JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 38, Number 1, January 2025, Pages 179–224
https://doi.org/10.1090/jams/1042
Article electronically published on January 19, 2024

THE SINGULARITY PROBABILITY OF A RANDOM SYMMETRIC MATRIX IS EXPONENTIALLY SMALL

MARCELO CAMPOS, MATTHEW JENSSEN, MARCUS MICHELEN, AND JULIAN SAHASRABUDHE

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 [45, 46] 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 [50], 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 led them to prove an elegant inverse Littlewood-Offord type result, inspired by [50], in a geometric setting.

Received by the editors August 27, 2021, and, in revised form, October 17, 2023.
2020 *Mathematics Subject Classification.* Primary 60B20, 15A18.
The first author was partially supported by CNPq. The second author was supported by a UKRI Future Leaders Fellowship MR/W007320/1. The third author was supported in part by NSF grants DMS-2137623 and DMS-2246624.

This perspective was then developed further in the 2018 breakthrough work of Tikhomirov [51], 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 [51] 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 advance on the problem was made recently by Jain, Sah and Sawhney [17], 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. [9, 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 [9], 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 [52], the latter obtaining a bound of the form $\exp(-n^c)$. Nguyen [31] developed the quadratic Littlewood-Offord theory introduced in [9], while Vershynin [52] 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 [12] and applied by Ferber and Jain [11] 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 [8] then improved this bound to

$$\mathbb{P}(\det(A) = 0) \leqslant \exp(-c\sqrt{n}), \tag{3}$$

by proving a"rough" inverse Littlewood-Offord theorem, inspired by the theory of hypergraph containers (see [2,41]). This bound was then improved by Jain, Sah and Sawhney [18], 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 $-c\sqrt{n\log n}$.

The convergence of these results onto the exponent of $-c\sqrt{n \log n}$ 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 [8], if one is restricted to working in the top-half of A one cannot obtain an exponent better than $-c\sqrt{n \log n}$. 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 random symmetric 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 2).

1.1. **Inverse Littlewood-Offord theory.** For $v \in \mathbb{R}^n$ and X uniform in $\{-1, 1\}^n$, we define the concentration function (one of several to come) as

$$\rho(v) = \max_{b \in \mathbb{R}} \mathbb{P}\big(\langle v, X \rangle = b\big). \tag{5}$$

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 [10]: if $v \in \mathbb{R}^n$ has all non-zero coordinates then

$$\rho(v) \leqslant \rho((1, \dots, 1)) = O(n^{-1/2}).$$

This was then developed by Sárközy and Szemerédi [40], 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 [42] who determined the *exact* maximum in this case. A higher-dimensional version of this problem also received considerable attention and was studied by several authors [15, 20, 21, 39] before it was ultimately resolved in the work of Frankl and Füredi [14] (see also [48]).

Of these early results, the most important for us here is the work of Halász [16] 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 [50], 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 [47, 50], and Nguyen and Vu [30, 32] proved that if v is such that $\rho(v) > n^{-C}$ then all but $O(n^{1-\varepsilon})$ 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 [12], and the "weak" inverse Littlewood-Offord theorems of Campos, Mattos, Morris and Morrison [8] 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, 34, 52], who showed that if $\rho(v) \gg e^{-cn}$ then there exists $\phi > 0$ with $|\phi| = O(1/\rho(v))$ for which the dilated vector ϕv is exceptionally close to the integer lattice $\mathbb{Z}^n$. In particular, Rudelson and Vershynin define the following important notion. 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 ϕv is within $\sqrt{\alpha d}$ of a non-zero integer point. That is,

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

where $d(x,S)$ denotes $\inf_{s\in S}\{\|x-s\|_2\}$ (not to be confused with the dimension d). Note here that we have excluded the origin from $\mathbb{Z}^d$ in the definition since $\phi v \approx 0$ does not tell us anything interesting about v. Indeed, given *any* $v \in \mathbb{S}^{d-1}$, one could always set $\phi < \sqrt{\alpha d}$ and obtain $d(\phi v, \mathbb{Z}^d) \leqslant d(\phi v, 0) \leqslant \sqrt{\alpha d}$, 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 (29)). Here we state the theorem of Rudelson and Vershynin in a slightly less general form than they prove.

Theorem 1.2. *For $d \in \mathbb{N}$, $\alpha \in (0,1)$ and $t > 0$, let $v \in \mathbb{S}^{d-1}$ satisfy $D_\alpha(v) > 16/t$. If $X \sim \{-1,1\}^d$ is uniform then*

$$\mathbb{P}\big(|\langle X, v\rangle| \leqslant t\big) \leqslant Ct + 2e^{-c\alpha d}.$$

Here $C, c > 0$ are absolute constants.

Thus we can think of $D_\alpha(v)$ as a measure of the arithmetic structure of v; a small value of $D_\alpha(v)$ corresponds to more structure, a large value of $D_\alpha(v)$ to less.

Our Littlewood-Offord theorem shows that a similar conclusion can be obtained in the presence of a large number ($k \approx n$) of additional "soft" constraints on the random walk. We prove the following result, which is in fact weaker than what we really need (see Lemma 4.1), but captures its essence.

Theorem 1.3. *For $d \in \mathbb{N}$, $\alpha \in (0,1)$ and $t > 0$, let $v \in \mathbb{S}^{d-1}$ satisfy $D_\alpha(v) > 16/t$. For $0 \leqslant k \leqslant d$, let W be a $k \times d$ matrix with orthonormal rows. If $X \sim \{-1,1\}^d$ is uniform then*

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant t \text{ and } \|WX\|_2 \leqslant c\sqrt{k}\right) \leqslant Cte^{-ck} + 2e^{-c\alpha d}, \tag{6}$$

where $C, c > 0$ are absolute constants.

Note that if $k = 0$ then our theorem reduces to Rudelson and Vershynin's theorem, stated above. Here we interpret $\|WX\|_2 \leqslant c\sqrt{k}$ as encoding the k "soft" constraints and $|\langle X, v\rangle| \leqslant t$ as the "hard" constraint. It is also useful to think of

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

$t \approx \rho(v)$, although we actually apply this theorem with t chosen with respect to a related notion, tailored specifically to our application.

To understand the numerology of this theorem, it is perhaps best to think of it as a result that allows us to "decouple" the hard constraint from the k soft constraints. It says something to the effect of, if $D_\alpha(v) > 16/t$ then

$$\mathbb{P}\left(|\langle X, v\rangle| \leqslant t \text{ and } \|WX\|_2 \leqslant c\sqrt{k}\right) \leqslant C \cdot \mathbb{P}\big(|\langle X, v\rangle| \leqslant t\big) \cdot \mathbb{P}\big(\|WX\|_2 \leqslant c\sqrt{k}\big). \tag{7}$$

Given this, we see that Rudelson and Vershynin's theorem and the Hanson-Wright inequality allow us to deal with these two quantities in isolation. These say that

$$\mathbb{P}\big(|\langle X, v\rangle| \leqslant t\big) \leqslant Ct + e^{-c\alpha d} \qquad \text{and} \qquad \mathbb{P}\big(\|WX\|_2 \leqslant c\sqrt{k}\big) \leqslant e^{-ck}, \tag{8}$$

thus explaining the form of the conclusion of Theorem 1.3.

While we don't prove exactly (7), the main difficulty for us lies in decoupling the soft and hard constraints, which is ultimately achieved by a somewhat complicated geometric argument on the Fourier side and will consume our focus in Sections 4, 5 and 6.

It is useful to compare our Theorem 1.3 to a mutlidimensional version of Theorem 1.2 proved by Rudelson and Vershynin Theorem 7.5 in [38]. Using their theorem, one could prove a version of our Theorem 1.3 if one added the additional assumption that $D_\alpha(u)$ is large for *all* unit vectors u that are obtained as certain[2] linear combinations of v with the rows of W. This is insufficient for us as Theorem 1.3 assumes no information about the the structure of the space spanned by the rows of W.

2. Proof sketch and our novel "inversion of randomness" technique

Here we sketch the proof of Theorem 1.1, assuming our Littlewood-Offord theorem (Theorem 1.3) and show how it fits into our novel "inversion of randomness" technique, which allows us to overcome the barrier encountered in previous works. We highlight this main new idea in Section 2.3 after warming-up with some more general discussion of our approach.

Throughout this section we keep our discussion loose and impressionistic and only take up our careful study in the following sections.

2.1. Setup.

A matrix is singular if and only if there exists $v \in \mathbb{S}^{n-1}$ such that $Av = 0$. A central challenge in studying the singularity probability of discrete random matrices lies in the fact that different v have vastly different probabilities of being in the kernel of A. For example, it is easy to see that if

$$v = 2^{-1/2}(1, 1, 0, \dots, 0) \qquad \text{then} \qquad \mathbb{P}(Av = 0) = 2^{-n}. \tag{9}$$

If $v = n^{-1/2}(1, \dots, 1)$ it is *significantly* harder to determine the corresponding probability, but one's first guess actually resembles the truth; the probability that the first entry of Av is 0 is $\Theta(n^{-1/2})$, the probability a simple random walk returns to 0 after n steps. Thus, boldly assuming the approximate independence of the rows, we expect that if

$$v = n^{-1/2}(1, \dots, 1) \quad \text{then} \qquad \mathbb{P}(Av = 0) \approx (Cn)^{-n/2}, \tag{10}$$

[2]Specifically, if we let w_i be the rows of W, we are interested in all linear combinations of the form $\theta_0 v + \sum_{i=1}^{k} \theta_i w_i$, where $|\theta_0| \leqslant C/\varepsilon$ and $|\theta_i| < C$, for $i = 1, \dots, k$

a very different result from (9). But these are both very structured and special examples. The opposite extreme comes from a *random* vector $v \sim \mathbb{S}^{n-1}$ on the unit sphere. Here we have to be a bit careful since there are only a finite number of possible kernel vectors of a discrete random matrix and thus it is natural to instead consider the probability that a random vector is ε-far from the kernel, for some well-chosen $\varepsilon > 0$. Again this case is not easy to establish rigorously, but in a similar way to the above, for $\varepsilon > e^{-cn}$, we expect if

$$v \sim \mathbb{S}^{n-1} \qquad \text{is random then} \qquad \mathbb{P}\big(\|Av\|_2 \leqslant \varepsilon\sqrt{n}\big) \approx (C\varepsilon)^n, \tag{11}$$

with high probability, where $C > 0$ represents a constant that is unimportant for us.

