ij}| \leqslant \frac{\delta}{8\sqrt{d}}$ for all i, j. We may therefore apply Theorem E.1 with $B = W - U$ and $K = \frac{\delta}{8\sqrt{d}}$ to get

$$\mathbb{P}_W (\|W - U\| \leqslant 8\delta) \geqslant 3/4\,. \tag{113}$$

Thus W satisfies properties (1),(2),(3), each with probability $> 3/4$ and therefore there must exist a W satisfying all three properties. □

References

[1] Z.-D. Bai and Y.-Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[2] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.

[3] B. Bollobás. *Random graphs*. Number 73. Cambridge university press, 2001.

[4] C. Borell. Inequalities of the Brunn–Minkowski type for Gaussian measures. *Probability theory and related fields*, 140(1-2):195–205, 2008.

[5] J. Bourgain, V. H. Vu, and P. M. Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[6] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. Singularity of random symmetric matrices revisited. *arXiv preprint arXiv:2011.03013*, 2020.

[7] M. Campos, L. Mattos, R. Morris, and N. Morrison. On the singularity of random symmetric matrices. *Duke Mathematical Journal*, 170(5):881–907, 2021.

[8] K. P. Costello, T. Tao, and V. Vu. Random symmetric matrices are almost surely nonsingular. *Duke Math. J.*, 135(2):395–413, 2006.

[9] P. Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[10] A. Ferber and V. Jain. Singularity of random symmetric matrices—a combinatorial approach to improved bounds. *Forum Math. Sigma*, 7:Paper No. e22, 29, 2019.

[11] A. Ferber, V. Jain, K. Luh, and W. Samotij. On the counting problem in inverse Littlewood–Offord theory. *Journal of the London Mathematical Society*.

[12] A. Ferber, V. Jain, and Y. Zhao. On the number of Hadamard matrices via anti-concentration. *arXiv preprint arXiv:1808.07222*, 2018.

[13] P. Frankl and Z. Furedi. Solution of the Littlewood-Offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.

[14] J. R. Griggs, J. C. Lagarias, A. M. Odlyzko, and J. B. Shearer. On the tightest packing of sums of vectors. *European Journal of Combinatorics*, 4(3):231–236, 1983.

[15] G. Halász. On the distribution of additive arithmetic functions. *Acta Arithmetica*, 1(27):143–152, 1975.

[16] V. Jain, A. Sah, and M. Sawhney. On the smallest singular value of symmetric random matrices. *arXiv preprint arXiv:2011.02344*, 2020.

[17] V. Jain, A. Sah, and M. Sawhney. Singularity of discrete random matrices I. *arXiv preprint arXiv:2010.06553*, 2020.

[18] V. Jain, A. Sah, and M. Sawhney. Singularity of discrete random matrices II . *arXiv preprint arXiv:2010.06554*, 2020.

[19] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[20] G. Katona. On a conjecture of Erdős and a stronger form of Sperner's theorem. *Studia Sci. Math. Hungar*, 1:59–63, 1966.

[21] D. J. Kleitman. On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Advances in Mathematics*, 5(1):155–157, 1970.

[22] J. Komlós. On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar.*, 2:7–21, 1967.

[23] J. Komlós. On the determinant of random matrices. *Studia Sci. Math. Hungar.*, 1968.

[24] J. E. Littlewood and A. C. Offord. On the Number of Real Roots of a Random Algebraic Equation. *J. London Math. Soc.*, 13(4):288–295, 1938.

[25] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.

[26] A. E. Litvak and K. E. Tikhomirov. Singularity of sparse bernoulli matrices. *arXiv preprint arXiv:2004.03131*, 2020.

[27] G. V. Livshyts. The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding. *arXiv preprint arXiv:1811.07038*, 2018.

[28] G. V. Livshyts, K. Tikhomirov, and R. Vershynin. The smallest singular value of inhomogeneous square random matrices. *The Annals of Probability*, 49(3):1286–1309, 2021.

[29] M. W. Meckes. Concentration of norms and eigenvalues of random matrices. *Journal of Functional Analysis*, 211(2):508–524, 2004.

[30] H. Nguyen and V. Vu. Optimal inverse Littlewood-Offord theorems. *Adv. Math.*, 226(6):5298–5319, 2011.

[31] H. H. Nguyen. Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Math. J.*, 161(4):545–586, 2012.

[32] H. H. Nguyen and V. H. Vu. Small probability, inverse theorems, and applications. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 409–463. János Bolyai Math. Soc., Budapest, 2013.

[33] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[34] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[35] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1576–1602. Hindustan Book Agency, New Delhi, 2010.

[36] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18, 2013.

[37] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617, 2015.
[38] A. Sali. Stronger form of an m-part Sperner theorem. *European Journal of Combinatorics*, 4(2):179–183, 1983.
[39] A. Sárközy and E. Szemeredi. Über ein problem von Erdös und Moser. *Acta Arithmetica*, 11(2):205–208, 1965.
[40] D. Saxton and A. Thomason. Hypergraph containers. *Inventiones mathematicae*, 201(3):925–992, 2015.
[41] R. P. Stanley. Weyl groups, the hard Lefschetz theorem, and the Sperner property. *SIAM Journal on Algebraic Discrete Methods*, 1(2):168–184, 1980.
[42] M. Talagrand. A new look at independence. *The Annals of Probability*, 24(1):1 – 34, 1996.
[43] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.
[44] T. Tao and V. Vu. On random ± 1 matrices: singularity and determinant. *Random Structures Algorithms*, 28(1):1–23, 2006.
[45] T. Tao and V. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.
[46] T. Tao and V. Vu. A sharp inverse Littlewood-Offord theorem. *Random Structures & Algorithms*, 37(4):525–539, 2010.
[47] T. Tao and V. Vu. The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica*, 32(3):363–372, 2012.
[48] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.
[49] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.
[50] K. Tikhomirov. Singularity of random Bernoulli matrices. *Ann. of Math. (2)*, 191(2):593–634, 2020.
[51] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures Algorithms*, 44(2):135–182, 2014.
[52] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
[53] V. Vu. Random discrete matrices. In *Horizons of combinatorics*, volume 17 of *Bolyai Soc. Math. Stud.*, pages 257–280. Springer, Berlin, 2008.
[54] V. H. Vu. Recent progress in combinatorial random matrix theory. *Probability Surveys*, 18:179–200, 2021.

Instituto de Matemática Pura e Aplicada (IMPA).
Email address: marcelo.campos@impa.br

University of Birmingham, School of Mathematics.
Email address: m.jenssen@bham.ac.uk

University of Illinois at Chicago. Department of Mathematics, Statistics and Computer Science.
Email address: michelen.math@gmail.com

University of Cambridge. Department of Pure Mathematics and Mathematics Statistics.
Email address: jdrs2@cam.ac.uk