Thus we see that there is a great variety in how different directions contribute to the singularity probability. For us the key task is to understand how "many" of each of these different directions there are. For example there are about n^2 different vectors v of type (9), and this multiplied by the probability that one of these vectors is in the kernel represents the conjectured asymptotic for the singularity probability. On the other hand, there are about 2^n vectors of the type (10), thus the expected contribution of these vectors to the singulairty probability is significantly less than (9).

The crux comes with estimating the quantity of vectors that fail to be of type (11), for each $\varepsilon > e^{-cn}$: we would like to say that extremely few v deviate from this heuristic, at any given scale $\varepsilon > 0$. Here it does not quite make sense to *count* the number of such offending vectors, since there are infinitely many; rather, we "capture" these vectors by building efficient ε-nets for them. From this point of view, this is the main technical content of this paper.

2.2. **Definition of the ε-nets.** To define our ε-nets we would like to associate each vector $v \in \mathbb{S}^{n-1}$ with a *scale* $\varepsilon = \varepsilon(v)$. Essentially, though we define things a bit differently in the proof, we define the *scale* of a vector v to be the maximum $\varepsilon \in (0, 1)$ for which

$$\mathbb{P}\big(\|Av\|_2 \leqslant \varepsilon\sqrt{n}\big) \geqslant (L\varepsilon)^n, \tag{12}$$

where L is a large constant $L \gg C$. Intuitively speaking, the scale of the vector v is the largest granularity at which our heuristic (11) fails. Importantly, we can prove that at this scale we also have the reverse inequality $\mathbb{P}\big(\|Av\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant (CL\varepsilon)^n$.

Vectors that have scales that are smaller than e^{-cn} can be dealt with using now-standard ideas dating back to Costello, Tao and Vu [9]. As such, our focus will be on eliminating vectors with scales $\varepsilon > e^{-cn}$. It will also be easy, in light of previous work, to ignore "compressible" vectors, that is, vectors that have almost all of their ℓ_2-mass on $o(n)$ coordinates. Thus we can restrict to vectors which have at least $\Omega(n)$ coordinates of magnitude $\Theta(n^{-1/2})$. Let $\mathbb{S}_0^{n-1}$ denote this subset of the sphere; without loss of generality, we can assume that these coordinates are the first d and that $d/n = \Theta(1)$, but chosen to be sufficiently small.

For each $\varepsilon > e^{-cn}$, we would like to build an ε-net for all $v \in \mathbb{S}_0^{n-1}$ at scale ε. Our first move is to start with a decent ε-net for *all of* $\mathbb{S}_0^{n-1}$, which we will call Λ_ε, and then define a subset $\mathcal{N}_\varepsilon \subset \Lambda_\varepsilon$, which will serve as our desired ε-net. We note that the most efficient ε-nets for the whole of $\mathbb{S}_0^{n-1}$ are of size $(C/\varepsilon)^n$, which is vastly too large for us and thus $\mathcal{N}_\varepsilon$ must be substantially smaller than Λ_ε. Indeed,

we need something like

$$|\mathcal{N}_\varepsilon| \leqslant L^{-2n}|\Lambda_\varepsilon| \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n, \tag{13}$$

since we will be taking a union bound over $|\mathcal{N}_\varepsilon|$ events of the form $\|Av\|_2 \leqslant \varepsilon\sqrt{n}$, each with has probability at most $(CL\varepsilon)^n$, from the remark below (12).

We now prepare for the definition of $\mathcal{N}_\varepsilon$. For this we first introduce a different model of a random symmetric matrix, that is slightly cleaner to work with, and which we will be able to "swap" for A, in the proof. We define the random matrix

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H^T \\ H & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}, \tag{14}$$

where H is a $(n-d)\times d$ random matrix with iid entries that are 1/4-lazy, meaning that $H_{i,j} = 0$ with probability 3/4 and $H_{i,j} = \pm 1$ with probability 1/8. The key property that we have here is that for all v,

$$\mathbb{P}\big(\|Av\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant C^n \cdot \mathbb{P}\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big),$$

which we establish on the Fourier side, akin to [19]. We now crucially define[3] our ε-net

$$\mathcal{N}_\varepsilon = \big\{v \in \Lambda_\varepsilon : \mathbb{P}\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \geqslant (L\varepsilon)^n\big\}. \tag{15}$$

It turns out that it is not too hard to show that this is an ε-net; to do so, we simply adapt some now-standard random rounding techniques [27] to this higher dimensional setting. The real challenge lies in estimating the *size* of $\mathcal{N}_\varepsilon$. For this we take a *probabilistic* vantage point (inspired by [51]) and it is this new source of randomness that helps us "recover" some of the randomness lost due to the symmetry of A. To prove (13), it is enough to show, for $v \in \Lambda_\varepsilon$ chosen uniformly at random, that

$$\mathbb{P}_{v\in\Lambda_\varepsilon}\big(v \in \mathcal{N}_\varepsilon\big) = \mathbb{P}_{v\in\Lambda_\varepsilon}\Big(\mathbb{P}_M\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \geqslant (L\varepsilon)^n\Big) \leqslant (C/L^2)^n. \tag{16}$$

(see Lemma 8.3 for the precise statement.) To get a feel for how we tackle this, let us consider the event $\|Mv\|_2 \leqslant \varepsilon\sqrt{n}$. Indeed recalling the definition (14) of M, we have that

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

and so to control the event $\|Mv\|_2 \leqslant \varepsilon\sqrt{n}$, it is enough to control the intersection of events

$$\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n} \qquad \text{and} \qquad \|H^T v_{[d+1,n]}\|_2 \leqslant \varepsilon\sqrt{n}. \tag{17}$$

Note that if we simply ignore the second event and bound

$$\mathbb{P}_M\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant \mathbb{P}_H\big(\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big),$$

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

To prove (16) we use a *second moment* argument. For now, however, we will limit ourselves to a discussion of the first moment and then comment on the extra

[3]We actually use a slightly smaller net, see (27) for the formal definition.

complications in working with the second moment. In particular, we outline a proof of the inequality

$$\mathbb{E}_{v\in\Lambda_\varepsilon}\mathbb{P}_M\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant (C\varepsilon)^n, \tag{18}$$

which implies that $|\mathcal{N}_\varepsilon| \leqslant (C/L)^n|\Lambda_\varepsilon| \leqslant (C/\varepsilon L)^n$, by Markov's inequality:

$$\frac{|\mathcal{N}_\varepsilon|}{|\Lambda_\varepsilon|} = \mathbb{P}_{v\in\Lambda_\varepsilon}\big(v\in\mathcal{N}_\varepsilon\big) = \mathbb{P}_{v\in\Lambda_\varepsilon}\Big(\mathbb{P}_M\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \geqslant (L\varepsilon)^n\Big) \leqslant \left(\frac{C}{L}\right)^n.$$

This falls short of (13), for which we will need to control the second moment, but is a good starting point.

2.3. **Rank splitting and inversion of randomness.** In understanding (18) we come to our novel "inversion of randomness" technique that allows us to weave the randomness of v into our arguments. The idea is to use the *randomness in* H to control the first event at (17) and to use the *randomness in* $v\in\Lambda_\varepsilon$ to control the second. To get this to work, we crucially partition the outcomes of H, based on a robust notion of rank. Indeed, let $\mathcal{E}_k$ be the event that all but k of the singular values of H are "healthy"

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

where $\sigma_1(H) \geqslant \cdots \geqslant \sigma_d(H)$ denote the singular values of H. The point of this definition is that it allows us to get some mileage out of the second event at (17). At this point it is useful to point out that we may assume that the coordinates of $v\sim\Lambda_\varepsilon$ are iid random variables, which follows from an easy covering argument of Λ_ε with product sets, as in [51]. Now, if $H\in\mathcal{E}_k$ is a *fixed* matrix, we prove, using only the randomness in $v_{[d+1,n]}$, that

$$\mathbb{P}_{v_{[d+1,n]}}\big(\|H^T v_{[d+1,n]}\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant (C\varepsilon)^{d-k}. \tag{19}$$

We prove (19) by adapting the main result of [37]. On the other hand, using only the randomness in H, we bound $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon\sqrt{n})$ from above by

$$\sum_{k=0}^{d} \mathbb{P}_H\Big(\|H^T v_{[d+1,n]}\|_2 \leqslant \varepsilon\sqrt{n} \mid \big\{\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big\}\cap\mathcal{E}_k\Big)\mathbb{P}_H\Big(\big\{\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big\}\cap\mathcal{E}_k\Big). \tag{20}$$

So to prove (18), we average (20) over all $v\in\Lambda_\varepsilon$ and use (19) to bound the first term in each summand to obtain

$$\mathbb{E}_{v\in\Lambda_\varepsilon}\mathbb{P}_M\big(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\big) \leqslant (C\varepsilon)^d\cdot\sum_{k=0}^{d}(C\varepsilon)^{-k}\cdot\mathbb{E}_{v\in\Lambda_\varepsilon}\mathbb{P}_H\Big(\big\{\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big\}\cap\mathcal{E}_k\Big), \tag{21}$$

where we have used the independence of $v_{[d]}$ from $v_{[d+1,n]}$.

In dealing with the remaining probabilities in the sum at (21) we use our new inverse Littlewood-Offord theorem, Theorem 1.3. We first note that

$$\mathbb{P}_H\Big(\big\{\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big\}\cap\mathcal{E}_k\Big) \leqslant \mathbb{P}_H\Big(\big\{\|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}\big\}\cap\big\{\sigma_{d-k+1}(H) < c\sqrt{n}\big\}\Big), \tag{22}$$

and then observe that since the rows of H are independent, the probability on the right-hand side of (22) should approximately factor as the product of "one-dimensional" events, corresponding to each row. In particular, we show the right-hand side of (22) is at most

(23)
$$C^{n-d} \cdot \max_{u_1,\dots,u_k} \left(\mathbb{P}_X \left(|\langle X, v_{[d]} \rangle| \leqslant \varepsilon |\langle X, u_1 \rangle| \leqslant cn^{-1/2}, \dots, |\langle X, u_k \rangle| \leqslant cn^{-1/2} \right) \right)^{n-d},$$

using a (considerably easier) ε-net argument along with a tensorization argument. Here $X \sim \{-1, 0, 1\}^d$ is distributed as a row of H and the maximum is taken over all orthonormal k-tuples $u_1, \dots, u_k \in \mathbb{R}^d$ which correspond to the k orthonormal singular directions of H that witness the event $\sigma_{d-k+1}(H) < c\sqrt{n}$.

We now observe that the probability in (23) is exactly the sort of quantity that we can bound with our Littlewood-Offord theorem. There is a slight wrinkle here in that we need to ensure $D_\alpha(v_{[d]}) > 16/\varepsilon$, but this is a technicality we can deal with earlier in the proof by directly bounding the probability a random $v \in \Lambda_\varepsilon$ has $D_\alpha(v_{[d]}) \leqslant 16/\varepsilon$. Thus we can apply Theorem 1.3 to bound (23) and hence obtain

$$\mathbb{P}_H \left(\{ \|Hv_{[d]}\|_2 \leqslant \varepsilon\sqrt{n} \} \cap \{ \sigma_{d-k+1}(H) \leqslant c\sqrt{n} \} \right) \leqslant (C\varepsilon e^{-ck})^{n-d}. \tag{24}$$

We then apply this bound to each term in (21), by way of (22), to see that

$$\mathbb{E}_{v \in \Lambda_\varepsilon} \mathbb{P}_M \left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n} \right) \leqslant (C\varepsilon)^n,$$

where we have used that $k \leqslant d \leqslant n$, d/n is small compared to $1/C$ and c and that $\varepsilon > e^{-cn}$. This proves (18), as desired.

As we discussed above, this gives a bound in the direction of (16) but falls short of our desired bound of $(C/L^2)^n$. For this, we instead study the second moment,

$$\mathbb{E}_v \left[\mathbb{P}_M \left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n} \right) \right]^2. \tag{25}$$

Here, we decompose this quantity analogously to the above, to show that (25) is bounded above by a quantity of the form

$$\mathbb{E}_{v \in \Lambda_\varepsilon} \mathbb{E}_{H_1} \mathbb{P}_{H_2} \left(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}, \|H_2 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}, \text{ and } \|H_3^T v_{[d+1,n]}\|_2 \leqslant 2\varepsilon\sqrt{n} \right),$$

where H_1, H_2 are independent copies of H and $H_3 := [H_1, H_2]$ is the concatenation of these two matrices. We then proceed in much the same way as above, treating H_3 in place of H. We shall also require a more complicated form of our Littlewood-Offord theorem, where we allow two "hard" constrains corresponding to the first two events in (25). Ultimately, we arrive at the bound

$$\mathbb{E}_{v \in \Lambda_\varepsilon} \left[\mathbb{P}_M \left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n} \right) \right]^2 \leqslant (C\varepsilon)^{2n},$$

which implies the desired conclusion at (16).

2.4. **Outline of the paper.** In the next section we formally introduce the central definitions and notions that will be used throughout this paper. The remainder of the paper is then roughly divided into three parts. The first part consists of Sections 4–7. Sections 4–6 are dedicated to proving our conditioned inverse Littlewood-Offord result, Lemma 4.1, which is the "real" version of Theorem 1.3. This theorem is properly introduced in Section 4 where we go on to set up the problem on the Fourier side. In Section 5, we establish the key geometric results

we need for navigating the Fourier side of the problem, before completing the proof of Lemma 4.1 in Section 6.

In Section 7, the final section of this first part, we set ourselves up for the next part of the paper by using Lemma 4.1 to establish the crucial inequality described at (24), the formal statement of which takes the form of Theorem 7.1. Theorem 7.1 is the only result we carry forward into later sections.

The second part of the paper consists of Sections 8 and 9. In Section 8, we obtain our crucial bound on the size of our net $\mathcal{N}_\varepsilon$ by carrying out our "inversion of randomness" scheme, as outlined in Section 2.3. Section 9 contains the less exciting proof that $\mathcal{N}_\varepsilon$ is in fact a net for Σ_ε.

In the final section, Section 10, we pull together the various elements of this proof, state the reductions we use from previous work and complete the proof of Theorem 1.1.

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

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

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

where $\rho(v)$ is as defined at (5). 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 10 for the details.

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 $D \subseteq [n]$, $|D| = d$, define

(26)
$$\mathcal{I}(D) := \left\{ v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D \right\},$$

where $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants, fixed throughout the paper and defined in Section 3.1. We will set $d := c_0^2 n/2$, where c_0 is defined below in Section 3.1. Now set

$$\mathcal{I} := \bigcup_D \mathcal{I}(D),$$

where the union is over all $D \subseteq [n]$, $|D| = d$. The case of non-flat v is already taken care of in the work of Vershynin [52] (see Section 10) 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 $\mu/2$).

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

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

in the style of [51] (where it is called the *threshold*). Note we are defining $\mathcal{T}_L$ *relative to the matrix* M, rather than our original distribution A. Now define our partition of $\mathcal{I}([d]) \cap \Sigma$. For $\varepsilon \in (0,1)$, let

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

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

$$\Sigma \cap \mathcal{I}([d]) \subseteq \bigcup_{\varepsilon > \gamma^4/(2^{12}L)} \Sigma_\varepsilon.$$

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[4]

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

which is a natural net for $\mathcal{I}([d])$. Here $\mathcal{I}'(D)$ is similar to $\mathcal{I}(D)$ but with slightly looser constraints:

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

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 bounded operator norm. 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 and define

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

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 10.5), however some care is needed as we are mostly concerned with far less likely events.

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

$$\mathcal{N}_\varepsilon := \left\{v \in \Lambda_\varepsilon : \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n\right\}. \tag{27}$$

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

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

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 $|\mathcal{N}_\varepsilon|/|\Lambda_\varepsilon|$ is smaller than $\approx L^{-2n}$, where L is a large constant.

3.1. **Discussion of constants and parameters.** We will treat the constants κ_0, κ_1 (seen at (26)) as absolute throughout the paper, and we allow other absolute constants C, C', ... to depend on these exact quantities. In particular, we set $\kappa_0 = \rho/3$ and $\kappa_1 = \delta^{-1/2} + \rho/6$, where δ, ρ are as in Lemma 10.2 (which is a lemma from [52]). 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 $c_0 = \min\{2^{-24}, \rho\delta^{1/2}/2\}$ where, again, δ, ρ are as in Lemma 10.2. 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 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.

4. 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 Littlewood-Offord type theorem, Lemma 4.1, which works in the presence of a large number ($k \approx n$) of relatively soft constraints on our random walk. As we will see, the proof of Lemma 4.1 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 1.3) and a vector $Y \in \mathbb{R}^d$, we define the Y-augmented matrix W_Y as

$$W_Y = \left[W, \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix}, \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} \right]. \tag{28}$$

Here $Y \approx v/t$ will be a re-scaled version of v from Theorem 1.3. We define, for $\alpha \in (0,1)$, the *least common denominator* of a vector $v \in \mathbb{R}^d$ to be

$$D_\alpha(v) := \inf\left\{ \phi > 0 : \ \|\phi \cdot v\|_{\mathbb{T}} \leqslant \min\left\{ \phi\|v\|_2/2, \sqrt{\alpha d} \right\} \right\}, \tag{29}$$

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 (29) 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 Littlewood-Offord type theorem, which is our "real" (and strengthened) version of Theorem 1.3, from Section 1.1.

Lemma 4.1. *For $d\in\mathbb{N}$ and α, $\mu\in(0,1]$, let $0\leqslant k\leqslant 2^{-10}\alpha d$ and $t\geqslant\exp(-2^{-8}\mu\alpha d)$. For $0<c_0\leqslant 2^{-22}\mu$, 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 $\tau\sim\mathcal{Q}(2d,\mu)$ and $D_\alpha(Y)>16$ then

$$\mathcal{L}\left(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\right)\leqslant (Rt)^2\exp(-c_0k), \tag{30}$$

where $R=2^{33}c_0^{-2}\mu^{-1/2}$.

Before we start working towards the proof of Lemma 4.1, we make a few informal remarks on its statement and its connection to Theorem 1.3. The main difference to note is that there are now two "hard" constraints encoded in the left-hand side of (30); these are, in the notation of Theorem 1.3,

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

The "soft" constraints are, as above, encoded as the columns $w_1,\ldots,w_k$ 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 4.1, in contrast with the statement of Theorem 1.3, 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 4.1 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 $c>0$.

As a warm-up for the reader, we show how Lemma 4.1 easily implies Theorem 1.3.

Proof of Theorem 1.3. Let $\alpha,t\in(0,1)$, $v\in\mathbb{S}^{d-1}$ with $D_\alpha(v)>16/t$ and W be a $k\times d$ matrix with orthonormal rows. Let $Y=(2^{-23}/t)v$ and note we have $\|Y\|_2=2^{-22}t^{-1}$ and $D_\alpha(Y)>16$. Now let $X,X'\sim\{-1,1\}^d$ be iid uniform random variables and let $\tau=(X,X')$. We bound the square of quantity at (6) above by

$$\begin{aligned}\mathbb{P}\left(|\langle Y,X\rangle|\leqslant c_0/2,\ \|WX\|\leqslant\sqrt{c_0k/2}\right)^2 &\leqslant \mathbb{P}\big(\langle Y,X\rangle^2+\langle Y,X'\rangle^2+\|W\tau\|_2^2\\ &\leqslant c_0(k+1)\big).\end{aligned}$$

We now define W' to be the $k\times 2d$ matrix formed by of concatenating two copies of W. We note that $\|W'\|=\sqrt{2}$ and $\|W'\|_{\mathrm{HS}}=\sqrt{2k}$. We then easily see that

$$\mathbb{P}\left(\langle Y,X\rangle^2+\langle Y,X'\rangle^2+\|W\tau\|_2^2\leqslant c_0(k+1)\right)\leqslant\mathcal{L}\left(W'_Y\tau, c_0^{1/2}\sqrt{k+1}\right).$$

We now apply Lemma 4.1 with $t' = t + \exp(2^{-8}\alpha d)$, $\mu = 1$ and $c_0 = 2^{-22}$ to see

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

Now using that $Y = (c_0/2t)v$ and letting $C = R = 2^{77}$ and $c = c_0/2 = 2^{-23}$ we obtain

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant t \text{ and } \|WX\|_2 \leqslant c\sqrt{k}\right) \leqslant Cte^{-ck} + 2e^{-c\alpha d},$$

as desired. □

For the remainder of this section, we take some first steps towards the proof of Lemma 4.1. 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 5, we make a more significant reduction about the geometry of our level set. In Section 6 we prove the key Lemma 6.1, the statement of which is very similar to that of Lemma 4.1, but with a more complicated quantity replacing the right-hand side of (30). Finally, with one further step, we conclude Section 6, with the proof of Lemma 4.1.

4.1. **Passing to the Fourier side.** To prove Lemma 4.1 we will prove the contrapositive; assume (30) 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 4.1 is to use the lower bound in the negation of (30) to obtain a lower bound on a level set of an appropriate Fourier transform. This manoeuvre was pioneered by Halász [16] 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 (30) with the Gaussian volume of a level set.

Lemma 4.2. *Let $\beta > 0$, $\nu \in (0, 1]$, 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 4.2 along with the negation of (30), 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*

$$\Gamma_{r,s} := \left\{\theta \in \mathbb{R}^{k+2} : \ \left\|\theta_{[k]}\right\|_2 \leqslant r, |\theta_{k+1}| \leqslant s \text{ and } |\theta_{k+2}| \leqslant s\right\}. \tag{31}$$

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

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S_{W_Y}(m) \neq \emptyset, \tag{32}$$

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 6.4–6.6). 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 (32) fails.

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

Fact 4.3. 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. First note that if $8s^2 > 1$ then the statement holds trivially and so we may assume $8s^2 \leqslant 1$. 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

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

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

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

where we have used that $(x_i - s)^2 \geqslant p_i^2/8$, which holds since $s < 1$. 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"

$$S(\theta_{[k]}) := \left\{(\theta_{k+1}, \theta_{k+2}) \in \mathbb{R}^2 : (\theta_{[k]}, \theta_{k+1}, \theta_{k+2}) \in S\right\}. \tag{33}$$

Lemma 4.4 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 4.4. *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 4.3) 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 4.1.

5. Inverse Littlewood-Offord II: A geometric inequality

We now turn to make a more intricate and subtle reduction from that seen in Section 4.2, 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 4.2.

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 6) a little easier to spot. Given sets $A, B \subseteq \mathbb{R}^k$, we let $A - B = \{a - b : a \in A, b \in B\}$ and define $A + B$ similarly.

Lemma 5.1. *For $k \in \mathbb{N}$ and $s > 0$, let $S \subset \mathbb{R}^{k+2}$ be a measurable set which satisfies*

$$8s^2 e^{-k/8} + 64s^2 \max_{a,b,y} \left(\gamma_k(F_y(S; a, b) - F_y(S; a, b))\right)^{1/4} < \gamma_{k+2}(S). \tag{34}$$

Then there is an $x \in S$ so that[5]

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S \neq \emptyset. \tag{35}$$

To prove this lemma, we will need a few facts about Gaussian space, which we collect in Sections 5.1 and 5.2, before moving on to prove Lemma 5.1 in Section 5.3.

5.1. 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 5.2. *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 $B := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leqslant r\}$.

[5]Note, in particular, that Lemma 5.1 says that *if* (34) is satisfied then we must have $s < 16$.

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

$$\gamma_{k+2}(S\cap B) = \sum_{y\in\{0\}^k\times 16\mathbb{Z}^2} \gamma_{k+2}(S\cap(\Gamma_{r,8}+y)). \tag{36}$$

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

$$\int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\theta\in S\cap(\Gamma_{r,8}+y)\big] e^{-\pi\|\theta\|_2^2/2}\,d\theta = \int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\phi\in(S-y)\cap\Gamma_{r,8}\big] e^{-\pi\|\phi+y\|_2^2/2}\,d\phi. \tag{37}$$

Rewriting the exponent in the integrand at (37)

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

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

$$\begin{aligned}\gamma_{k+2}(S\cap(\Gamma_{r,8}+y)) &\\ \leqslant \exp\Big(-\frac{\pi}{2}y_{k+1}^2 - \frac{\pi}{2}y_{k+2}^2 + 8\pi|y_{k+1}| + 8\pi|y_{k+2}|\Big)&\,\gamma_{k+2}((S-y)\cap\Gamma_{r,8}).\end{aligned} \tag{38}$$

So, apply (38) to (36) 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 5.3. $\gamma_k\big(\{x\in\mathbb{R}^k : \|x\|_2^2\geqslant k\}\big)\leqslant\exp(-k/8)$.

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. □

5.2. **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 5.4 (Borell)**.** *Let $A, B \subseteq \mathbb{R}^k$ be Borel sets. Then*

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

Proof. In [4] Theorem 5.4 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. □

We will use the following simple consequence of Theorem 5.4.

Lemma 5.5. *Let $A \subseteq \mathbb{R}^k$ be Borel sets. Then*

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

Proof. By Theorem 5.4, we have

$$\gamma_k(A-A) \geqslant \Phi(2\Phi^{-1}(\gamma_k(A))) = \Phi(2x), \tag{39}$$

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

$$\Phi(2x) = \mathbb{P}(Z \leqslant 2x) = \mathbb{P}\left(Z_1+Z_2+Z_3+Z_4 \leqslant 4x\right) \geqslant \mathbb{P}(Z \leqslant x)^4 = \Phi(x)^4, \tag{40}$$

where Z_j are i.i.d. copies of $Z \sim \mathcal{N}(0,1)$. Combining (39) and (40) completes the proof. □

5.3. **Proof of Lemma 5.1.** With these pieces now in place, we can move on to prove Lemma 5.1, our key geometric lemma on the Fourier side.

Proof of Lemma 5.1. Write $r = \sqrt{k}$ for simplicity. We prove the contrapositive and assume for every $x \in S$ we have

$$(\Gamma_{2r,16} \setminus \Gamma_{2r,s} + x) \cap S = \emptyset. \tag{41}$$

We recall that

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

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 (33):

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

We may write

$$\gamma_{k+2}(S \setminus B) = \int_{\|\theta_{[k]}\|_2 \geqslant r} \gamma_2\left(S(\theta_{[k]})\right)\, d\gamma_k \tag{42}$$

and thus

$$\gamma_{k+2}(S \setminus B) \leqslant \left(\max_{\theta_{[k]} \in \mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right)\right) \gamma_k\big(\{\|\theta_{[k]}\|_2 \geqslant r\}\big). \tag{43}$$

Lemma 4.4 and (41) show that

$$\max_{\theta_{[k]}\in\mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right) \leqslant 8s^2. \tag{44}$$

Fact 5.3 bounds

$$\gamma_k\big(\{\|\theta_{[k]}\|_2 \geqslant r\}\big) \leqslant \exp(-k/8) \tag{45}$$

and so from (43), (44) and (45) we learn

$$\gamma_{k+2}(S \setminus B) \leqslant 8s^2 e^{-k/8}. \tag{46}$$

Step 2 (Upper bound for $\gamma_{k+2}(S \cap B)$). By Lemma 5.2, there exists $x \in S$ and $h \in \Gamma_{r,8}$ such that

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-x) \cap \Gamma_{2r,16} + h). \tag{47}$$

Now since $x \in S$, we use (41) to deduce that

$$(S-x) \cap \Gamma_{2r,16} \subseteq (S-x) \cap \Gamma_{2r,s} \tag{48}$$

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

$$(S-x) \cap \Gamma_{2r,s} + h = (S-x+h) \cap (\Gamma_{2r,s} + h) = (S-y) \cap (\Gamma_{2r,s} + h). \tag{49}$$

Thus by (47), (48) and (49), we have

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)). \tag{50}$$

Bound

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant \int_{|a-h_{k+1}|,|b-h_{k+2}|\leqslant s} \gamma_k\big(F_y(S;a,b)\big)\, d\gamma_2 \tag{51}$$

and apply Lemma 5.5 to obtain

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant 4s^2 \max_{a,b,y}(\gamma_k(F_y(S;a,b) - F_y(S;a,b)))^{1/4}. \tag{52}$$

Combining (50) and (52) gives

$$\gamma_{k+2}(S \cap B) \leqslant 64s^2 \max_{a,b,y}\left(\gamma_k\left(F_y(S;a,b) - F_y(S;a,b)\right)\right)^{1/4}. \tag{53}$$

Putting Step 1 and Step 2 together: (53) together with (46) 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. □

6. Inverse Littlewood-Offord III: Comparison to a lazier walk and the proof of Lemma 4.1

In Section 5 we proved our key geometric ingredient, Lemma 5.1, to deal with the geometry of our level set (as seen in Section 4.1). We now use this lemma to take the following big step towards Lemma 4.1.

Lemma 6.1. *For $d \in \mathbb{N}$ and $\alpha, \mu \in (0,1]$, let $0 \leqslant k \leqslant 2^{-8}\alpha d$ and $t \geqslant \exp(-2^{-8}\mu\alpha d)$. For $0 < c_0 \leqslant 2^{-22}\mu$, 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 $\tau \sim \mathcal{Q}(2d,\mu)$ and $\tau' \sim \mathcal{Q}(2d, 2^{-7}\mu)$ and $\beta \in [c_0/2^{10}, \sqrt{c_0}]$, $\beta' \in (0, \sqrt{c_0})$.*

If

(54)
$$\mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}) \geqslant (Rt)^2 \exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4}$$

then $D_\alpha(Y) \leqslant 16$. *Here we have set* $R = 2^{32}c_0^{-2}\mu^{-1/2}$.

Of course, Lemma 6.1 looks quite a bit like Lemma 4.1 save for quantity

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k), \tag{55}$$

on the right-hand side of (54). 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 [9], which is used to deal with very unstructured vectors.

We prove this lemma in Section 6.2 after laying out a few facts on level sets in Section 6.1. We will then conclude this section in Section 6.3 with a proof of Lemma 4.1, by combining Lemma 6.1 with one further ingredient to bound (55).

6.1. **Working with level sets.** To prepare for the proof of Lemma 6.1, 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 4, Lemma 4.2. 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 6.2. *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 for all $t \geqslant 0$, we have*

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

We remark that we impose laziness $\nu \in (0, 1/4]$ here to make the characteristic function of $W^T\tau$ non-negative.

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",

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

Fact 6.3. 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

$$F_y(S_W(m); a, b) - F_y(S_W(m); a, b) \subseteq F_0(S_W(4m); 0, 0). \tag{56}$$

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 (56), 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 (56). □

6.2. **Proof of Lemma 6.1.** We may now turn to proving Lemma 6.1, our big step towards Lemma 4.1.

Proof of Lemma 6.1. Apply Lemma 4.2, with parameter μ, 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

$$e^{-\frac{\mu m}{2} + 2\beta^2 k} \gamma_{k+2}(S) \geqslant \mathcal{L}(W_Y^T \tau, \beta\sqrt{k+1}). \tag{57}$$

Thus (57) together with our hypothesis (54) gives a lower bound

$$\gamma_{k+2}(S) \geqslant \frac{1}{4} e^{\frac{\mu m}{2} + 2\beta^2 k} (Rt)^2 T^{1/4}, \tag{58}$$

where we have set

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

where we recall that $\tau' \sim \mathcal{Q}(2d, 2^{-7}\mu)$. We now make the following important designations,

$$r_0 := \sqrt{k} \qquad \text{and} \qquad s_0 := 2^{16} c_0^{-1} (\sqrt{m} + \sqrt{k}) t. \tag{59}$$

Recall from (31) 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 6.4. There exists $x \in S \subseteq \mathbb{R}^{k+2}$ so that[6]

$$\left(\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x\right) \cap S \neq \emptyset. \tag{60}$$

Proof of Claim 6.4. We look to apply Lemma 5.1 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 $e^{\mu m} T$, thus giving a lower bound on $\gamma_{k+2}(S)$ and allowing us to apply Lemma 5.1. Use Fact 6.3 to see that for any y, a, b, we have

$$F_y(S; a, b) - F_y(S; a, b) \subseteq F_0(S_{W_Y}(4m); 0, 0). \tag{61}$$

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 $\tau' \sim \mathcal{Q}(2d, 2^{-7}\mu)$ and $\beta' \in (0, 1/2)$ is as in the hypothesis. Thus we may apply Lemma 6.2 (with $\nu = 2^{-7}\mu$ and $t = 4m$) along with (61) to obtain

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

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

We may combine this with the fact that $T \geqslant \exp(-\beta'^2 k) \geqslant e^{-k/4}$, since $\beta' \leqslant 1/2$, to get

$$T^{1/4} \geqslant \frac{1}{2} e^{-\mu m/4}(e^{-k/16} + M^{1/4}). \tag{62}$$

So combining (62) with (58) gives

$$\gamma_{k+2}(S) \geqslant (1/8)e^{\mu m/4 + 2\beta^2 k}(Rt)^2(e^{-k/16} + M^{1/4}) > 64 s_0^2(e^{-k/16} + M^{1/4}), \tag{63}$$

allowing us to apply Lemma 5.1 and complete the proof of the claim. The last inequality at (63) follows from a simple check. First note that

$$s_0^2 = 2^{32} c_0^{-2}(\sqrt{m} + \sqrt{k})^2 t^2 < 2^{33}(k+m)(t/c_0)^2. \tag{64}$$

Now use (64) and the facts that $R = \mu^{-1/2} c_0^{-2} 2^{32}$ and $\beta \geqslant 2^{-10} c_0$ to bound

$$64 s_0^2 \leqslant 2^{39} t^2 c_0^{-2}(2^{20} c_0^{-2}\beta^2 k + 4\mu^{-1}(\mu m/4)) \leqslant \frac{1}{8}(Rt)^2 e^{\mu m/4 + 2\beta^2 k}$$

thus showing the second inequality at (63) and finishing the proof of the claim. □

We now observe the simple consequence of Claim 6.4.

Claim 6.5. We have that $S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0}) \neq \emptyset$.

Proof of Claim 6.5. By Claim 6.4, 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 6.3. □

We now conclude the proof of Lemma 6.1 with Claim 6.6.

Claim 6.6. 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 6.6. 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

$$\psi_{k+1}\begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2}\begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \in B_{2d}(p, 2\sqrt{m}) - W\psi_{[k]} \subseteq B_{2d}(p, 2\sqrt{m} + 4\sqrt{k}), \tag{65}$$

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 (65) onto the first d coordinates, to obtain

$$\psi_{k+1} Y \in B_d(p_{[d]}, 2\sqrt{m} + 4\sqrt{k}). \tag{66}$$

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

$$\frac{\psi_{k+1}\|Y\|_2}{2} > \frac{s_0\|Y\|_2}{2} \geqslant \left(\frac{2^{15}(\sqrt{m} + \sqrt{k})t}{c_0}\right)\left(2^{-10}\frac{c_0}{t}\right) > (2\sqrt{m} + 4\sqrt{k}), \tag{67}$$

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 (58), the bounds $\gamma_{k+2}(S) \leqslant 1$, $T \geqslant e^{-\beta'^2 k}$ and our assumption $t \geqslant \exp(-2^{-8}\mu\alpha d)$ to see that

$$e^{-\mu m/2} \geqslant \gamma_{k+2}(S)e^{-\mu m/2} \geqslant \frac{1}{4}(Rt)^2 e^{2\beta^2 k - \beta'^2 k/4} \geqslant \exp(-2^{-5}\mu\alpha d),$$

where we have used $R^2 \geqslant 4$, $k \leqslant 2^{-7}\alpha d$ and $\beta' \leqslant \sqrt{c_0}$ for the last inequality. It follows that $m \leqslant 2^{-4}\alpha d$ and so by (66) and (67) 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 6.6. □

Let ψ and $i \in \{k+1, k+2\}$ be as guaranteed by Claim 6.6. Then $\psi_i \leqslant 16$, since $\psi \in \Gamma_{2r_0,16}$, 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 6.1. □

6.3. **Proof of Lemma 4.1.** Before turning to prove Lemma 4.1, 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 [13, 16, 36] (and we will encounter another in Subsection 8.2), we state here a variant of the Hanson-Wright inequality with an explicit constant. A proof can be found in Appendix D of [7], the arXiv version of this paper, and is a consequence of a classical concentration inequality due to Talagrand [43].

Lemma 6.7. *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*

$$\mathbb{P}(\|W^T\sigma\|_2 \leqslant \delta\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k). \tag{68}$$

We now turn to prove Lemma 4.1.

Proof of Lemma 4.1. Setting $\beta' := 4\sqrt{c_0}$, we look to apply Lemma 6.1. For this, note that the hypotheses in Lemma 4.1 imply the hypotheses in Lemma 6.1 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 6.1 (in the contrapositive) to obtain

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (2^{32}c_0^{-2}\mu^{-1/2}t/2)^2 e^{4\beta^2 k}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + e^{-\beta'^2 k}\right)^{1/4}. \tag{69}$$

To deal with the right-hand side, we apply Lemma 6.7 to take care of the quantity involving $\tau' \in \{-1,0,1\}^{2d}$, our $\nu = 2^{-7}\mu$ lazy random vector. Note that $4\sqrt{c_0} \leqslant 2^{-9}\sqrt{\mu} \leqslant \sqrt{\nu}/16$, and that our given W satisfies $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Thus we may apply Lemma 6.7, with $\delta = \beta'$ and $\sigma = \tau'$, to see

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k). \tag{70}$$

Plugging this into the right-hand side of (69) yields

$$\begin{aligned}\exp(4\beta^2 k)&\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4}\\ &\leqslant 2\exp(4c_0 k - 2^{-14}\nu k) + 2\exp(2c_0 k - 4c_0 k)\\ &\leqslant 4\exp(-c_0 k).\end{aligned}$$

Putting this together with (69), yields

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

as desired. □

7. Inverse Littlewood-Offord for conditioned random matrices

In this section we lift the main result of the previous sections (Lemma 4.1) 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 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 Theorem 7.1.[7]

Theorem 7.1. *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 $D_\alpha(r_n X) > 16$ then

(71)

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

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 7.1, notice that if we only consider the "soft" constraints on the singular values (without the constraints imposed by X) we would expect something like

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n}\right) \approx c^{nk}, \tag{72}$$

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.[8]

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

$$\mathbb{P}\big(\|H_1X\|_2, \|H_2X\|_2 \leqslant n\big) \geqslant \left(\frac{R}{N}\right)^{2n-2d},$$

then $D_\alpha(n^{-1/2}X) = O(1)$. Thus Theorem 7.1 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.

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

[8] 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 (72) in the literature. However, it is not so hard to deduce (72) from the Hanson-Wright inequality [36] along with a "random rounding" step similar to that in Appendix E in [7].

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

Fact 7.2. 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 7.3. *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

(73)
$$\mathbb{P}\big(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\big) \leqslant \exp\big(2\beta^2(k+1)(n-d)\big)\,\mathbb{E}_H e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2}.$$

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

$$\mathbb{E}_H e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2} = \prod_{i=1}^{n-d}\mathbb{E}_{\tau_i} e^{-2\|W^T\tau_i\|^2/\beta^2} = \Big(\mathbb{E}_\tau e^{-2\|W^T\tau\|^2/\beta^2}\Big)^{n-d}. \tag{74}$$

Observe now that

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

Splitting the integral on the right-hand side gives

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

We then appeal to Fact 7.2 to write

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

Here the first integral is $\leqslant 1$, while the second integral is $\leqslant 8$ and thus

$$\mathbb{E}_\tau e^{-2\|W^T\tau\|^2/\beta^2} \leqslant 9\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big). \tag{75}$$

Combining line (75) with (74) and (73) gives

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

and the result follows. □

7.2. **Approximating matrices W with nets.** Note that in Theorem 7.1, 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 $\mathbb{R}^{[2d]\times[k]}$. A proof can be found in Appendix E of [7], the arXiv version of this paper.

Lemma 7.4. *For $k \leqslant d$ and $\delta \in (0, 1/2)$, there exists $\mathcal{W} = \mathcal{W}_{2d,k} \subset \mathbb{R}^{[2d]\times[k]}$ with $|\mathcal{W}| \leqslant (2^6/\delta)^{2dk}$ so that for any $U \in \mathcal{U}_{2d,k}$, any $r \in \mathbb{N}$ and $r \times 2d$ matrix A there exists $W \in \mathcal{W}$ 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 (28)) the augmented matrix

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

7.3. **Proof of Theorem 7.1.** We recall a standard fact from linear algebra, reworded to suit our context.

Fact 7.5. 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 $w_1, \ldots, w_k \in \mathbb{R}^{2d}$ 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 7.1. Write $Y := \frac{c_0}{16\sqrt{n}} \cdot X$. We use Fact 7.5 to upper bound the left-hand-side of (71) as

$$\begin{aligned}\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n) \\ \leqslant \mathbb{P}(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant 3c_0\sqrt{n(k+1)}/16).\end{aligned}$$

Set $\delta := c_0/16$, and let $\mathcal{W}$ be the net for $\mathcal{U}_{2d,k}$, given by Lemma 7.4.

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 7.4 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

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

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, by Lemma 7.4,

$$|\mathcal{W}| \leqslant (2^6/\delta)^{2dk} \leqslant \exp(32dk\log c_0^{-1}) \leqslant \exp(c_0k(n-d)/4),$$

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

$$\sum_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right) \leqslant e^{c_0k(n-d)/4} \max_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right). \tag{76}$$

Let $W \in \mathcal{W}$ be such that the maximum in (76) is attained, apply Lemma 7.3 with $\beta := \sqrt{c_0}/2$ to obtain

$$\mathbb{P}(\|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}) \leqslant \left(2^5 e^{c_0k/2}\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big)\right)^{n-d}. \tag{77}$$

We now look to apply Lemma 4.1. We define $t := 16/(c_0N) \geqslant \exp(-2^{-9}\alpha d)$ and $R_0 := 2^{-7}c_0R = 2^{-7}c_0(2^{39}c_0^{-3}) = 2^{32}c_0^{-2}$ so that we have

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

By the construction of $\mathcal{W}$ in Lemma 7.4 we have $\|W\| \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$. We also have $k \leqslant 2^{-10}\alpha d$ and $D_\alpha(\frac{c_0}{16\sqrt{n}}X) = D_\alpha(Y) > 16$, therefore we may apply Lemma 4.1 to see that

$$\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big) \leqslant (R_0t)^2\exp(-c_0k) \leqslant \left(\frac{R}{8N}\right)^2\exp(-c_0k).$$

Substituting this bound in (77) 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

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

This completes the proof of Theorem 7.1. □

8. 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

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

In particular, our main goal of this section will be to prove Theorem 8.1 on $|\mathcal{N}_\varepsilon|$.

Theorem 8.1. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-24}$, let $n \geqslant L^{64/c_0^2}$, let $d \in [c_0^2 n/4, c_0^2 n]$ and let $\varepsilon > 0$ be such that $\log \varepsilon^{-1} \leqslant nL^{-32/c_0^2}$. Then*

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

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 [51], by working with the intersection of $\mathcal{N}_\varepsilon$ with a selection of "boxes" which cover (an appropriately re-scaled) Λ_ε.

Definition 8.2. For $d, n, N \in \mathbb{N}$ with $d \leqslant n$ and $\kappa > 1$, 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 8.1, in this probabilistic framework. Indeed, almost all of the work in proving Theorem 8.1 goes into proving the following variant for boxes.

Lemma 8.3. *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.

8.1. **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 7.1, which requires an upper bound on the $D_\alpha(X)$ for application.

We point out that in Lemma 8.4, 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 8.4 when $K = 16$.

Lemma 8.4. *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*

$$\mathbb{P}_X\left(D_\alpha(r_n X) \leqslant K\right) \leqslant (2^{20}\alpha)^{d/4}, \tag{78}$$

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

Proof. If $D_\alpha(r_nX) \leqslant K$ then let $\psi \in (0, K]$ be the minimum[9] in the definition of least common denominator. Set $\phi := r_n\psi$ and observe that ϕ satisfies

$$\|\phi X\|_{\mathbb{T}} \leqslant \sqrt{\alpha d} \quad \text{and} \quad \phi \in [(2\kappa N)^{-1}, r_n K]. \tag{79}$$

To see the bound $\phi \geqslant (2\kappa N)^{-1}$, note that if $\phi < (2\kappa N)^{-1}$ then each coordinate of ϕX 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 (29), this would imply

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

which is a contradiction. Thus the bounds in (79) hold.

Now to calculate the probability in (78), 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 (79) then there exists ϕ_i for which

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

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

$$\|\phi_i X\|_{\mathbb{T}} \leqslant \|\phi X\|_{\mathbb{T}} + \|(\phi_i - \phi)X\|_2 \leqslant \sqrt{\alpha d} + |\phi_i - \phi| \cdot \sqrt{d}(\kappa N) \leqslant 2\sqrt{\alpha d}. \tag{80}$$

Thus we have that

$$\mathbb{P}_X(D_\alpha(r_nX) \leqslant K) \leqslant \sum_{i\in I} \mathbb{P}_X\left(\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d}\right). \tag{81}$$

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

$$\mathbb{P}_X(D_\alpha(r_nX) \leqslant K) \leqslant 2^d \sum_{i\in I} \prod_{j=1}^{d/2} \mathbb{P}_{X_j}\left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}\right). \tag{82}$$

We now claim that

$$\mathbb{P}_{X_j}\left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}\right) \leqslant 32\sqrt{\alpha}. \tag{83}$$

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

$$\begin{aligned}\mathbb{P}_{X_j}\left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}\right) &\leqslant \sum_{p=-T_i}^{T_i} \mathbb{P}_{X_j}\left(|X_j - p\phi_i^{-1}| \leqslant 4\sqrt{\alpha}\phi_i^{-1}\right) \\ &\leqslant \frac{(2T_i+1)(8\alpha^{1/2}\phi_i^{-1}+1)}{2(\kappa-1)N},\end{aligned}$$

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 (80) along with the assumption $\kappa \geqslant 2$. Also note that $8\alpha^{1/2}\phi_i^{-1} \geqslant 1$ since $\phi \leqslant r_nK \leqslant d^{-1/2}K$, allowing us to conclude (83).

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

Now, plugging (83) into (82) and bounding $|I| \leqslant (2KN/\alpha + 1) \leqslant 3^d$ completes the proof of Lemma 8.4. □

8.2. **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 8.5 will actually be H^T in Section 8.3.

Lemma 8.5. *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*

$$\mathbb{P}_X\big(\|HX\|_2 \leqslant n\big) \leqslant \left(\frac{Cn}{dc_0N}\right)^{2d-k},$$

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 8.6. *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 8.5.

Proof of Lemma 8.5. Since H^TH is a symmetric $(n-d) \times (n-d)$ matrix with $\mathrm{rk}(H) \leqslant 2d$, by the spectral theorem we have $H^TH = \sum_{i=1}^{2d} \sigma_i(H)^2 v_i v_i^T$, 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

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

Therefore

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \mathbb{P}_X(\|PX\|_2 \leqslant 16c_0^{-1}\sqrt{n}). \tag{84}$$

We now apply Theorem 8.6 to the orthogonal projection P, with $K = 16c_0^{-1}\sqrt{n/(2d-k)}$,

$$\mathbb{P}_X(\|PX\|_2 \leqslant K\sqrt{2d-k}) \leqslant \left(\frac{Cn}{dc_0N}\right)^{2d-k}, \tag{85}$$

which together with (84) completes the proof of Lemma 8.5. □

8.3. **Proof of Lemma 8.3.** We take a moment to prepare the ground for the proof of Lemma 8.3. We express our random matrix M, as in the statement of Lemma 8.3, 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

$$\mathcal{A}_1 := \left\{H : \|H_1X_{[d]}\|_2 \leqslant n \text{ and } \|H_2X_{[d]}\|_2 \leqslant n\right\}$$

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

$$\mathcal{A}_2 := \left\{H : \|H^TX_{[d+1,n]}\|_2 \leqslant 2n\right\}.$$

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

Fact 8.7. 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

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

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

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

which is at most

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

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

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

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

$$\alpha := 2^{13}L^{-8n/d}, \tag{86}$$

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

$$T = T(\mathcal{B}) := \left\{X \in \mathcal{B} : D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) > 16\right\}. \tag{87}$$

Now set $K := 16$ and note that Lemma 8.4 implies that if X is chosen uniformly from $\mathcal{B}$ and $n \geqslant L^{64/c_0^2} \geqslant 2^8/\alpha$ we have

$$\mathbb{P}_X(X \notin T) = \mathbb{P}_X\big(D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) \leqslant 16\big) \leqslant \left(2^{33} L^{-8n/d}\right)^{d/4} \leqslant \left(\frac{2}{L}\right)^{2n}. \tag{88}$$

Proof of Lemma 8.3. 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

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

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

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X\left(f(X) \geqslant (L/N)^n\right) + (2/L)^{2n} \leqslant (N/L)^{2n}\mathbb{E}_X f(X)^2 + (2/L)^{2n}, \tag{89}$$

where the last inequality follows from Markov's inequality. So to prove Lemma 8.3, it is enough to prove $\mathbb{E}_X f(X)^2 \leqslant 2(R/N)^{2n}$.

From Fact 8.7 we may write

$$\mathbb{P}_M(\|MX\|_2 \leqslant n)^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2) = \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \tag{90}$$

and so

$$f(X)^2 \leqslant \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{1}(X \in T). \tag{91}$$

We now look to apply Lemma 7.1 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 7.1). 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 7.1,

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

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

$$f(X)^2 \leqslant \sum_{k=0}^{\alpha' d} \mathbb{P}_H(\mathcal{A}_2 \mid \mathcal{A}_1 \cap \mathcal{E}_k) \exp(-c_0 nk/4)\left(\frac{R_0}{N}\right)^{2n-2d} + \exp(-c_0 \alpha' dn/4). \tag{92}$$

We now consider the quantities $g_k(X) := \mathbb{P}_H(\mathcal{A}_2 \mid \mathcal{A}_1 \cap \mathcal{E}_k)$ appearing in (92). Indeed,

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

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 8.5 to see that

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

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

$$\mathbb{E}_X[g_k(X)] \leqslant \left(\frac{R}{2N}\right)^{2d-k}. \tag{93}$$

We apply $\mathbb{E}_X$ to (92) and then use (93) 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

$$\mathbb{E}_X f(X)^2 \leqslant 2\left(\frac{R}{2N}\right)^{2n}. \tag{94}$$

Combining (94) with (89) completes the proof of Lemma 8.3. □

8.4. **Proof of Theorem 8.1.** The main work of this section is now complete with the proof of Lemma 8.3. 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

$$\mathcal{I}'([d]) := \left\{v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in [d]\right\},$$
$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2}\mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

and that $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants defined in Section 3.

Lemma 8.8. *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*

$$\Lambda_\varepsilon \subseteq \bigcup_{\mathcal{B}\in\mathcal{F}} (4\varepsilon n^{-1/2})\mathcal{B}, \tag{95}$$

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 $(\ell_{d+1},\dots,\ell_n) \in \mathbb{Z}^n_{\geqslant 0}$ we define the box $B(\ell_{d+1},\dots,\ell_n) := J^d \times \prod_{j=d+1}^n I_{\ell_j}$ and the family of boxes

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

We claim that $\mathcal{F}$ is the desired family. For this, we first show the inclusion at (95). 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 $X \in B(\ell_{d+1},\dots,\ell_n)$. 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 $X \in B(\ell_{d+1}, \dots, \ell_n)$. We now observe that $B(\ell_{d+1}, \dots, \ell_n) \in \mathcal{F}$, 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 (95).

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

$$\sum_{j=0}^{8n/(\kappa_0^2 4^t)} \binom{n}{j} \leqslant \left(\frac{e\kappa_0^2 4^t}{8}\right)^{8n/(\kappa_0^2 4^t)} \leqslant e^{8n/(\kappa_0^2 2^t)}$$

choices for these values of i if $8/(\kappa_0^2 2^t) \leqslant 1$ and at most 2^n choices otherwise. Hence, there are at most

$$2^{n\log_2(8/\kappa_0^2)} \cdot \prod_{t \geqslant \log_2(8/\kappa_0^2)} e^{8n/(\kappa_0^2 2^t)} \leqslant (8/\kappa_0^2)^n \cdot e^{2n} < \kappa^n$$

such sequences $(\ell_{d+1}, \dots, \ell_n)$.

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 8.8 to apply Lemma 8.3 to deduce Theorem 8.1, the main result of this section.

Proof of Theorem 8.1. Apply Lemma 8.8 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

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

and so

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

By rescaling by $\sqrt{n}/(4\varepsilon)$ and applying Lemma 8.3, we have

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

Here the application of Lemma 8.3 is justified as $0 < c_0 \leqslant 2^{-24}$, $c_0^2 n/2 \leqslant d \leqslant c_0^2 n$; $\kappa \geqslant 2$; we have $\log 1/\varepsilon \leqslant n/L^{32/c_0^2}$ and therefore

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

as specified in Lemma 8.3, 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

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

where $C = \kappa^2 R^2 c_0^6$, thus completing the proof of Theorem 8.1. □

9. 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 9.2, which, in a similar spirit to Lemma 7.4, is based on randomized rounding. For this, we recall that we defined

$$\Sigma_\varepsilon = \{v \in \mathcal{I}([d]) : \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon]\} \subset \mathbb{S}^{n-1}, \tag{96}$$

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

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

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. Lemma 9.1 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. We note that the proof makes use of the fact that for fixed $v \in \mathbb{R}^n$, the characteristic function of Mv is non-negative since the entries of M are sufficiently lazy. This is similar to the replacement step in the work of Kahn Komlós and Szemerédi [19] and subsequent works [5, 46]. However, here we only need to "break even", whereas they are looking for a substantial gain at this step.

Lemma 9.1. *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 9.2 which tells us that $\mathcal{N}_\varepsilon$ is a net for Σ_ε.

Lemma 9.2. *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

$$\mathbb{P}(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n. \tag{97}$$

We first show that *all* u satisfy the upper bound at (97). 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 9.1 bounds

$$\mathcal{L}(Av, 5\varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n. \tag{98}$$

We now show that

$$\mathbb{E}_u \mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (1/2)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \geqslant (1/2)(2\varepsilon L)^n, \tag{99}$$

where the last inequality holds by the fact $v \in \Sigma_\varepsilon$. From (99), it follows that there exists $u \in \Lambda_\varepsilon$ satisfying (97).

So to prove the first inequality in (97), 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

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

Taking expectations with respect to u gives

$$\mathbb{E}_u \mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \big(1 - \mathbb{E}_u \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n} \mid \mathcal{E})\big)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \tag{100}$$

and exchanging the expectations reveals that it is enough to show

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

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

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

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

$$\mathbb{P}_r(\|Mr\|_2 \geqslant 2\varepsilon\sqrt{n}) \leqslant (2\varepsilon\sqrt{n})^{-2}\mathbb{E}_r\|Mr\|_2^2 \leqslant 1/4. \tag{101}$$

Putting (101) together with (100) proves (99), completing the proof of (97). □

10. 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

$$q_n(\gamma) := \max_{w \in \mathbb{R}^n} \mathbb{P}_A(\exists v \in \mathbb{R}^n \setminus \{0\} : Av = w,\ \rho(v) \geqslant \gamma), \tag{102}$$

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 (102) 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 10.1. *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 of the arXiv version of this paper [7], although an almost identical lemma can be found in [8], which collected elements from [9, 11, 31].

10.1. **Non-flat vectors.** Here we note a lemma due to Vershynin [52] 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 [52, Proposition 4.2], Vershynin provides Lemma 10.2 which allows us to disregard all compressible vectors.

Lemma 10.2. *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.

Lemma 10.3 of Rudelson and Vershynin [33, Lemma 3.4] tells us that incompressible vectors are "flat" for a constant proportion of coordinates.

Lemma 10.3. *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 10.2. We also fix $c_0 = \min\{2^{-24}, \rho\delta^{1/2}/2\}$.

Lemma 10.4 is what we will apply in the proof of Theorem 1.1.

Lemma 10.4. *For $n \in \mathbb{N}$, let $d < c_0^2 n$. Then*

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

Proof. Apply Lemma 10.3 along with the definitions of κ_1, κ_2 and $\mathcal{I}$ to see $\mathbb{S}^{n-1} \setminus \mathcal{I} \subseteq \mathrm{Comp}(\delta, \rho)$. Clearly we may assume that $\|w\|_2 = 1$ or $w = 0$. Now take a $c\sqrt{n}$-net $\mathcal{X}$ for $\{tw\}_{0 < t \leqslant 4\sqrt{n}}$ of size at most $8c^{-1}$. Then

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

Union bounding over $\mathcal{X}$ and applying Lemma 10.2 completes the lemma. □

10.2. **Proof of Theorem 1.1.** As we noted in Section 3, 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 10.5. *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 [44, 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 10.5 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 10.6. *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 3, 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 8.1. By possibly decreasing c we may also assume that it is at most half the constant from Lemma 10.4 (which we note depends only on c_0). We also let $c_0 > 0$ be as defined above and $d := \lceil c_0^2 n/2 \rceil$.

From Lemma 10.1 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 3, and note that

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

Since $d = \lceil c_0^2 n/2 \rceil$, by Lemma 10.4 and Lemma 10.5, we have

(103)
$$q_n(\gamma) \leqslant \max_{w \in \mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I} \cap \Sigma :\ Av \in \{tw\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) + 64c^{-1}e^{-2cn}$$

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 (103) is

$$(104) \quad \leqslant 2^n \max_{D \in [n]^{(d)}} \max_{w \in \mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}(D) \cap \Sigma :\ Av \in \{tw\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right)$$

$$(105) \quad = 2^n \max_{w \in \mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}([d]) \cap \Sigma :\ Av \in \{tw\}_{t>0}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right),$$

where the last line holds by symmetry of the coordinates. Thus it is enough to show that the maximum at (105) is at most 4^{-n}.

Now, for $v \in \Sigma$ we have $\rho(v) \geqslant \gamma$ and so, by Lemma 10.6, 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 (96) of $\Sigma_\varepsilon = \Sigma_\varepsilon([d])$ from Section 3, 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

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

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

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

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

$$Q_\varepsilon := \max_{w \in \mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \Sigma_\varepsilon : \ \|Av - w\|_2 \leqslant \varepsilon\sqrt{n}\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right) \leqslant 2^{-4n}. \tag{106}$$

Let $w \in \mathbb{R}^n$ be such that the maximum at (106) is attained. Now, since $\varepsilon < \kappa_0/8$ for $v \in \Sigma_\varepsilon$, we apply Lemma 9.2, 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

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

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

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

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 7.2 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 8.1 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 8.1 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. □

Appendix A. The Proofs of two Esseen-type lemmas

In this section we prove our two Esseen-type lemmas, Lemma 4.2 and Lemma 6.2, 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

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

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

$$-\log\left(1-\mu+\mu\cos(2\pi x)\right) \leqslant 32\mu\|x\|_{\mathbb{T}}^2, \tag{107}$$

and for $\mu \in [0,1]$

$$-\log|1-\mu+\mu\cos(2\pi x)| \geqslant \mu\|x\|_{\mathbb{T}}^2 \tag{108}$$

from which we deduce that for $\mu \in [0,1/4]$

$$\varphi(\theta) \geqslant \exp\left(-32\mu\left\|W\theta\right\|_{\mathbb{T}}^2\right), \tag{109}$$

and for $\mu \in [0,1]$

$$|\varphi(\theta)| \leqslant \exp\left(-\mu\left\|W\theta\right\|_{\mathbb{T}}^2\right). \tag{110}$$

We now note a standard fact regarding Fourier inversion (see [49] 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 4.2 and Lemma 6.2.** 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 4.2.

Proof of Lemma 4.2. Let $w \in \mathbb{R}^\ell$. We apply Markov's inequality to obtain

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

As above, let φ be the characteristic function of $W^T\tau$. We apply Fact A.1 and (110) to obtain

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

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

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

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

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

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 6.2 proceeds in much the same way.

Proof of Lemma 6.2. Let us set $X = \|W^T\tau\|_2$ and write

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

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

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

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 (109) we obtain

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

Similar to the proof of Lemma 4.2, we write

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

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 6.2. □

Appendix B. Relating A to the zeroed out matrix M

In this section we prove Lemma 9.1 and Lemma 10.6. 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

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

and

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

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

$$\psi_v(\xi)\leqslant\chi_v(2\xi), \tag{111}$$

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

$$\mathbb{P}(\|Mv\|_2\leqslant t\sqrt{n})+\sqrt{n}\int_t^\infty\exp\left(-\frac{s^2n}{t^2}\right)\mathbb{P}(\|Mv\|_2\leqslant s\sqrt{n})\,ds, \tag{112}$$

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

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

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

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

as desired. □

Proof of Lemma 9.1. Apply Markov's inequality to bound

$$\mathbb{P}(\|Av-w\|_2\leqslant t\sqrt{n})\leqslant\exp(\pi n/2)\mathbb{E}\exp\left(-\pi\|Av-w\|_2^2/2t^2\right). \tag{113}$$

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

$$\mathbb{E}_A\exp\left(-\pi\|Av-w\|_2^2/2t^2\right)=\int_{\mathbb{R}^n}e^{-\pi\|\xi\|_2^2}\cdot e^{-2\pi it^{-1}\langle w,\xi\rangle}\psi_v(t^{-1}\xi)\,d\xi. \tag{114}$$

Rescaling, applying (111) and non-negativity of χ_v yields that the RHS of (114) 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 10.6 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 10.6. 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

(115)
$$\rho_\varepsilon(v)^n \leqslant \max_{w\in\mathbb{R}^n} \mathbb{P}(\|Bv-w\|_2 \leqslant \varepsilon\sqrt{n}) \leqslant \max_{w\in\mathbb{R}^n} \exp(\pi n/2)\mathbb{E}\exp\left(-\pi\|Bv-w\|_2^2/2\varepsilon^2\right).$$

Apply Fact A.1 to write

(116)
$$\mathbb{E}\exp\left(-\pi\|Bv-w\|_2^2/2\varepsilon^2\right) = \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\cdot e^{-2\pi i\varepsilon^{-1}\langle w,\xi\rangle}\prod_{1\leqslant j,k\leqslant n}\cos(2\pi\varepsilon^{-1}v_j\xi_k)\,d\xi$$

and use Hölder's inequality to bound the RHS of (116)

$$(117)\quad \leqslant \left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2/3}\,d\xi\right)^{3/4}\left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2}\prod_{1\leqslant j,k\leqslant n}\cos(2\pi\varepsilon^{-1}v_j\xi_k)^4\,d\xi\right)^{1/4}.$$

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 (117) is

$$(118)\qquad \leqslant \left(\frac{3}{2}\right)^{3n/8}\left(2^{-n/2}\int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\chi_v(\sqrt{2}\varepsilon^{-1}\xi)\,d\xi\right)^{1/4}$$
$$\leqslant \left(\frac{27}{128}\right)^{n/8}\left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4}.$$

Taken together, lines (115), (116), (117), (118) tell us that

$$(119)\qquad \rho_\varepsilon(v)^n \leqslant (3/2)^{3n/8}(\exp(\pi/2)/\sqrt{2})^n\left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4}.$$

Now apply Fact B.1 to bound $\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right) \leqslant (9L\varepsilon)^n$ and so $\rho_\varepsilon(v)^n \leqslant (2^{12}L\varepsilon)^{n/4}$, as desired. □

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. We would also like to thank the anonymous referees for their careful reading of this paper and for their helpful comments.

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*, Ann. Probab. **16** (1988), no. 4, 1729–1741. MR958213

[2] József Balogh, Robert Morris, and Wojciech Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709, DOI 10.1090/S0894-0347-2014-00816-X. MR3327533

[3] Béla Bollobás, *Random graphs*, Springer, 1998.

[4] Christer Borell, *Inequalities of the Brunn-Minkowski type for Gaussian measures*, Probab. Theory Related Fields **140** (2008), no. 1-2, 195–205, DOI 10.1007/s00440-007-0062-5. MR2357675

[5] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), no. 2, 559–603, DOI 10.1016/j.jfa.2009.04.016. MR2557947

[6] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe, *Singularity of random symmetric matrices revisited*, Proc. Amer. Math. Soc. **150** (2022), no. 7, 3147–3159, DOI 10.1090/proc/15807. MR4428895

[7] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe, *The singularity probability of a random symmetric matrix is exponentially small*, `arXiv:2105.11384` (2021).

[8] Marcelo Campos, Letícia Mattos, Robert Morris, and Natasha Morrison, *On the singularity of random symmetric matrices*, Duke Math. J. **170** (2021), no. 5, 881–907, DOI 10.1215/00127094-2020-0054. MR4255046

[9] Kevin P. Costello, Terence Tao, and Van Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), no. 2, 395–413, DOI 10.1215/S0012-7094-06-13527-5. MR2267289

[10] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902, DOI 10.1090/S0002-9904-1945-08454-7. MR14608

[11] Asaf Ferber and Vishesh Jain, *Singularity of random symmetric matrices—a combinatorial approach to improved bounds*, Forum Math. Sigma **7** (2019), Paper No. e22, 29, DOI 10.1017/fms.2019.21. MR3993806

[12] Asaf Ferber, Vishesh Jain, Kyle Luh, and Wojciech Samotij, *On the counting problem in inverse Littlewood-Offord theory*, J. Lond. Math. Soc. (2) **103** (2021), no. 4, 1333–1362, DOI 10.1112/jlms.12409. MR4273471

[13] Asaf Ferber, Vishesh Jain, and Yufei Zhao, *On the number of Hadamard matrices via anti-concentration*, Combin. Probab. Comput. **31** (2022), no. 3, 455–477, DOI 10.1017/s0963548321000377. MR4410720

[14] P. Frankl and Z. Füredi, *Solution of the Littlewood-Offord problem in high dimensions*, Ann. of Math. (2) **128** (1988), no. 2, 259–270, DOI 10.2307/1971442. MR960947

[15] Jerrold R. Griggs, Jeffrey C. Lagarias, Andrew M. Odlyzko, and James B. Shearer, *On the tightest packing of sums of vectors*, European J. Combin. **4** (1983), no. 3, 231–236, DOI 10.1016/S0195-6698(83)80017-1. MR725071

[16] Gábor Halász, *On the distribution of additive arithmetic functions*, Acta Arith. **27** (1975), 143–152, DOI 10.4064/aa-27-1-143-152. MR369292

[17] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Singularity of discrete random matrices*, Geom. Funct. Anal. **31** (2021), no. 5, 1160–1218, DOI 10.1007/s00039-021-00580-6. MR4356701

[18] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *On the smallest singular value of symmetric random matrices*, Combin. Probab. Comput. **31** (2022), no. 4, 662–683, DOI 10.1017/s0963548321000511. MR4439776

[19] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240, DOI 10.2307/2152887. MR1260107

[20] Gy. Katona, *On a conjecture of Erdős and a stronger form of Sperner's theorem*, Studia Sci. Math. Hungar. **1** (1966), 59–63. MR205864

[21] Daniel J. Kleitman, *On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors*, Advances in Math. **5** (1970), 155–157 (1970), DOI 10.1016/0001-8708(70)90038-1. MR265923

[22] J. Komlós, *On the determinant of* $(0,1)$ *matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21. MR221962

[23] J. Komlós, *On the determinant of random matrices*, Studia Sci. Math. Hungar. **3** (1968), 387–399. MR238371

[24] J. E. Littlewood and A. C. Offord, *On the Number of Real Roots of a Random Algebraic Equation*, J. London Math. Soc. **13** (1938), no. 4, 288–295, DOI 10.1112/jlms/s1-13.4.288. MR1574980

[25] M. Kac, *On the average number of real roots of a random algebraic equation*, Bull. Amer. Math. Soc. **49** (1943), 314–320, DOI 10.1090/S0002-9904-1943-07912-8. MR7812

[26] Alexander E. Litvak and Konstantin E. Tikhomirov, *Singularity of sparse Bernoulli matrices*, Duke Math. J. **171** (2022), no. 5, 1135–1233, DOI 10.1215/00127094-2021-0056. MR4402560
[27] Galyna V. Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, J. Anal. Math. **145** (2021), no. 1, 257–306, DOI 10.1007/s11854-021-0183-2. MR4361906
[28] Galyna V. Livshyts, Konstantin Tikhomirov, and Roman Vershynin, *The smallest singular value of inhomogeneous square random matrices*, Ann. Probab. **49** (2021), no. 3, 1286–1309, DOI 10.1214/20-aop1481. MR4255145
[29] Mark W. Meckes, *Concentration of norms and eigenvalues of random matrices*, J. Funct. Anal. **211** (2004), no. 2, 508–524, DOI 10.1016/S0022-1236(03)00198-8. MR2057479
[30] Hoi Nguyen and Van Vu, *Optimal inverse Littlewood-Offord theorems*, Adv. Math. **226** (2011), no. 6, 5298–5319, DOI 10.1016/j.aim.2011.01.005. MR2775902
[31] Hoi H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Math. J. **161** (2012), no. 4, 545–586, DOI 10.1215/00127094-1548344. MR2891529
[32] Hoi H. Nguyen and Van H. Vu, *Small ball probability, inverse theorems, and applications*, Erdős centennial, Bolyai Soc. Math. Stud., vol. 25, János Bolyai Math. Soc., Budapest, 2013, pp. 409–463, DOI 10.1007/978-3-642-39286-3_16. MR3203607
[33] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633, DOI 10.1016/j.aim.2008.01.010. MR2407948
[34] Mark Rudelson and Roman Vershynin, *Smallest singular value of a random rectangular matrix*, Comm. Pure Appl. Math. **62** (2009), no. 12, 1707–1739, DOI 10.1002/cpa.20294. MR2569075
[35] Mark Rudelson and Roman Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proc. ICM. Volume III, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602. MR2827856
[36] Emmanuel Rio, *On McDiarmid's concentration inequality*, Electron. Commun. Probab. **18** (2013), no. 44, 11, DOI 10.1214/ECP.v18-2659. MR3070910
[37] Emmanuel Rio, *Small ball probabilities for linear images of high-dimensional distributions*, Int. Math. Res. **2015** (2015), no. 19, 9594–9617.
[38] Mark Rudelson and Roman Vershynin, *No-gaps delocalization for general random matrices*, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776, DOI 10.1007/s00039-016-0389-0. MR3579707
[39] A. Sali, *Stronger form of an M-part Sperner theorem*, European J. Combin. **4** (1983), no. 2, 179–183, DOI 10.1016/S0195-6698(83)80048-1. MR705971
[40] A. Sárközi and E. Szemerédi, *Über ein Problem von Erdős und Moser* (German), Acta Arith. **11** (1965), 205–208, DOI 10.4064/aa-11-2-205-208. MR182619
[41] David Saxton and Andrew Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992, DOI 10.1007/s00222-014-0562-8. MR3385638
[42] Richard P. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property*, SIAM J. Algebraic Discrete Methods **1** (1980), no. 2, 168–184, DOI 10.1137/0601021. MR578321
[43] Michel Talagrand, *A new look at independence*, Ann. Probab. **24** (1996), no. 1, 1–34, DOI 10.1214/aop/1042644705. MR1387624
[44] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012, DOI 10.1090/gsm/132. MR2906465
[45] Terence Tao and Van Vu, *On random ± 1 matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), no. 1, 1–23, DOI 10.1002/rsa.20109. MR2187480
[46] Terence Tao and Van Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628, DOI 10.1090/S0894-0347-07-00555-3. MR2291914
[47] Terence Tao and Van Vu, *A sharp inverse Littlewood-Offord theorem*, Random Structures Algorithms **37** (2010), no. 4, 525–539, DOI 10.1002/rsa.20327. MR2760363
[48] Terence Tao and Van Vu, *The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi*, Combinatorica **32** (2012), no. 3, 363–372, DOI 10.1007/s00493-012-2716-x. MR2965282

[49] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006, DOI 10.1017/CBO9780511755149. MR2289012

[50] Terence Tao and Van H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169** (2009), no. 2, 595–632, DOI 10.4007/annals.2009.169.595. MR2480613

[51] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), no. 2, 593–634, DOI 10.4007/annals.2020.191.2.6. MR4076632

[52] Roman Vershynin, *Invertibility of symmetric random matrices*, Random Structures Algorithms **44** (2014), no. 2, 135–182, DOI 10.1002/rsa.20429. MR3158627

[53] Van Vu, *Random discrete matrices*, Horizons of combinatorics, Bolyai Soc. Math. Stud., vol. 17, Springer, Berlin, 2008, pp. 257–280, DOI 10.1007/978-3-540-77200-2_13. MR2432537

[54] Van H. Vu, *Recent progress in combinatorial random matrix theory*, Probab. Surv. **18** (2021), 179–200, DOI 10.1214/20-ps346. MR4260513

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK

Email address: mc2482@cam.ac.uk

Department of Mathematics, King's College London, Strand, London, WC2R 2LS, UK

Email address: matthew.jenssen@kcl.ac.uk

Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, 851 S. Morgan Street, Chicago, IL 60607-7045, USA

Email address: michelen.math@gmail.com

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK

Email address: jdrs2@cam.ac.uk